

This is a repository copy of *Participant Use of Artificial Intelligence in Online Focus Groups: An Experiential Account*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/219171/>

Version: Published Version

Article:

Stafford, Lucy, Preston, Catherine Elizabeth Jane orcid.org/0000-0001-7158-5382 and Pike, Alex orcid.org/0000-0003-1972-5530 (2024) Participant Use of Artificial Intelligence in Online Focus Groups: An Experiential Account. *International Journal of Qualitative Methods*. ISSN 1609-4069

<https://doi.org/10.1177/16094069241286417>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Participant Use of Artificial Intelligence in Online Focus Groups: An Experiential Account

International Journal of Qualitative Methods

Volume 23: 1–12

© The Author(s) 2024

DOI: 10.1177/16094069241286417

journals.sagepub.com/home/ijqLucy Stafford¹ , Catherine Preston¹, and Alexandra C. Pike¹

Abstract

Large language models (LLMs), one application of artificial intelligence, experienced a surge in users between 2022–2023. During this time, we were conducting online focus groups in which participants insisted on responding using the chat box feature. Based on several chat box responses, we became concerned they were LLM generated. Out of the 42 participants who typed a chat box response during a focus group, we identify 9 as potentially providing LLM generated answers and present their responses with the highest similarity score to an LLM answer. Given the growth and improvement in LLMs, we believe that this issue is likely to increase in frequency. In response to this, in this article we reflect on (1) strategies to prevent participants from using LLMs, (2) indicators LLMs may be being used, (3) the fallibility of identifying LLM generated responses, (4) philosophical frameworks that may permit LLM responses to be incorporated into analyses, and (5) procedures researchers may follow to evaluate the influence of LLM responses on their results.

Keywords

artificial intelligence, large language models, online focus groups, interpretivism, post-positivism

Introduction

Large language models (LLMs) are one application of generative artificial intelligence. Generative artificial intelligence refers to technology that generates human-like content in response to prompts (Lim et al., 2023). Their responses depend on the data it is trained on. LLMs are trained on web text and can respond to a prompt with text, and in some cases, images (Wu et al., 2023; Yang et al., 2023). More specifically, LLMs are trained to recognise statistical patterns in vast amounts of existing data, such as that available on the internet (Kasneci et al., 2023). Once trained, the model can then be given an input (i.e. prompt), such as a question or request, to respond to. To respond to an input, the model first pulls out tokens from the input. Tokens refer to a unit of text, which can be as small as a singular character and as large as one word, depending on the language and tokenisation method used. Input tokens are then converted into a unique number, and the model predicts the most probable next unique number, which is then decoded into a token and will appear in the output as human-readable text (Trott, 2024). One example of an LLM is

ChatGPT. The number of people using ChatGPT increased from 1 million in November 2022 (DeVon, 2023) to 180.5 million users in August 2023 (Tong, 2023). At the time ChatGPT was experiencing a surge in users, we were conducting online focus groups investigating how sensations from the body made people feel about its appearance. In 10 of the 12 focus groups we conducted, several participants insisted on answering using the chat box feature rather than their microphone.

Adapting qualitative methods of data collection to an online setting is beneficial in several ways. Moving qualitative methods online limits the need for participants to travel,

¹Department of Psychology, University of York, UK

Corresponding Author:

Lucy Stafford, Department of Psychology, University of York, York YO10 5DD, UK.

Email: lucy.stafford@york.ac.uk

Data Availability Statement included at the end of the article.



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

increasing accessibility and geographical reach (Pellicano et al., 2024). From the researcher perspective, limiting participant travel is particularly beneficial for focus groups, which include multiple participants engaging in a real-time group discussion focussed on a facilitator's questions (Guest et al., 2023). This is because the risk of travel disruption is negated, eliminating a source of focus group non-attendance and in turn cancellation (Stewart & Shamdasani, 2017). One problem with online focus groups is that participants, even if explicitly asked to use their microphones, might attend focus groups without doing so (Sharma et al., 2024). This is something that we encountered during our own online focus groups. Participants cited microphone problems, poor internet connection, and concerns over their voice being recognised outside the session by other attendees as reasons to not use their microphone. A number of participants then contributed using the chat box. Reasons why participants may use the chat box feature to respond to focus group questions rather than their microphone include feeling more comfortable when disclosing sensitive information (Walther & Boyd, 2002), speech difficulties (Williams et al., 2012), social anxiety (Yarmand et al., 2021), and a fear of being overheard by family/housemates (Morris et al., 2021). Based on several chat box responses, we encountered a novel concern not yet discussed in online focus groups: could participants be using the chat box to provide LLM generated responses?

In this article, we provide evidence that participants in our online focus groups may have been providing LLM generated responses, detail preventative measures to discourage participants from using LLMs or at least use them appropriately, evidence indicators that LLMs may be being used, discuss the fallibility of identifying LLM generated responses, examine scientific philosophical frameworks that may permit LLM responses to be incorporated into analyses, and describe the procedures researchers may follow to evaluate the influence of LLM responses on their results.

Methods

Participant Recruitment

Participants expressed their interest in taking part in a focus group via an online questionnaire. We wanted to recruit people who identified as having an eating disorder, gastric disorder, or neither disorder. We did not ask for proof of diagnosis as we wanted to honour lived experience and establish a relationship built on trust between the researcher and participant. People can experience eating disorders without an official diagnosis as they are difficult to diagnose in the first place (Dalle Grave, 2011). For example, people with an eating disorder might not show the stereotypical signs of disordered eating needed for an official diagnosis, such as meeting the low weight requirement for a diagnosis of anorexia nervosa (Tse et al., 2022). People can also experience gastric disorder symptoms without an

official diagnosis due to the long time it takes to receive one (Blackwell et al., 2021).

The questionnaire link was posted to relevant subreddits (with moderator's permission), closed Facebook groups, and Twitter/X. Physical posters were also distributed around the university campus. The aim of recruitment was to conduct online focus groups to explore how hunger, satiation, and fullness are experienced in the body and how they impact feelings towards the bodily appearance. This included asking participants how they physically and emotionally experienced states of hunger, satiation, and fullness and how this made them feel about their body. Focus groups were chosen as the appropriate methodology so we could capture a range of bodily experiences during a singular session (Rabiee, 2004).

Materials

Participants were given the option to provide demographic details (age, gender identity, highest level of education, ethnicity, weight, and height (for BMI to be calculated)). They also had the choice to answer questionnaires that would allow us to gauge the severity of their disorder. For eating disorder participants, this included the Eating Disorder Examination Questionnaire-6 (Fairburn, 2008). For gastric disorder participants, this included the Gastrointestinal Quality of Life Index (Eypasch et al., 1995). Information on these measures can be found in the [Supplementary Materials](#).

Procedure

Participants read an information sheet and gave informed consent to take part in the expression-of-interest survey hosted by Qualtrics (Provo, UT). In this, they created a unique identifier code made up of the last 2 letters of their first name, the last 2 digits of their mobile phone number, the last two letters of the street they live on, and the last two digits of their birth year so their questionnaire responses could be linked to the focus group they attended. They were also given the option to provide demographic details and answer questionnaires that would allow us to gauge the severity of their disorder (see Materials section). They were then redirected to another Qualtrics (Provo, UT) online questionnaire which allowed us to collect their email address (so it was not directly linked to their responses on the expression-of-interest survey) and provide their availability. They were then given access to the debrief document which detailed the aims of this project and a list of resources if they needed further support.

Participants were then selected to take part in a focus group based on the availability they had given in the online questionnaire. They were sent an invite via email which provided them with the date and time of the focus group, the focus group link, a Zoom help link, what to expect during a session, and were all encouraged to use their cameras and microphones during the session. Data were collected May-June 2023 (inclusive) and they could claim a £10 Amazon voucher for

attending a focus group. Full ethical approval for the questionnaire and focus group was gained from the University of York Psychology Ethics Committee April 2023 (ref: 23012). Respondents gave written consent before starting the questionnaire and attending a focus group.

Twelve 60-min focus groups with 3–12 people (average = 6.5) were conducted online via Zoom. This wide range in the size of the focus groups was a result of discrepancies between the number of participants who confirmed their attendance and the number of participants who then showed up. The sessions started with an introduction and ethical reminders from the researcher. Participants were encouraged to use their microphone and camera if comfortable/possible. A moderator, one of the supervisors of the project, joined to collect participant's unique identifier codes (created in the expression-of-interest survey) via the chat box. This is so we could connect our focus group participants to their answers on the expression-of-interest survey, allowing us to contextualise our research findings. The researcher then started recording the session and asking questions on the topic guide. They were asked the following questions in which chat box responses were analysed:

- *When you have eaten to the point where your hunger has been satisfied but your stomach does not feel uncomfortable, how does this make you feel emotionally?*
- *When you are hungry, how does this make you feel emotionally?*
- *When you have eaten to the point where you could not eat anymore, how does this impact how you feel about your body and body size?*
- *When you have eaten to the point where your hunger has been satisfied but your stomach does not feel uncomfortable, how does this make you feel about your body and body size?*
- *When you have eaten to the point where you could not eat anymore, how does this make you feel emotionally?*
- *When you are hungry, how does this make you feel about your body and body size?*

After 1 hour, participants were thanked and asked if there was anything else that they would like to add to the discussion. They were also told they would receive an email containing instructions on how to redeem their e-gift card and a debrief document. After online focus groups, participants were sent a follow up email containing the debrief details and instructions on how to get their e-gift card. It also asked if there was anything they would like to add to their responses and for feedback.

Participants

Participants were eligible to take part if they were aged 18+, live in the UK, fluent in English, identified as having an eating disorder (eating disorder groups), identified as having a gastric

disorder (gastric disorder groups), or identified as having no eating or gastric disorder (no disorder groups). Exclusion criteria involved having been involuntarily committed to eating disorder treatment in the last 6 months (in-patient or outpatient). People who had experienced involuntary care were excluded from this research due to concerns about worsening their condition (Sala et al., 2023).

1022 people accessed the expression-of-interest survey to take part in an online focus group. 489 provided an email address to be invited to a focus group. 78 took part in an online focus group. 57 focus group participants (ED $n = 24$; GD $n = 15$, ND $n = 18$) chose to answer the demographic questions, and hence could be matched to their responses on the expression-of-interest survey. See sample demographics for our online focus groups participants in Table 1. We collected these demographic details to contextualise our findings and assess the diversity of the sample. It was important that we assessed age, gender identity, educational level, and ethnic origin as these factors can influence how someone feels about their body; the topic being investigated where the data for this analysis came from (Gluck & Geliebter, 2002; Kozar & Damhorst, 2009; McLaren & Kuh, 2004; Richburg & Stewart, 2024).

Data Analysis and Results

We did not use AI detectors in this analysis as the responses from participants were not long enough for LLM generated responses to be reliably detected (Chakraborty et al., 2023). For example, Turnitin, the most robust AI detector currently available (Weber-Wulff et al., 2023), requires at least 350 words. Therefore, we compared participant responses to ChatGPT's answers to the same questions (Rahman & Watanobe, 2023).

In 10 of the 12 focus groups conducted, 42 participants out of the total 78 participants who took part in a focus group typed at least one response to one of the above questions in the chat box. These responses and ChatGPT's responses to the same questions can be downloaded from <https://osf.io/4nzwh/>. To make this analysis as unbiased as possible, we asked ChatGPT the exact same question (see questions in Procedure section), without any requests to make it sound human or to shorten it (i.e. 'paraphrase' it).

Similarity scores between each participants' typed response and the ChatGPT answer were calculated using the levenshteinSim function in the RecordLinkage package (Winkler, 1990) in R version 4.3.1 (R Core Team, 2013). As string comparison is case sensitive, participant and ChatGPT answers were transformed into lowercase strings. An average similarity score for each participant across the 6 open questions asked in our focus groups was then calculated, allowing us to identify suspicious participants. The number of questions answered by each participant via the chat box varied between 1 and 6, meaning the average similarity score for participants who typed an answer to just one response was based on that

Table 1. Focus Group Participant Demographic Characteristics by Population.

	ED <i>n</i> = 24	GD <i>n</i> = 15	ND <i>n</i> = 18
Age in years (mean)	26.29 (SD = 4.03; range = 20.00–48.00) Undisclosed <i>n</i> = 4	27.90 (SD = 8.39; range = 20.00–48.00) Undisclosed = 5	23.00 (SD = 2.32; range = 20.00–26.00) Undisclosed <i>n</i> = 5
BMI (mean)	22.95 (SD = 5.62; range = 13.59–36.42) Undisclosed <i>n</i> = 8	22.40 (SD = 8.22; range = 19.38–30.22) Undisclosed = 5	23.89 (SD = 4.94; range = 19.49–38.10) Undisclosed <i>n</i> = 5
University Undergraduate degree or higher (%)	73.68 Undisclosed <i>n</i> = 5	70 Undisclosed <i>n</i> = 5	55.56 Undisclosed <i>n</i> = 5
White English/Welsh/Scottish/Northern Irish/British/ European (%)	61.91 Undisclosed <i>n</i> = 3	46.2 Undisclosed <i>n</i> = 2	40 Undisclosed <i>n</i> = 3
Identified as female (%)	65 Undisclosed <i>n</i> = 4	54.55 Undisclosed <i>n</i> = 4	84.62 Undisclosed <i>n</i> = 5

Note. For brevity we have reported the percentages of respondents categorised within the majority group for Highest educational level, Ethnic origin, and Gender identity. Undisclosed refers to people who chose not to answer the question.

lone response. Participants with an average similarity score above 10% were identified as likely LLM responses in line with the text similarity rate proposed by the [British Medical Journal \(2023\)](https://www.bmj.com/about-bmj/publishing-model) to suggest a redundant publication (<https://www.bmj.com/about-bmj/publishing-model>). We acknowledge that 10% may appear a low similarity score to identify suspicious participants, but there are four factors that we believe reduced the similarity scores. First, we do not know which LLM was used by participants. Second, it appears that participants were pasting only parts of the ChatGPT response (see data at <https://osf.io/4nzwh/>). Third, we do not know the exact prompts participants used. Fourth, LLMs are programmed to produce a different answer even when the same question is asked (Cowen & Tabarrok, 2023). With these justifications considered, 9 out of the 42 participants (21.43%) who typed a response to at least 1 of the 6 questions analysed were identified as having provided responses that, on average, were equal to or more than 10% similar to ChatGPT answers. For brevity, we have included the answer with the highest similarity from each of these 9 participants below.

Focus Group 2, Speaker 6. It helps foster a positive body image and appreciation for my body's ability to communicate its needs effectively. Feeling satisfied without discomfort emphasizes the importance of listening to my body rather than focusing solely on body size.

- Similarity score for this response: 12.85%

Focus Group 5, Speaker 4. When I'm hungry, it can affect my emotions in different ways. Sometimes, I feel a sense of frustration or irritability because my body is signalling that it needs nourishment. Other times, I might feel a bit anxious or unsettled until I can satisfy my hunger. However, I also recognize that hunger is a natural bodily sensation and try to address it calmly and responsibly.

- Similarity score for this response: 15.49%

Focus Group 6, Speaker 2. Well, the feeling of being extremely full after eating to my limits is a mix of physical discomfort and a sense of satisfaction. On one hand, there's a heaviness and bloated sensation that can be uncomfortable, almost as if my stomach is stretched to its capacity. This physical discomfort might make me feel a bit lethargic or even slightly nauseous.

- Similarity score for this response: 16.48%

Focus Group 8, Speaker 4. Hunger motivates me to prioritize self-care and provide my body with the nourishment it requires. Instead of focusing on body size, I focus on nourishing my body with balanced and nutritious meals, which contributes to my overall well-being. Hunger prompts me to approach food with mindfulness. Instead of using hunger as an opportunity to criticize or judge my body, I focus on making nourishing choices and listening to what my body truly needs.

- Similarity score for this response: 16.75%

Focus Group 8, Speaker 3. When I feel hungry, it's a reminder that my body has its natural way of signaling its need for nourishment. I try to view hunger as a normal physiological response rather than associating it with negative feelings about my body or size. I feel I might reduce in size and weight if I continue staying hungry for long.

- Similarity score for this response: 13.46%

Focus Group 8, Speaker 5. Feeling satisfied without discomfort positively impacts my overall emotional well-being. It eliminates any guilt or negative feelings that may arise from overeating or under-eating, allowing me to enjoy a balanced relationship with food and nourishment.

- Similarity score for this response: 13.09%

Focus Group 8, Speaker 2. Rather than feeling heavy or weighed down, I experience a sense of lightness after a satisfying but not

overly filling meal. It's a pleasant feeling that allows me to continue my activities without feeling sluggish or lethargic. Feeling comfortably full without discomfort uplifts my mood and contributes to a positive outlook on eating and my overall well-being.

- Similarity score for this response: 17.31%

Focus Group 8, Speaker 1. For me, instead of dwelling on the negative feelings, I try to focus on practising self-care and engaging in activities that promote a healthy mindset. This might involve engaging in physical activities I enjoy, finding ways to relax and de-stress, or reminding myself of the other positive aspects of my body beyond just its size. Building a positive body image is an ongoing process that involves self-acceptance, self-care, and cultivating a healthy relationship with food.

- Similarity score for this response: 16.23%

Focus Group 8, Speaker 6. When I have eaten to the point where I couldn't eat anymore, it can sometimes have an impact on how I feel about my body and body size. It's important to note that this feeling can vary from person to person, and everyone's experience may be different. In some instances, overeating can lead to feelings of discomfort or guilt, particularly if I've overindulged or eaten in a way that doesn't align with my personal health goals.

- Similarity score for this response: 15.92%

One source of similarity in these answers comes from the construction of the opening sentence. Many of the responses start with a sentence with two independent clauses. For example, *"Rather than feeling heavy or weighed down, I experience a sense of lightness after a satisfying but not overly filling meal"* (Focus group 8, speaker 2), *"When I feel hungry, it's a reminder that my body has its natural way of signaling its need for nourishment."* (Focus group 8, speaker 3), and *"When I have eaten to the point where I couldn't eat anymore, it can sometimes have an impact on how I feel about my body and body size."* (Focus group 8, speaker 6).

It is also worth noting that several responses also start by rephrasing the question they were asked, in particular starting with the word "when". This includes, *"When I have eaten to the point where I couldn't eat anymore, it can sometimes have an impact on how I feel about my body and body size."* (Focus group 8, speaker 6), *"When I feel hungry, it's a reminder that my body has its natural way of signaling its need for nourishment."* (Focus group 8, speaker 3), *"When I'm hungry, it can affect my emotions in different ways."* (Focus group 5, speaker 4).

Another similarity, pointed out by a reviewer, concerns the American spellings in responses, highlighted in bold in the following extracts *"Hunger motivates me to **prioritize** self-care and provide my body with the nourishment it requires...Instead of using hunger as an opportunity to **criticize** or judge my body, I focus on making nourishing choices and*

listening to what my body truly needs." (Focus group 8, speaker 4), *"When I feel hungry, it's a reminder that my body has its natural way of **signaling** its need for nourishment"* (Focus group 8, speaker 3), *"Feeling satisfied without discomfort **emphasizes** the importance of listening to my body rather than focusing solely on body size."* (Focus group 2, speaker 6). ChatGPT commonly uses American spelling, potentially indicating the use of an LLM given we requested that our participants lived in the UK so we could process their e-gift card.

Another commonality in these answers concerns the inclusion of an alternative point of view in responses. Examples of this include, *"Rather than feeling heavy or weighed down, I experience a sense of lightness after a satisfying but not overly filling meal."* (Focus group 8, speaker 2), *"For me, instead of dwelling on the negative feelings, I try to focus on practising self-care and engaging in activities that promote a healthy mindset."* (Focus group 8, speaker 1), and *"I try to view hunger as a normal physiological response rather than associating it with negative feelings about my body or size."* (Focus group 8, speaker 3)

A final observation refers to the use of soft, cautionary language, using words such as "may" and "might". For example, *"Other times, I might feel a bit anxious or unsettled until I can satisfy my hunger."* (Focus group 5, speaker 4), *"This physical discomfort might make me feel a bit lethargic or even slightly nauseous."* (Focus group 6, speaker 2), and *"It eliminates any guilt or negative feelings that may arise from overeating or under-eating."* (Focus group 8, speaker 5). We would also like to point to the inclusion of a full cautionary disclaimer, *"It's important to note that this feeling can vary from person to person, and everyone's experience may be different."* (Focus group 8, speaker 6).

Discussion

Twenty-one percent of participants who typed a response to at least 1 of the 6 questions asked were identified as potentially providing LLM responses. We will now discuss why participants using LLMs is problematic, provide indicators that participants may be using LLMs, and discuss how researchers could resolve this issue.

Is it Problematic for Participants to Use LLMs to Provide Responses?

Online focus groups may be particularly vulnerable to LLM use due to the special ethical considerations that come with them. As focus groups include a group of people that could potentially recognise each other (Sim & Waterfield, 2019), researchers often cannot have microphone use as a strict requirement for participation, instead only being able to encourage usage (Sharma et al., 2024). As a result, participants can choose to respond via a chat box function. These chat boxes allow users to paste messages into them,

which we believe is how participants are providing LLM responses.

We understand that LLMs might be useful for participants to contribute to an online discussion who struggle to verbalise their thoughts and feelings. Then, in the best case scenario, they are using aspects of the generated answer that capture their experience. Participants who do not have English as their first language may be using LLMs to better communicate their experiences (Shahriar & Hayawi, 2023). Indeed, this may be the case considering we were striving for diversity and inclusivity. It may also be true for one of our recruited populations in particular: participants with eating disorders. This is because eating disorders are associated with difficulties in recognising and interpreting feelings (alexithymia; Westwood et al., 2017). The impact of malnutrition on the cognition of two of our populations (those with eating disorders or gastric disorders) may also mean LLMs are used by participants to help articulate their thoughts (Himmerich et al., 2021; Lin & Micic, 2021). We acknowledge that most participants are willing research collaborators who want to provide meaningful data, but may need the assistance of LLMs to help them do this. In this case, LLM use does not necessarily make them fraudulent or imposter participants. Future research might consider investigating participants' motivations for using LLMs for more insight as to whether these participants are fraudulent or not. Researchers could also ask that participants disclose use of an LLM, like what is done in academic journals (Editorials, 2023), and provide guidance on how they would like them to be used. If participants are asking LLMs the exact question being asked in focus groups, this could include asking that participants (1) only use the parts of the answers that captures their experience, (2) make edits to the answers to make them more relevant to their experience, and (3) include more information in the initial prompt to better personalise their answer (Lingard, 2023).

However, participants using LLMs to provide a response can be problematic. For one, the data produced by LLMs is not 'new'. LLMs are trained on existing data, meaning their responses are essentially the patterns they have detected in data that has been previously collected (Thirunavukarasu et al., 2023). Thus, LLM responses can be considered as a rewording of the data that already exists, surely failing to provide any new insights into a research topic. Second, LLMs provide answers representative of Western, Educated, Industrialised, Rich, and Democratic (WEIRD) participants (Atari et al., 2023; Cowgill et al., 2020). Thus, if non-WEIRD participants use LLMs, they might not be providing an answer that best sums up their own authentic experience. This bias for WEIRD-like responses by LLMs most likely arises from training data representative of mostly WEIRD societies (Atari et al., 2023). Third, LLMs are unlikely to provide answers that capture the experiences of clinical groups. Given the relative scarcity of this data in an open 'format' (De Lusignan et al., 2014), it is hard to imagine they have undergone extensive training to be able to respond in a manner characteristic of a

particular clinical group. Fourth, LLMs can produce responses that stereotype certain populations. It is important to note that most LLMs will implement safety measures to prevent harmful responses. For instance, safety measures can be implemented to stop LLMs from responding to obvious damaging requests (Ayyamperumal & Ge, 2024). Indeed, when given a *direct* request to respond to a harmful prompt, ChatGPT will refuse to answer (Yu et al., 2024). Thus, obvious intentions to push damaging stereotypes will be halted by LLMs. However, LLMs can produce stereotyping responses without even being requested to do so (Deshpande et al., 2023; Gehman et al., 2020).

How can Researchers Prevent, Identify, and Deal With Large Language Model Responses in Their Data?

There are many excellent articles that recommend actions to prevent research data being infiltrated by fraudulent participants (see Davies et al., 2023; Pullen Sansfaçon et al., 2024). However, as noted above, participants who use LLMs may not necessarily be fraudulent. To our knowledge, just one other paper has provided recommendations on how to prevent the use of LLMs by research participants, and this was for online qualitative surveys (Gibson & Beattie, 2024). However, to our knowledge, there are no recommendations as to how researchers can prevent participants using LLMs in focus groups. Hence, we provide steps researchers might take to prevent participants from using LLMs.

One way researchers could prevent participants using LLMs is to tighten participation requirements. Researchers could make it clear that in order to participate, participants must have a working microphone and not respond using the chat box feature. Researchers could also make it explicit that participants should not use LLMs for their responses, or provide guidance about what could be used (i.e. to help articulate their own experiences and feelings rather than to generate answers they think the researchers are looking for). However, there are a number of reasons why these preventative measures may not be feasible, including ethical, accessibility, and practicality reasons. In focus groups, individuals by definition must interact with strangers. Using a microphone may impact individuals' anonymity, which is of particular importance when sensitive questions are being asked. Further, the target population may have a higher representation of difficulties speaking aloud, for example stutters or tics. Thus, requiring the use of a microphone might reduce the diversity of participants and exclude the experiences of these individuals. It is also unclear how a researcher should proceed if participants attend focus groups without a working microphone. It is not ethically or practically possible to ensure participants use their microphones. The only step a researcher could take would be to remove participants from the session. However, this could impact on how other participants experience the session and their willingness to respond (Drysdale et al., 2023). It may also reduce the numbers

who attend a focus group to an unsustainable level, perhaps warranting cancellation. This would be unfair for compliant participants. Given these drawbacks, imposing strict participation requirements may not be suitable.

A related alternative to imposing stricter participation requirements is to change the settings of the online video conferencing software. The chat box could be disabled for participants so they must use their microphone. There are, however, significant problems with this solution. Disabling participant use of a chat box reduces accessibility for people who have speech difficulties, as mentioned above. Further, participants may also have hearing difficulties, making a chat box useful for the researcher to paste their questions into. If a chat box is disabled, this reduces accessibility and excludes certain experiences from the research findings.

The harm that strict participation requirements and changes in chat box settings pose to accessibility means researchers may instead prevent participants from using LLMs through screening procedures. Researchers could invite prospective participants to a screening interview, a short video call before the date of data collection. This may help participants feel more comfortable using their microphones, and means anyone without a working microphone can be removed as a potential participant (Ridge et al., 2023). At least with a microphone on, researchers could tell if participants were reading out written answers (i.e. no pauses, repetition, or fillers) that have perhaps come from an LLM. However, there is a chance that participants could just read the LLM answer and then respond with their remembered points (rather than reading it out line for line). This would be harder to detect, but may be preceded by a long pause whilst the participant pastes the prompt into an LLM, waits for the response, reads the response, and then answers using the microphone. Like the above preventative measures, these too are flawed. Even if participants in the screening call have their microphone and camera on, they may still choose not to have them on in the actual focus group session. The screening call may have made them comfortable having their camera and/or microphone on around the researcher, but does not resolve concerns about discussing sensitive topics around other participants. Thus, a better screening method may be to run a screening focus group. Participants may then have the chance to become more comfortable with the other people who will be attending, and thus hopefully feel comfortable having their camera and microphone on when it comes to the data collection session. Although useful in theory, the practicality of running screening focus groups make them less appealing. Like screening interviews, screening focus groups demand more researcher time, and additional participant payment. Screening focus groups may also pose an additional burden on participant time. Further, even if participants in the screening call have their microphone on, they may still choose not to have it on in the actual focus group session.

Given the important limitations of the preventative measures detailed above, researchers may not choose to use them. The focus then turns to detecting LLM generated responses. For responses longer than 350 words, a researcher may consider using Turnitin to detect LLM responses (Weber-Wulff et al., 2023). However, this is not suitable for shorter answers like those in the current study. Further, a drawback of the analysis included in this paper is that it compares participant responses to just one response from an LLM, and LLMs can produce many different responses to the same question (Cowen & Tabarrok, 2023). This makes the analysis less sensitive in detecting AI responses. Therefore, we detail indicators that a participant may be providing LLM generated responses.

The first sign a participant may be using ChatGPT is tone change. We had participants switch from an informal tone (e.g. spelling mistakes, no punctuation, incorrect spellings, grammatical errors) to a more formal tone (e.g. capitalisation, correct spellings, punctuation, and correct grammar). We had participants firstly answering (initial) questions in a very informal manner, but when the questions became more complex, they began giving answers that were very formal, with descriptive adjectives, correct spellings, punctuation, and capital letters where appropriate. Gibson and Beattie (2024) and Fleckenstein et al. (2024) have noted a lack of mistakes (or ‘typos’) as an indicator of an LLM response. Kabir et al. (2023) and Cui et al. (2023) also note that LLM responses are often very formal (unless asked not to be). We provide an example of this indicator from our own focus group responses below:

- **Interviewer:** *Do you experience gastric sensations in your day-to-day life?*
- **Participant typed:** *Yes, i do experience it*
- **Interviewer:** *When you have eaten to the point where you could not eat any more, what kind of sensation do you perceive from your gastric system?*
- **Same participant typed:** *When I reach the point of being unable to eat any more, I typically experience a sensation of fullness or satiety in my gastric system. My stomach feels distended or stretched, and I may feel a sense of pressure or discomfort in my abdominal area.*

The second clue a participant may be using an LLM is they are providing answers in a very short amount of time. Everyone has different processing and typing speeds, but providing a formal answer in such a short amount of time is suspicious, especially if they were previously taking the same amount of time or longer to provide very basic answers. In the example below, this participant took 21 seconds to think about the question and then supposedly type a 37-word answer out with correct grammar, spelling, and punctuation. This is surprising given the average words per minute to merely copy a sentence is 52 (Dhakal et al., 2018).

- **Interviewer at 12:24:36 pm:** *When you have eaten to the point where you could not eat anymore, how does this impact how you feel about your body and body size?*
- **Participant at 12:24:57 pm:** *It can make me feel a bit indulgent, like I've treated myself to something enjoyable. There's a certain pleasure in indulging in delicious food, even if it means reaching the point of being unable to eat more.*

A third red flag a participant may be using a LLM is that when they are asked to expand on their answer, they do not give more details (Pullen Sansfaçon et al., 2024; Sharma et al., 2024). An example of this can be found below:

- **Interviewer said:** *When you have eaten to the point where you could not eat anymore, how does this make you feel emotionally?*
- **Participant typed:** *Rather than allowing the sensation of fullness to trigger self-criticism or negative body image, I practice self-compassion and remind myself that listening to my body is a fundamental part of self-care. I strive to cultivate a mindset that values overall well-being and respects the natural signals my body sends.*
- **Interviewer said:** *How is listening to your body important to your self-care?*
- Participant did not reply.

The fourth sign a participant may be using an LLM is that they give vague, general answers that do not draw on any concrete, lived experience (Cotton et al., 2024; Gao et al., 2023; Gibson & Beattie, 2024; Rahman & Watanobe, 2023). For example, they do not describe a particular time they felt the emotion or sensations described. This is exemplified below:

- **Interviewer:** *When you have eaten to the point where you could not eat anymore, how does this impact how you feel about your body and body size?*
- **Participant typed:** *When I reach the point of being unable to eat any more, it can be a reminder to practice mindful eating and listen to my body's signals more attentively. While it might make me feel a bit dissatisfied with my body in the moment, I understand that it's essential to nurture a positive relationship with food and my body, focusing on balance and moderation.*

A drawback of the indicators mentioned above is that they may become less relevant as people become more adept at prompting LLMs. For example, they can be prompted to provide text with typos, which makes the first indicator less relevant (Ladha et al., 2023). Another significant problem with the indicators listed above is that they are subjective. It is near-impossible to definitely know if a participant is using an LLM to provide answers. Researchers looking for more concrete

evidence may ask the same questions they asked the participants to a LLM and look out for similarities between answers. However, it is rare that the exact same response will be given by a LLM, as they are programmed to give a different answer each time (Cowen & Tabarrok, 2023). Further, LLM detection tools have not yet proved robust enough to provide definite evidence of LLM use (Elkhatat et al., 2023). Together, the recommendations we give for researchers to detect LLM generated participant responses rely on combinations of imperfect predictors and researcher discretion.

Nevertheless, if a researcher believes they have detected an LLM generated answer from a participant, they may be unsure about what to do with this data. Our first instinct may be to remove suspect responses. However, because a researcher cannot definitely know if a response is LLM generated, this runs the risk of researchers 'cherry picking' the data to be included in the analysis. The traditional philosophy underlying qualitative research may offer an alternative perspective to help answer this dilemma. Interpretivism (or constructivism) is the major philosophy that underlies qualitative analysis (Petty et al., 2012). According to interpretivism, research data is subjective and socially constructed by the researcher and participants, meaning it represents *a* reality, not *the* reality of the phenomenon under investigation (Lincoln et al., 2011). If researchers accept that their data is representative just of the personalised interaction between themselves and the participants at that moment in time (i.e. *a* reality of many possible realities, not *the* only reality), they may be more forgiving of responses produced by LLMs. LLM generated responses from participants may not be considered false or noisy data, but instead part of a subjective reality (i.e. data) constructed by a participant using a LLM. Researchers may then consider conducting response validation (or member checks) with a member of each focus group who answered using the microphone to ensure that, even with the inclusion of the suspected LLM generated responses in the data set, the analysis captures how a phenomenon is subjectively experienced.

Although unorthodox in qualitative research, researchers might adopt a positivist perspective, whereby research data represents *the* reality or truth of the phenomena being explored (Lincoln et al., 2011). From this perspective, LLM generated responses should not be incorporated into the research data because they are not responses born out of direct experience with the phenomenon under investigation, and so do not represent the reality, or truth, of that phenomenon. Nevertheless, if we consider the fact that LLMs are trained on human data, a *post*-positivist researcher may be able to incorporate LLM responses into their dataset. This is because a *post*-positivist researcher understands that there is a truth, or objective reality of a phenomena, but it is difficult to access (Ponterotto, 2005). Therefore, the *post*-positivist researcher may believe LLM responses *can* represent the objective reality of the phenomenon as the responses are based on training data produced by humans, but it is tough identifying which responses do in fact represent the reality. *Post*-positivist

researchers could check if LLM generated responses in their data set represent an objective truth by conducting a qualitative version of a sensitivity analysis on existing data or new data. A sensitivity analysis on the existing data would include conducting analyses with and without the suspected LLM responses in the data set, and then comparing the findings of both analyses (e.g. themes and subthemes if doing a thematic analysis) to see if the data set containing the LLM responses is producing findings different to the data set excluding LLM responses. A sensitivity analysis on new data would include collecting data in-person and then comparing findings from the in-person data to findings from the online data that included suspected LLM responses. If there are substantial differences in the findings between data sets, a researcher may decide that LLM responses do not reflect the truth of the phenomena being studied, and so exclude these responses from analysis. LLM responses may not reflect the reality of the phenomena if it has wrongly predicted their output.

Conclusions

The popularity of LLMs has massively increased. From our recent experience, we believe participants are using LLMs to take part in online focus groups. Our analysis found that this may indeed be the case; 9 out of 42 participants who typed a response during a focus group were found to potentially be sending LLM generated messages. We note some similarities in the answers of these participants, including the construction of the opening sentence, starting a response by paraphrasing the question, American spellings, the inclusion of an alternative point of view, and cautionary language. Participants using LLMs could be problematic in several ways, with implications for research in terms of whether it provides new insights, captures the experiences of clinical populations and people outside of WEIRD societies, and the spread of misinformation and harmful stereotypes. Therefore, we have provided measures to prevent LLM use by participants. In recognising the drawbacks of these measures, we have also detailed potential indicators that participants are using LLMs. However, we have also acknowledged that identifying data produced by LLMs is fallible and could result in ‘cherry picking’ data. We thus note philosophical frameworks that may allow researchers to incorporate LLM data (interpretivism, post-positivism) into their findings and how to decide whether to do so (member checks, sensitivity analysis).

Acknowledgements

We wish to thank the participants who took part in the project repurposed for this manuscript.

Author Contributions

L.S conceived the study, conducted the focus groups, analysed the data, wrote the original draft, and reviewed and edited the final manuscript. C.P and A.C.P helped moderate the focus groups,

supported the analysis, revised and edited the manuscript, and supervised the associated project. All authors read and agreed to the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the ESRC PhD scholarship ES/P000746/1.

Ethical Statement

Ethical Approval

The data re-used for this article obtained ethical approval from the University of York Departmental Ethics Committee (ref: 23012).

Informed Consent

Written informed consent for publication was provided by the participant(s).

ORCID iD

Lucy Stafford  <https://orcid.org/0000-0002-3204-7685>

Data Availability Statement

Please see the associated data and analysis scripts at: <https://osf.io/4nzwH/>.

Supplemental Material

Supplemental material for this article is available online.

References

- Atari, M., Xue, M. J., Park, P. S., Blasi, D., & Henrich, J. (2023). Which humans? <https://doi.org/10.31234/osf.io/5b26t>
- Ayyamperumal, S. G., & Ge, L. (2024). *Current state of LLM risks and AI guardrails*. arXiv. <https://doi.org/10.48550/arXiv.2406.12934>
- Blackwell, J., Saxena, S., Jayasooriya, N., Bottle, A., Petersen, I., Hotopf, M., Alexakis, C., & Pollok, R. C., POP-IBD study group. (2021). Prevalence and duration of gastrointestinal symptoms before diagnosis of inflammatory bowel disease and predictors of timely specialist review: A population-based study. *Journal of Crohn's and Colitis*, 15(2), 203–211. <https://doi.org/10.1093/ecco-jcc/jjaa146>
- British Medical Journal. (2023). *Publishing Mode; Preprints previously published material and duplication*. Redundant Publication. <https://www.bmj.com/about-bmj/publishing-model>
- Chakraborty, S., Bedi, A. S., Zhu, S., An, B., Manocha, D., & Huang, F. (2023). *On the possibilities of ai-generated text detection*. arXiv. <https://doi.org/10.48550/arXiv.2304.04736>

- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education & Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Cowen, T., & Tabarrok, A. T. (2023). How to learn and teach economics with large language models, including Chat GPT. <https://ssrn.com/abstract=4391863>
- Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020, July). Biased programmers? Or biased data? A field experiment in operationalizing AI ethics. In *Proceedings of the 21st ACM conference on economics and computation* (pp. 679–681). Association for Computing Machinery. <https://doi.org/10.1145/3391403.3399545>
- Cui, W., Zhang, L., Wang, Q., & Cai, S. (2023). *Who said that? Benchmarking social media AI detection*. arXiv.
- Dalle Grave, R. (2011). Eating disorders: Progress and challenges. *European Journal of Internal Medicine*, 22(2), 153–160. <https://doi.org/10.1016/j.ejim.2010.12.010>
- Davies, M. R., Monssen, D., Sharpe, H., Allen, K. L., Simms, B., Goldsmith, K. A., Byford, S., Lawrence, V., & Schmidt, U. (2023). Management of fraudulent participants in online research: Practical recommendations from a randomized controlled feasibility trial. *International Journal of Eating Disorders*, 57(6), 1311–1321. <https://doi.org/10.1002/eat.24085>
- De Lusignan, S., Mold, F., Sheikh, A., Majeed, A., Wyatt, J. C., Quinn, T., Cavill, M., Gronlund, T. A., Franco, C., Chauhan, U., Blakey, H., Kataria, N., Barker, F., Ellis, B., Koczan, P., Arvanitis, T. N., McCarthy, M., Jones, S., & Rafi, I. (2014). Patients' online access to their electronic health records and linked online services: A systematic interpretative review. *BMJ Open*, 4(9), Article e006021. <https://doi.org/10.1136/bmjopen-2014-006021>
- Deshpande, A., Murahari, V., Rajpurohit, T., Kalyan, A., & Narasimhan, K. (2023). *Toxicity in chatgpt: Analyzing persona-assigned language models*. arXiv. <https://doi.org/10.48550/arXiv.2304.05335>
- DeVon, C. (2023, November 30). *On ChatGPT's one-year anniversary, it has more than 1.7 billion users—here's what it may do next*. CNBC. <https://www.cnbc.com/2023/11/30/chatgpts-one-year-anniversary-how-the-viral-ai-chatbot-has-changed.html>
- Dhakal, V., Feit, A. M., Kristensson, P. O., & Oulasvirta, A. (2018, April). Observations on typing from 136 million keystrokes. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–12). ACM. <https://doi.org/10.1145/3173574.3174220>
- Drysdale, K., Wells, N., Smith, A. K., Gunatillaka, N., Sturgiss, E. A., & Wark, T. (2023). Beyond the challenge to research integrity: Imposter participation in incentivised qualitative research and its impact on community engagement. *Health Sociology Review*, 32(3), 372–380. <https://doi.org/10.1080/14461242.2023.2261433>
- Editorials, N. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613(7945), 612. <https://www.nature.com/articles/d41586-023-00191-1>
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19(1), 17. <https://doi.org/10.1007/s40979-023-00140-5>
- Eypasch, E., Williams, J. I., Wood-Dauphinee, S., Ure, B. M., Schmullig, C., Neugebauer, E., & Troidl, H. (1995). Gastrointestinal quality of life Index: Development, validation and application of a new instrument. *Journal of British Surgery*, 82(2), 216–222. <https://doi.org/10.1002/bjs.1800820229>
- Fairburn, C. G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press.
- Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. D., Köller, O., & Möller, J. (2024). Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. *Computers and Education: Artificial Intelligence*, 6(5), Article 100209. <https://doi.org/10.1016/j.caeai.2024.100209>
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine*, 6(1), 75. <https://doi.org/10.1038/s41746-023-00819-6>
- Gelman, S., Gururangan, S., Sap, M., Choi, Y., & Smith, N. A. (2020). *Realtotoxicityprompts: Evaluating neural toxic degeneration in language models*. arXiv. <https://doi.org/10.48550/arXiv.2009.11462>
- Gibson, A. F., & Beattie, A. (2024). More or less than human? Evaluating the role of AI-as-participant in online qualitative research. *Qualitative Research in Psychology*, 21(2), 175–199. <https://doi.org/10.1080/14780887.2024.2311427>, <https://www.tandfonline.com/doi/pdf/10.1080/14780887.2024.2311427>
- Gluck, M. E., & Geliebter, A. (2002). Racial/ethnic differences in body image and eating behaviors. *Eating Behaviors*, 3(2), 143–151. [https://doi.org/10.1016/S1471-0153\(01\)00052-6](https://doi.org/10.1016/S1471-0153(01)00052-6)
- Guest, G., Namey, E., O'Regan, A., Godwin, C., & Taylor, J. (2023). Comparing interview and focus group data collected in person and online. <https://doi.org/10.25302/05.2020.ME.1403117064>
- Himmerich, H., Kan, C., Au, K., & Treasure, J. (2021). Pharmacological treatment of eating disorders, comorbid mental health problems, malnutrition and physical health consequences. *Pharmacology & Therapeutics*, 217(2), Article 107667. <https://doi.org/10.1016/j.pharmthera.2020.107667>
- Kabir, S., Udo-Imeh, D. N., Kou, B., & Zhang, T. (2023). *Who answers it better? An in-depth analysis of ChatGPT and stack overflow answers to software engineering questions*. arXiv. <https://doi.org/10.48550/arXiv.2308.02312>
- Kasneeci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneeci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning*

- and Individual Differences, 103(102274), Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kozar, J. M., & Damhorst, M. L. (2009). Comparison of the ideal and real body as women age: Relationships to age identity, body satisfaction and importance, and attention to models in advertising. *Clothing and Textiles Research Journal*, 27(3), 197–210. <https://doi.org/10.1177/0887302X08326351>
- Ladha, N., Yadav, K., & Rathore, P. (2023). AI-generated content detectors: Boon or bane for scientific writing. *Indian Journal of Science and Technology*, 16(39), 3435–3439. <https://doi.org/10.17485/IJST/v16i39.1632>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management in Education*, 21(2), Article 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Lin, A., & Micic, D. (2021). Nutrition considerations in inflammatory bowel disease. *Nutrition in Clinical Practice*, 36(2), 298–311. <https://doi.org/10.1002/ncp.10628>
- Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In *The sage handbook of qualitative research* (4th ed., pp. 97–128). Sage Publications.
- Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on Medical Education*, 12(1), 261–270. <https://doi.org/10.5334/pme.1072>
- McLaren, L., & Kuh, D. (2004). Women's body dissatisfaction, social class, and social mobility. *Social Science & Medicine*, 58(9), 1575–1584. [https://doi.org/10.1016/S0277-9536\(03\)00209-0](https://doi.org/10.1016/S0277-9536(03)00209-0)
- Morris, M. E., Kuehn, K. S., Brown, J., Nurius, P. S., Zhang, H., Sefidgar, Y. S., Xu, X., Riskin, E. A., Dey, A. K., Consolvo, S., & Mankoff, J. C. (2021). College from home during COVID-19: A mixed-methods study of heterogeneous experiences. *PLoS One*, 16(6), Article e0251580. <https://doi.org/10.1371/journal.pone.0251580>
- Pellicano, E., Adams, D., Crane, L., Hollingue, C., Allen, C., Almendinger, K., Botha, M., Haar, T., Kapp, S. K., & Wheeley, E. (2024). Letter to the editor: A possible threat to data integrity for online qualitative autism research. *Autism*, 28(3), 786–792. <https://doi.org/10.1177/13623613231174543>
- Petty, N. J., Thomson, O. P., & Stew, G. (2012). Ready for a paradigm shift? Part 1: Introducing the philosophy of qualitative research. *Manual Therapy*, 17(4), 267–274. <https://doi.org/10.1016/j.math.2012.03.006>
- Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology*, 52(2), 126–136. <https://doi.org/10.1037/0022-0167.52.2.126>
- Pullen Sansfaçon, A., Gravel, E., & Gelly, M. A. (2024). Dealing with scam in online qualitative research: Strategies and ethical considerations. *International Journal of Qualitative Methods*, 23, 1–11. <https://doi.org/10.1177/16094069231224610>
- Rabiee, F. (2004). Focus-group interview and data analysis. *Proceedings of the Nutrition Society*, 63(4), 655–660. <https://doi.org/10.1079/PNS2004399>
- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 5783. <https://doi.org/10.3390/app13095783>
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Richburg, A., & Stewart, A. J. (2024). Body image among sexual and gender minorities: An intersectional analysis. *Journal of Homosexuality*, 71(2), 319–343. <https://doi.org/10.1080/00918369.2022.2114399>
- Ridge, D., Bullock, L., Causer, H., Fisher, T., Hider, S., Kingstone, T., Gray, L., Riley, R., Smyth, N., Silverwood, V., Spiers, J., & Southam, J. (2023). Imposter participants' in online qualitative research, a new and increasing threat to data integrity? *Health Expectations: An International Journal of Public Participation in Health Care and Health Policy*, 26(3), 941–944. <https://doi.org/10.1111/hex.13724>
- Sala, M., Keshishian, A., Song, S., Moskowitz, R., Bulik, C. M., Roos, C. R., & Levinson, C. A. (2023). Predictors of relapse in eating disorders: A meta-analysis. *Journal of Psychiatric Research*, 158(3), 281–299. <https://doi.org/10.1016/j.jpsychires.2023.01.002>
- Shahriar, S., & Hayawi, K. (2023). *Let's have a chat! A conversation with ChatGPT: Technology, applications, and limitations*. arXiv. <https://doi.org/10.48550/arXiv.2302.13817>
- Sharma, P., McPhail, S. M., Kularatna, S., Senanayake, S., & Abell, B. (2024). Navigating the challenges of imposter participants in online qualitative research: Lessons learned from a paediatric health services study. *BMC Health Services Research*, 24(1), 724. <https://doi.org/10.1186/s12913-024-11166-x>
- Sim, J., & Waterfield, J. (2019). Focus group methodology: Some ethical challenges. *Quality and Quantity*, 53(6), 3003–3022. <https://doi.org/10.1007/s1135-019-00914-5>
- Stewart, D. W., & Shandasani, P. (2017). Online focus groups. *Journal of Advertising*, 46(1), 48–60. <https://doi.org/10.1080/00913367.2016.1252288>
- Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, 29(8), 1930–1940. <https://doi.org/10.1038/s41591-023-02448-8>
- Tong, A. (2023, September 7). *Exclusive: ChatGPT traffic slips again for third month in a row*. Reuters. <https://www.reuters.com/technology/chatgpt-traffic-slips-again-third-month-row-2023-09-07/>
- Trott, S. (2024, May 2). *Tokenization in large language models, explained*. Over the Counter. <https://seantrott.substack.com/p/tokenization-in-large-language-models>
- Tse, A., Xavier, S., Trollope-Kumar, K., Agarwal, G., & Lokker, C. (2022). Challenges in eating disorder diagnosis and management among family physicians and trainees: A qualitative study. *Journal of Eating Disorders*, 10(1), 45. <https://doi.org/10.1186/s40337-022-00570-5>

- Walther, J. B., & Boyd, S. (2002). Attraction to computer-mediated social support. *Communication Technology and Society: Audience Adoption and Uses*, 153188(2).
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1), 26. <https://doi.org/10.1007/s40979-023-00146-z>
- Westwood, H., Kerr-Gaffney, J., Stahl, D., & Tchanturia, K. (2017). Alexithymia in eating disorders: Systematic review and meta-analyses of studies using the Toronto Alexithymia Scale. *Journal of Psychosomatic Research*, 99, 66–81. <https://doi.org/10.1016/j.jpsychores.2017.06.007>
- Williams, S., Clausen, M. G., Robertson, A., Peacock, S., & McPherson, K. (2012). Methodological reflections on the use of asynchronous online focus groups in health research. *International Journal of Qualitative Methods*, 11(4), 368–383. <https://doi.org/10.1177/160940691201100405>
- Winkler, W. E. (1990). *String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage*. U.S. Bureau of the Census.
- Wu, S., Fei, H., Qu, L., Ji, W., & Chua, T. S. (2023). *Next-gpt: Any-to-any multimodal ll*. arXiv. <https://doi.org/10.48550/arXiv.2309.05519>
- Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., & Hu, X. (2023). Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data*, 18(6), 1–32. <https://doi.org/10.1145/3649506>
- Yarmand, M., Solyst, J., Klemmer, S., & Weibel, N. (2021, May). “It feels like I am talking into a void”: Understanding interaction gaps in synchronous online classrooms. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–9). ACM. <https://doi.org/10.1145/3411764.3445240>
- Yu, Z., Liu, X., Liang, S., Cameron, Z., Xiao, C., & Zhang, N. (2024). *Don't listen to me: Understanding and exploring jailbreak prompts of large language models*. arXiv.