

# Voucher Now Please: How to Navigate AI Bots and Fraudulent and Mischievous Responders in Online Research

**Authors:** Black, Suzanne R., Connelly, Louise, Osborne, Nicola, and Terras, Melissa.

**Date:** September 2024

**Contact:** [designinformatics@ed.ac.uk](mailto:designinformatics@ed.ac.uk)

**License:** Creative Commons Attribution 4.0 International (CC-BY) 

**Citation:** Black, Suzanne R., Connelly, Louise, Osborne, Nicola, and Terras, Melissa. (2024). 'Voucher Now Please: How to Navigate AI Bots and Fraudulent and Mischievous Responders in Online Research'. *Zenodo*. <https://doi.org/10.5281/zenodo.13356359>

## INTRODUCTION

Using the internet to recruit participants and gather data for research purposes has become more and more prevalent, accelerated by both the restrictions on travel and meeting prompted by the Covid-19 pandemic and the ease of accessing large amounts of cheap data. While problems of fraudulent responses and low-quality data are not new, the delivery mode of the internet and recent developments in generative artificial intelligence (AI) have enabled new challenges to flourish. In this paper, we focus on the specific challenges of automated responses to surveys, interviews and online workshops by 'bots' (software robots), which often rely on artificial intelligence tools. The consequences of gathering bot-generated data – whether knowingly or not – are significant, ranging from drawing conclusions from compromised data to wasting researcher time and funds. The forms of bot engagement, and the detectability of the presence of automated responses or participants is becoming increasingly complex as AI becomes more sophisticated and ubiquitous. We offer recommendations for researchers who are designing their research projects about how to mitigate the challenges posed by AI to the integrity of their research.

Research conducted on or with participants recruited via the internet has many benefits. These include lower costs compared to in-person data collection (Storozuk et al., 2020)

and greater participation due to geography and anonymity (Jackson et al., 2023). Conducting research online can also enable access to diverse and hard-to reach populations (Bybee et al., 2022) and those from “minority ethnic backgrounds who are often under-represented in research” (Woolfall, 2023, p. 421). However, the infiltration of bot-generated data into online returns is becoming a pressing issue, which relates to the known issue of mischievous or fraudulent responders contributing to data gathering studies (Przybylski, 2016). Bots can be used in some practical applications (e.g. customer care) or as a benign efficiency measure (Ferrara et al., 2016), but there is growing evidence of their use as “malicious software applications programmed to complete automated tasks online” (Storozuk et al., 2020, p. 472) with random (Xu et al., 2022) or AI-generated responses. When bots are deployed, data may be corrupted, potentially invalidating the research. In particular, the use of online research to access minority populations becomes a weakness making it difficult to understand issues faced by minority groups (Cimpian & Timmer, 2020). The existing literature by researchers who have encountered bots judges that “[t]he new, disturbing, reality is that bot-activity in survey data is inevitable” (Storozuk et al., 2020, p. 478).

In this paper, we give an overview of the current advice around mitigating these challenges. While best practice for avoiding bot content in online surveys exists, there are fewer extant guidelines for avoiding this issue related to interviews, focus groups and workshops held online (which speaks to the rapid, recent development in moving image and sound synthesis in generative AI, Ramponi, 2023). Additionally, the tactics of bot makers escalate in proportion to the tactics used to mitigate them. Therefore, this is a rapidly changing field in which researchers need to be aware of the latest challenges and how to tackle them. We also describe the authors’ difficulties with bots being used during data gathering activities and combine this research and experience to offer a list of practical solutions to mitigate the problem. This case study uses a participatory action research methodology (Cornish et al., 2023) to engage with a research-related issue and generate practical solutions and guidelines.

Furthermore, we have been careful about the language we use around this problem. Acosta’s analysis of how digitally savvy east Asian media fans organise online shows that the term ‘bot’ can be used to signal otherness in ways that echo racist and orientalist discourses (2023). Pellicano et al. (2023), who undertake autism research, point out the potential harm in using the term ‘imposters’ when many autistic people

often feel like imposters themselves, and they instead use the term ‘scammer’ for individuals who attempt to partake in research – and hence gain participation incentives – despite not meeting the eligibility criteria. In this paper we take care to distinguish bots, which are deployed by human agents and have no agency of their own, and which cannot be eligible research participants, from scammer individuals providing fraudulent, inaccurate or mischievous data.

## THE PROBLEM OF BOTS

The problem of unwanted data, from a combination of bots, generative AI or malicious or mischievous responders, in online research is part of a broader discussion on disinformation. Bots and AI-generated content are made for a variety of reasons, from the seemingly harmless, although still environmentally harmful (Kollanyi, 2016), to the outright malicious (Dunham & Melnick, 2008). Much of the literature on online misinformation (Geers et al., 2024) and disinformation (Simion, 2023) identifies strategies to disrupt western society and western discourse with the aim of changing public and political opinion (Chan, 2024) and destabilising trust in institutions, democracy and policy (Bennett & Livingston, 2023; Colomina, 2022). These strategies have been exacerbated by the use of AI-generated content (Karinshak & Jin, 2023; Shoaib et al., 2023) with the result that “[t]he internet is filling up with ‘zombie content’ designed to game algorithms and scam humans” (Purtill, 2024); the incursion of vast amounts of AI-generated data is a very real threat to the western information environment.

It is difficult to estimate the scale of the problem of fraudulent and scammer responses to online data gathering instruments as there is no way to tell how many fraudulent responses have gone undetected. The situation may be dire, though. It is estimated that the introduction of as little as 5% invalid data can significantly alter study results (Storozuk et al., 2020, p. 478). In surveys, the presence of bots invalidates the data unless the bot responses can be identified and removed as “[w]ithout removing all of the bot-generated responses, the data set cannot be used to gain insight into the research question at hand” (Simone, 2019). Identifying bot responses is not a trivial activity as “[p]rogrammers have developed bots that will create a normal distribution across all responses...or if there are open-ended questions, they’ll extract language from the survey itself to compose more logical responses” (Simone, 2019). There are potential financial consequences (Jackson et al., 2023) if automatic compensation is in place as

well as the financial impact of significant loss of researcher time due to additional steps in validating the data. Furthermore, the introduction of illegitimate data has the potential to pollute the scholarly and humanistic record, which will then affect further ongoing studies as well as AI models trained upon these outputs.

In online data gathering situations, such as interviews, focus groups and workshops held over videoconferencing software, the use of bots by participants to simulate audio and chat participation can derail the session, cause confusion about who is a legitimate participant and who is not, and result in unusable data.<sup>1</sup> These issues arising from generative AI's growing capabilities in online synchronous spaces are distinct from, though related to, existing challenges arising from mischievous or fraudulent responders when recruiting for interviews. Pellicano et al. (2023) report the move from scammers infiltrating surveys to more recently trying to access interviews and focus groups by posing as autistic individuals or the parents of autistic children. The researchers discuss the problem of trying to distinguish between scammers and eligible participants while maintaining the trust of an under-served research population. Research participants are often screened by means of a survey and introducing inaccurate data at this stage affects the entire study; Woolfall (2023) found that 80% of those who registered interest for one study were fraudulent. The inclusion of such data can invalidate entire studies and, beyond that, the scholarly and humanistic record, and this is compounded when further information environments propagate the results of fraudulent studies or use them to train AI models themselves.

## IDENTIFYING BOTS

Researchers who have encountered bots and mischievous or fraudulent responders have implemented a range of strategies to identify and remove inaccurate data as well as preventative measures.

## SURVEYS

The following characteristics may indicate that bots have been used to populate survey data: a large number of responses at the same time (Storozuk et al., 2020; Woolfall, 2023), surveys completed more quickly than a human could (Storozuk et al., 2020),

---

<sup>1</sup> At the time of writing there is a known issue with note-taking bots accessing Zoom meetings and taking control of recordings. (See <https://community.zoom.com/t5/Zoom-Meetings/Help-AI-Notetaking-Bots-Overtaking-Recording/td-p/196448>)

nonsensical answers (Storozuk et al., 2020), respondents' answers indicating that they do not meet eligibility criteria (Storozuk et al., 2020), a large number of participants claiming to be from under-represented groups (Cimpian & Timmer, 2020; Pellicano et al., 2023), contradictory or vague answers (Pellicano et al., 2023; Woolfall, 2023), and frequent requests for compensation (Pellicano et al., 2023).

## INTERVIEWS

The following characteristics may indicate that fraudulent or mischievous responders have accessed online interviews, focus groups or workshops: difficulty scheduling interviews (Pellicano et al., 2023), reluctance to turn cameras on (Pellicano et al., 2023; Woolfall, 2023), poor internet connections (Pellicano et al., 2023), muffled audio (Woolfall, 2023) and vague responses (Woolfall, 2023).

## WORKSHOPS

In the authors' experience of running an online Zoom workshop on 8 August 2023 that was accessed by fraudulent responders who used bots to simulate plausible participation, we found that direct questions, such as asking about a participant's job, were answered in the chat with answers that contradicted sign-up information, participants did not show themselves on camera, audio participation during breakout groups was minimal, there was often a delay before answering questions, audio participation often only extended to claiming a bad internet connection, and there were immediate demands for compensation, both in the Zoom chat and by email after the workshop. This experience has similarities with the challenges experienced by other researchers when conducting synchronous online data gathering. However, in this case, we think the fraudulent responders escalated their tactics by using bots to generate the text and audio participation by using the audio input capabilities of ChatGPT.<sup>2</sup>

There are commonalities across all of the above online research modes and by both fraudulent or mischievous responders using bots or not. There was often a demand for payment, which fits with the presumed motivation of receiving compensation. Woolfall (2023) found that "following the interview, fraudulent participants may get in touch to ask about payment for their time. They will sometimes be persistent and send multiple emails" while "[i]t is very rare that a genuine participant will email to ask about monetary

---

<sup>2</sup> <https://openai.com/index/chatgpt-can-now-see-hear-and-speak/>

payment even when a voucher to compensate people for their time has been mentioned in the social media advert” (p. 421).

## CASE STUDY

### EXPERIENCE OF BOTS IN CREATIVE AI SURVEY AND ONLINE WORKSHOP

Three of the authors have firsthand experience with this problem. We conducted the [Creative AI Demonstrator project](#) (part of [Creative Informatics](#), the AHRC creative cluster based in Edinburgh and its regions) leading to a report on the use of AI in the creative industries in Scotland (Black et al., 2024). As part of this project we held an online Zoom workshop (8 August 2023) and a survey (19 June to 30 September 2023) that both attracted fraudulent responders using different methods.

### SURVEY

We identified a range of fraudulent responses in answers provided to our online survey, hosted by Jisc Online Surveys. These included answers that were merely strings of gibberish text (for example, one answer to “Please describe your Creative AI / Machine Learning work to date” was “4r3r34t34t”) as well as answers that seemed plausible at first but, when read carefully, did not answer the question or were composed of the terms used in the question (for example, in response to the same question, “Artificial intelligence robots utilize a variety of sensors to perceive the surrounding environment”). There was also a lot of repetition across the answers of supposedly different participants. We surmised that some of these answers had been provided by automated survey-taking software, possibly using ChatGPT as it has become a popular tool for generating text. One answer even identified the writer as an AI tool: “My design and development was done by a team of professional AI engineers and data scientists who used vast amounts of data and algorithms to train and optimize my models to improve the quality of my answers and the accuracy of my responses”.

Once the presence of fraudulent data was identified, the authors implemented measures to separate it from the other valuable data. This involved carefully reading a selection of the free text questions for each response to identify any responses where the majority of the answers were not within a range of plausible answers. This strategy was only possible because of the number of responses (less than 100), although it was

still time consuming. We identified and removed 65% of the responses as being bot-generated or otherwise not provided in good faith.

The reason for these fraudulent responses is unclear given that there was no direct financial reward beyond entry into a prize draw. One possibility is to train bots for surveys offering larger financial rewards (Simone, 2019). Also, the topic of the survey was creative uses of artificial intelligence, which may have attracted mischievous or playful responses.

## WORKSHOP

Our online workshop, held on 8 August 2023<sup>3</sup> to ascertain the response to AI by those in or adjacent to the creative industries, attracted a large number of sign-ups, with all 60 spaces quickly being booked. 35 of those who signed up attended on the day. However, it became apparent during the workshop that the majority of those who attended did so to claim the participation voucher we offered, rather than being genuinely interested in sharing their perspectives. This was particularly in evidence when the workshop was split into breakout groups. In this case, there were so many participants contributing very little or very low-quality responses, some of which we judged to be false (such as answers to eligibility criteria), that we had to discount all of the data from the workshop. As described by other researchers, these participants were focused on securing a voucher for participation, with one participant asking in the online chat “How about the voucher?” before the workshop had ended.<sup>4</sup>

The fact that we felt compelled to discard the responses from the online workshop was particularly unfortunate from an equality, diversity and inclusion perspective as the aim of undertaking a workshop online was to ensure participation in the research was accessible to those who are often less able to participate in in-person events. In our experience, this includes people living with physical disabilities or mental ill health, people with sensory sensitivities and/or anxiety that may make in-person group participation challenging, those with caring responsibilities, and those living in more

---

<sup>3</sup> See <https://web.archive.org/web/20240913114204/https://www.eventbrite.co.uk/e/exploring-the-potential-for-creative-ai-online-tickets-669202922957>

<sup>4</sup> The title of this paper, ‘Voucher Now Please’, is indicative of the type of demand we received via the workshop chat function and by email after the event.



rural and/or remote locations (an important consideration when engaging with Scotland as a whole).

## **STRATEGIES TO PREVENT/REMOVE BOTS**

Researchers who have encountered bots and fraudulent or mischievous responders have reported the strategies they used and suggested others that may be used to mitigate the problem. These fall into various categories: participant recruitment, participant verification, study design, compensation, and data validation. Some of these apply to surveys, some to interviews, focus groups and workshops, and some to all.

### **PARTICIPANT RECRUITMENT**

Although social media seems like a good way to reach populations, openly advertising studies with compensation attracts a high proportion of fraudulent participants. Instead, accessing closed social media groups with members who are in the target population, mailing groups or other trusted sources is more likely to attract legitimate participants (Jackson et al., 2023; Storozuk et al., 2020; Woolfall, 2023). Likewise, sharing survey links publicly makes it easy for those using bots to access the survey (Storozuk et al., 2020).

### **PARTICIPANT VERIFICATION**

It is suggested that steps are taken to eliminate as many fraudulent responders from contributing data in the first place. Tactics include screening email addresses to look for patterns such as numbers instead of a name and IP addresses to check participant location (Storozuk et al., 2020), requiring a phone number for verification (Loebenberg et al., 2023), and screening participants over a phone or video call with the camera on (Woolfall, 2023). For surveys, it is possible to track traffic to a survey URL by using tracking codes or link shorteners. It should be noted that there are multiple difficulties with these tactics: screening email addresses is not consistent, collecting IP addresses can impinge upon data privacy (and goes against the research and GDPR best practice of 'data minimisation') and may be prejudicial against those using VPNs, screening participants over a phone or video call requires a lot of labour from the research team and may disincentivise those who have communication needs around neurodivergence or require anonymity.



## STUDY DESIGN

There are various ways to incorporate checks for participant validity and data quality into surveys and interviews. In surveys, these include the use of CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) (Loebenberg et al., 2023; Storozuk et al., 2020), presenting text as an image (Storozuk et al., 2020), attention check questions to identify whether the respondent is human (Loebenberg et al., 2023; Storozuk et al., 2020), participant eligibility questions about the research topic (Jackson et al., 2023), qualitative questions in an otherwise quantitative survey (Griffin et al., 2022), 'honeypot' questions that are only visible to non-human survey completion software (Storozuk et al., 2020), and asking for the same information in multiple ways (Griffin et al., 2022) to check for consistency. Tools and platforms will have specific affordances that may be helpful, for example in online workshops and focus groups, waiting rooms can be used and participants can be removed by the host.

It should be noted that some of these may introduce accessibility issues, such as CAPTCHAs and text as images. Additionally, bot software is continually becoming more sophisticated, and some can now navigate tactics that hindered them previously, such as CAPTCHAs and honeypot questions (Storozuk et al., 2020). Text generation technologies such as ChatGPT are quickly becoming able to plausibly answer qualitative questions and, while they are still usually detectable by human readers, checking all answers is time consuming when dealing with large amounts of data.

Any mischievous or playful responses may still have value, for example, if the target research population is likely to play with or subvert research objectives, this may be valid despite being inconvenient.

## COMPENSATION

The issue of compensation for participants' time in research studies is a difficult one as monetary reward is likely the motivating factor for many fraudulent responders. Not automatically giving compensation to survey respondents until they have been judged to be human participants is a sensible precaution (Storozuk et al., 2020). It has been suggested that not advertising financial compensation on social media (Loebenberg et al., 2023), not mentioning compensation in advertising material about the study at all until the participant is taking part (Woolfall, 2023), or entering participants into a raffle

(Griffin et al., 2022) might help with this challenge. However, responsible researchers will want to compensate participants for their participation and ethics boards may not give approval for studies that are not clear about compensation in advance (Storozuk et al., 2020). While compensation is rarely more than a modest sum, its provision can be an important way to include a wider diversity of participants. Remuneration can be helpful when reaching out to those on lower or irregular incomes who may lose out on work, for example freelancers, or to offset costs associated with childcare or internet access. It can also help when reaching out to populations who are less familiar with, interested in, or represented by research.

## **DATA VALIDATION**

Once it has been determined that there is bot activity in a survey, it is recommended that the survey is closed as quickly as possible as giving bots access to a survey allows them to learn how to plausibly complete it (Storozuk et al., 2020) and increases the invalid data that must be identified and removed. This can be a time-consuming process and therefore it is important to prevent bots and fraudulent responders in advance as much as possible. Tactics for identifying bot activity include checking for inhumanly fast survey completion times (Storozuk et al., 2020), running statistical analyses on the data to check for expected distributions (Irish & Saba, 2023; Storozuk et al., 2020) and using machine learning methods to identify mischievous responders and predict eligibility status (Cimpian & Timmer, 2020). None of these methods are easy or guaranteed to be entirely accurate. The increasing sophistication and continual performance improvement of bot programs means that methods to detect them must also evolve. For example, it is difficult for automated programs to detect AI-generated text: OpenAI, the company who developed ChatGPT, created a classifier to distinguish between text written by humans and AIs, but removed the product in July 2023 as it was not performing well enough (Kirchner et al., 2023).

Detecting fraudulent participants during synchronous research modes like interviews, focus groups and workshops is likely to be done by researchers during the research activity. There are still consequences for data validation, especially where it is difficult to distinguish between legitimate and fraudulent data after the fact, such as a difficulty in reconciling transcripts with contributions made in the chat (Jackson et al., 2023).

## **FURTHER ADVICE**

The overwhelming advice from researchers who have encountered these problems is to use multiple strategies and to be transparent about the challenges. Woolfall (2023) argues that “it is important for research teams to set clear guidance and regularly discuss how they will prevent, identify and act on fraudulent research participation” (p. 422) as it benefits the research community to share information about how bots and the tactics of fraudulent responders are evolving, and which strategies work best to defeat them.

The presence of bots or otherwise fraudulent responses to online research is an ongoing problem. It not only has the potential to compromise the data collected in individual studies but also to undermine relations between researchers and current or future pools of participants. As Goodrich et al. (2023) caution, “we should remain mindful of the actual humans whom we hope to gather information from” (p. 779). As we have stressed, no one tactic can ward off or root out fraudulent or mischievous response data, and many tactics require care when they are implemented to reduce the potential for “methods or barriers to detect or reduce fraudulent responses” coming “at the expense of valid respondents” (Goodrich et al., 2023, p. 779). Judging whether a response has been submitted by a human or a bot brings with it a value system about what it means to be human; we must be careful not to define the parameters too narrowly and thereby implement measures that are prejudiced against groups like non-native English speakers, those who are neurodivergent, those without access to robust internet connections, or even those with impressive touch-typing speeds.

## **ETHICAL CONSIDERATIONS**

### **FOR RESEARCHERS**

Researchers need to ensure online ethical guidelines and practices are adhered to (see AoIR guidelines: franzke et al., 2020) and in addition, specific ethical considerations for using AI or potential impact from AI, should also be considered. Ultimately, the researcher is responsible for being knowledgeable and informing Research Ethics Committees (RECs) about how AI is being used or how it may impact or pose a risk, in order that the REC can adequately review the application. The challenge is that existing institutional and research frameworks, processes, or guidelines may not be equipped for this type of discussion and, therefore, the researcher must complete an ethics application to the best of their ability.

We are beginning to see statements appear in research outputs to indicate when AI is used deliberately in research, for example, the use of ChatGTP,<sup>5</sup> but it is also prudent to include a ‘declaration of confidence’ stating the likelihood that use of AI has influenced datasets and outputs. This will ensure that readers are aware of any potential impact on their own research if they rely on those outputs. It may also be helpful if the declaration includes “broader societal impact statements [that] can ensure researchers reflect on, and document, the full list of potential harms, risks and benefits their work may pose” (Ada Lovelace Institute, 2022, p. 72).

## THE ROLE OF RESEARCH ETHICS COMMITTEES

This article has highlighted a range of considerations for the researcher, but we must also stress the role of Research Ethics Committees (RECs) in this space. What aspects of bot use must researchers consider when applying for ethics approval, and what do RECs need to consider in relation to this fast-changing bot-ridden environment, where existing ethical processes or guidelines may not be suitable?

RECs are responsible for the governance of research (see UKRI, 2023), as they ensure research considers the minimisation of harm or risk, participants’ dignity, rights, well-being and safety; as well as ensuring the research aligns with ethical standards, legislation, best practice, and is of value or benefit to others or society. However, as this article has already highlighted, with the introduction of AI, this can be a challenging undertaking for RECs for various reasons (Knight et al., 2024).

Overall, research that intentionally has an AI component, or which may be impacted or disrupted by AI, requires researchers and RECs to consider several factors that may differ to more mainstream research processes, practices, or guidelines, if participants or data are to be protected and to ensure robust, ethical and reliable research is undertaken (Connelly et al., 2024; Knight et al., 2024). Recommendations for researchers and RECs who are undertaking AI research are beginning to emerge, but this needs continual horizon scanning to ensure it aligns with the ever-changing online environment (Ada Lovelace Institute, 2022). Connelly et al. (2024) provide further discussion of how RECs can approach the issue of AI in ethics applications.

---

<sup>5</sup> For example, Carvalho, de Moraes & Souza (2023) include a “declaration of generative AI and AI-assisted technologies in the writing process” (p.10).

## **SYNTHETIC DATA**

It should be noted that researchers are already experimenting with creating synthetic datasets generated using Large Language Models (LLMs) and designed to be statistically representative of real-world populations. However, the logistical and ethical barriers are immense, including the fact that “these research methods...conflict with central values of research involving human participants: representing, including and understanding those being studied” (Stokel-Walker, 2024). Moreover, the way that LLMs work – by aggregating the training data fed to them and offering up the most likely response to a given prompt – reflects and exacerbates the biases already present within society. Social research values “the idiosyncratic and irreverent aspects of human participation”, “the unexpected directions participants go in” and “the messy, emotional, lived experience of people’s perspectives”, all things that are outside of the scope of AI-generated data (Beattie & Gibson, 2024).

## **QUESTIONS TO CONSIDER WHEN DESIGNING ONLINE RESEARCH**

The landscape of AI is changing rapidly and there are no foolproof rules for avoiding the attention of fraudulent or mischievous responses to surveys, interviews and focus groups. We end by suggesting a list of questions to enable researchers to identify and minimise potential harms to data gathering activities. These questions are designed to stimulate reflection on the research design and envisage unwanted responses rather than as an exhaustive checklist.

### **GENERAL RESEARCH DESIGN:**

- Which population is this research designed to target?
- What are the most effective ways of reaching this population? Are there in-person locations, groupings, trusted partner organisations or online ‘walled gardens’ (e.g. email lists) where this population is communicating, which would be more effective than seeking participants on the open internet?
- What are the consequences of receiving participation from outside of that specific population?

- Are there any particular risks of exposing your target research participants to potential bot participants, especially if you are working with a more vulnerable or younger age group?
- Are there any particular risks you can envision in differentiating between human and potential bot participants? For instance, are you using platforms known to be particularly populated by bots (e.g. X/Twitter)? Are you working with a community that may give less expected or less conventionally articulated responses (e.g. non-native English speakers, individuals with neurodivergent characteristics)?
- How will you motivate or compensate participation in your research? Can you devise ways to communicate this as part of your recruitment process so that your work remains inclusive but does not attract malicious actors and/or bots?

### **FOR SURVEYS:**

- What are the consequences of receiving responses that are either fraudulent or mischievous in nature?
- Are the topic of your research and the platforms you plan to use likely to attract fraudulent or mischievous responses? It may be useful to undertake a literature review and talk to peers about previous issues they have encountered.
- What technology tools do you plan to use to undertake your survey, and what is their current position and/or track record on bot detection, spam prevention, etc?
- How will you craft questions to alert you to the fact that it may not be a human participant who has responded?
- How will you identify whether data has been submitted either fraudulently or mischievously?
- What plans do you have in place to identify and remove such responses? How will you report this in your reporting and publication of the work so that the criteria for exclusion are clear?

### **FOR ONLINE WORKSHOPS, FOCUS GROUPS AND INTERVIEWS:**

- What are the consequences of participation that is fraudulent or mischievous in nature?

- Are there any risks to your participants of potential bot participation? This may include exposure of personal information, information security risks, psychological safety, etc. These risks should be considered in all such research methods whether or not they take place online, but automation presents a different sense and potential scalability of exposure.
- How can you limit the participation of fraudulent or mischievous participants at the outset of the activity?
- How will you identify whether participants are participating in good faith and providing good faith responses, remembering that 'good faith' responses may still be hostile? Providing a code of conduct making clear what behaviours will and will not be tolerated can help ensure that there is a shared sense of safety and expectation.
- What plans do you have in place to identify and mitigate fraudulent participants during a live online data gathering event?
  - Do you have an adequate number of researchers or assistants supporting the event to enact these plans? How will you communicate potential identification of bot or fraudulent actors, e.g. through agreed 'back channels'?
  - How would you communicate any identification of bots or fraudulent or mischievous participants to your other legitimate participants to maintain trust and transparency?
  - Are you familiar with your organisational and/or legal obligations to report any malicious behaviour? For example if you believe someone's personally identifiable information has been accessed by a bad actor, do you know your relevant Data Protection Officer and the data breach reporting mechanisms you need to undertake?
- Are you able to swiftly remove any bad actors if required, and how will you handle any possible backlash to this action?



- Can you make a shift in platform or meeting ID/invitation if you need to relocate the session at the last minute, or during the session? How would you communicate this to participants?
- Do you know how to swiftly and effectively stop the live event, if needed? How would you communicate early termination of a session and ensure the safety and wellbeing of your participants?

### **DECLARING AI USE/INTERFERENCE:**

- Would it be useful to make a declaration of potential or actual AI use/interference in your ethics submission or subsequent publications?
- Will this information help readers of your work and those who may want to use your outputs to do so more accurately and with confidence?
- How will you share any negative or compromising experiences back to your peers, locally within your institution or to others in the field undertaking similar work?

### **DURING THE ETHICS APPROVAL PROCESS:**

- Research Ethics Committees (RECs) may not have the experience or expertise to understand the complexity, nuances, or potential risks of your research. Have you taken the opportunity to adequately inform RECs as part of your ethics submission of the related risks, societal impacts, and disruption of data collection, etc? Have you continued to reflect on these throughout the lifecycle of the research (design, data collection and analysis, and publication/presentation of findings)?
- In preparing a submission for the REC, would it be beneficial to seek additional peer review from a colleague or known researcher experienced in internet research and/or currently conducting research or research recruitment through online platforms?

## CONCLUSION

Using the internet to recruit participants and gather data for research purposes is a valuable method that has, recently, become more difficult due to the use of responses submitted either mischievously or fraudulently using bot- or AI-generated data. Such unwanted responses not only waste researcher time and other resources but also jeopardise the integrity of the data collected and any related findings and, more perilously, contribute to an information environment that is flooded with low-quality data and thereby impoverished.

In this paper, we describe the authors' experiences in dealing with fraudulent responses in their own data collection, collate and summarise the current advice around mitigating these challenges, and offer a list of guiding questions for researchers when designing their own data collection methods involving the internet. While it is the responsibility of researchers to ensure ethical guidelines and practices are adhered to, including identifying the ethical and methodological repercussions of AI use, Research Ethics Committees need to be able to respond to emerging ethical challenges to research methods. This perspective is explored in more detail in Connelly et al. (2024).

## REFERENCES

- Acosta, A. (2023, February 23). *Bots and Binaries: On the Failure of Human Verification - Post45*. <https://post45.org/2023/02/bots-and-binaries-on-the-failure-of-human-verification/>
- Ada Lovelace Institute (2022). Looking before we leap: ethical review processes for AI and data science research. <https://www.adalovelaceinstitute.org/report/looking-before-we-leap/>
- Beattie, A., & Gibson, A. (2024, March 17). Something felt ‘off’ – how AI messed with our human research, and what we learned. *The Conversation*. <http://theconversation.com/something-felt-off-how-ai-messed-with-our-human-research-and-what-we-learned-225555>
- Bennett, W. L., & Livingston, S. (2023). A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority. In S. Salgado & S. Papathanassopoulos (Eds.), *Streamlining Political Communication Concepts: Updates, Changes, Normalcies* (pp. 43–73). Springer International Publishing. [https://doi.org/10.1007/978-3-031-45335-9\\_4](https://doi.org/10.1007/978-3-031-45335-9_4)
- Black, S. R., Bilbao, S., Moruzzi, C., Osborne, N., Terras, M., & Zeller, F. (2024). *The Future of Creativity and AI: Views from the Scottish Creative Industries*. A Report from Creative Informatics. Zenodo. <https://doi.org/10.5281/zenodo.10805253>
- Bybee, S., Cloyes, K., Baucom, B., Supiano, K., Mooney, K., & Ellington, L. (2022). Bots and nots: Safeguarding online survey research with underrepresented and diverse populations. *Psychology & Sexuality*, 13(4), 901–911. <https://doi.org/10.1080/19419899.2021.1936617>
- Carvalho, A. F., de Moraes, I. O. B., & Souza, T. B. (2023). Profiting from cruelty: Digital content creators abuse animals worldwide to incur profit. *Biological Conservation*, 287. <https://doi.org/10.1016/j.biocon.2023.110321>
- Chan, J. (2024). Online astroturfing: A problem beyond disinformation. *Philosophy & Social Criticism*, 50(3), 507–528. <https://doi.org/10.1177/01914537221108467>

Connelly, L., Osborne, N., Black, S.R., & Terras, M. (2024). *Guidance for research ethics committees and researchers on designing research in the age of AI*. A report from Creative Informatics. Zenodo. <https://doi.org/10.5281/zenodo.13739835>

Cornish, F., Breton, N., Moreno-Tabarez, U., Delgado, J., Rua, M., de-Graft Aikins, A., & Hodgetts, D. (2023). Participatory Action Research. *Nature Reviews Methods Primers*, 3(1), 1–14. <https://doi.org/10.1038/s43586-023-00214-1>

Cimpian, J. R. & Timmer, J. D. (2020). Mischievous Responders and Sexual Minority Youth Survey Data: A Brief History, Recent Methodological Advances, and Implications for Research and Practice. *Archives of Sexual Behavior*, 49(4), 1097–1102. <https://doi.org/10.1007/s10508-020-01661-7>

Colomina, C. (2022). *Words as weapons: From disinformation to the global battle for the narrative*. Barcelona Centre for International Affairs. <https://www.cidob.org/en/publication/words-weapons-disinformation-global-battle-narrative>

Dunham, K., & Melnick, J. (2008). *Malicious Bots: An Inside Look into the Cyber-Criminal Underground of the Internet*. Auerbach Publications. <https://doi.org/10.1201/9781420069068>

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Commun. ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>

franzke, a. s., Bechmann, A., Zimmer, M., Ess, C. and the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0*. <https://aoir.org/reports/ethics3.pdf>

Geers, M., Swire-Thompson, B., Lorenz-Spreen, P., Herzog, S. M., Kozyreva, A., & Hertwig, R. (2024). The Online Misinformation Engagement Framework. *Current Opinion in Psychology*, 55. <https://doi.org/10.1016/j.copsyc.2023.101739>

Goodrich, B., Fenton, M., Penn, J., Bovay, J., & Mountain, T. (2023). Battling bots: Experiences and strategies to mitigate fraudulent responses in online surveys. *Applied Economic Perspectives and Policy*, 45(2), 762–784. <https://doi.org/10.1002/aepp.13353>

Griffin, M., Martino, R. J., LoSchiavo, C., Comer-Carruthers, C., Krause, K. D., Stults, C. B., & Halkitis, P. N. (2022). Ensuring survey research data integrity in the era of internet bots. *Quality & Quantity*, 56(4), 2841–2852. <https://doi.org/10.1007/s11135-021-01252-1>

Irish, K., & Saba, J. (2023). Bots are the new fraud: A post-hoc exploration of statistical methods to identify bot-generated responses in a corrupt data set. *Personality and Individual Differences*. <https://doi.org/10.1016/j.paid.2023.112289>

Jackson, A. M., Woo, J., Olson, M., Dalisay, F., Pokhrel, P., Muller, C. J., & Okamoto, S. K. (2023). Methodological Challenges in Web-Based Qualitative Research With Medically Underserved Populations. *Journal of Medical Internet Research*, 25(1). <https://doi.org/10.2196/44086>

Karinshak, E., & Jin, Y. (2023). AI-driven disinformation: A framework for organizational preparation and response. *Journal of Communication Management*, 27(4), 539–562. <https://doi.org/10.1108/JCOM-09-2022-0113>

Kirchner, J. H., Ahmad, L., Aaronson, S., & Leike, J. (2023, July 20). New AI classifier for indicating AI-written text. *OpenAI Blog*. <https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text>

Knight, S., Shibani, A. & Vincent, N. (2024). Ethical AI governance: mapping a research ecosystem. *AI Ethics*. <https://doi.org/10.1007/s43681-023-00416-z>

Kollanyi, B. (2016). Automation, Algorithms, and Politics| Where Do Bots Come From? An Analysis of Bot Codes Shared on GitHub. *International Journal of Communication*, 10(0). <https://ijoc.org/index.php/ijoc/article/view/6136>

Loebenberg, G., Oldham, M., Brown, J., Dinu, L., Michie, S., Field, M., Greaves, F., & Garnett, C. (2023). Bot or Not? Detecting and Managing Participant Deception When Conducting Digital Research Remotely: Case Study of a Randomized Controlled Trial. *Journal of Medical Internet Research*, 25(1). <https://doi.org/10.2196/46523>

Pellicano, E., Adams, D., Crane, L., Hollingue, C., Allen, C., Almendinger, K., Botha, M., Haar, T., Kapp, S. K., & Wheeley, E. (2023). Letter to the Editor: A possible threat to data integrity for online qualitative autism research. *Autism*. <https://doi.org/10.1177/13623613231174543>

Przybylski, A. K. (2016). Mischievous responding in Internet Gaming Disorder research. *PeerJ*, 4. <https://doi.org/10.7717/peerj.2401>

Purtill, J. (2024, February 27). A 'great flood' of AI noise is coming for the internet and it's swallowing Twitter first. *ABC News*. <https://www.abc.net.au/news/science/2024-02-28/twitter-x-fighting-bot-problem-as-ai-spam-floods-the-internet/103498070>

Ramponi, M. (2023, June 27). Recent developments in Generative AI for Audio. *AssemblyAI*. <https://www.assemblyai.com/blog/recent-developments-in-generative-ai-for-audio/>

Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. *2023 International Conference on Computer and Applications (ICCA)*, 1–7. <https://doi.org/10.1109/ICCA59364.2023.10401723>

Simion, M. (2023). Knowledge and Disinformation. *Episteme*, 1–12. <https://doi.org/10.1017/epi.2023.25>

Simone, M. (2019, November 25). How to Battle the Bots Wrecking Your Online Study. *Behavioral Scientist*. <https://behavioralscientist.org/how-to-battle-the-bots-wrecking-your-online-study/>

Stokel-Walker, C. (2024, March 22). *Can AI Replace Human Research Participants? These Scientists See Risks*. *Scientific American*. <https://www.scientificamerican.com/article/can-ai-replace-human-research-participants-these-scientists-see-risks/>

Storozuk, A., Ashley, M., Delage, V., & Maloney, E. A. (2020). Got Bots? Practical Recommendations to Protect Online Survey Data from Bot Attacks. *The Quantitative Methods for Psychology*, 16(5), 472–481. <https://doi.org/10.20982/tqmp.16.5.p472>

UKRI, UK Research and Innovation (2023). Research ethics guidance. <https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/>

Woolfall, K. (2023). Identifying and preventing fraudulent participation in qualitative research. *Archives of Disease in Childhood*, 108(6), 421–422. <https://doi.org/10.1136/archdischild-2023-325328>

Xu, Y., Pace, S., Kim, J., Iachini, A., King, L. B., Harrison, T., DeHart, D., Levkoff, S. E., Browne, T. A., Lewis, A. A., Kunz, G. M., Reitmeier, M., Utter, R. K., & Simone, M. (2022). Threats to Online Surveys: Recognizing, Detecting, and Preventing Survey Bots. *Social Work Research*, 46(4), 343–350. <https://doi.org/10.1093/swr/svac023>

## ACKNOWLEDGEMENTS

This work was funded by Creative Informatics, AHRC grant number [AH/S002782/1](https://doi.org/10.1093/swr/svac023) as part of the Creative AI Demonstrator project, co-funded by the AHRC and the Department of Culture, Media and Sport (DCMS).

Creative Informatics is a partnership across four organisations: the University of Edinburgh, Edinburgh Napier University, Codebase, and Creative Edinburgh. The programme is part of the Creative Industries Clusters Programme, managed by the Arts and Humanities Research Council (AHRC) as part of the Industrial Strategy. Creative Informatics is also funded by, and part of, the Data-Driven Innovation initiative of the Edinburgh and South-East Scotland City Region Deal. It also benefits from additional funding from the Scottish Funding Council, and the Department for Culture, Media & Sport (DCMS).

## HOW TO CITE THIS PAPER

Black, Suzanne R., Connelly, Louise, Osborne, Nicola, and Terras, Melissa. (2024). 'Voucher Now Please: AI Bots and Fraudulent and Mischievous Responders in Online Research'. A Report from Creative Informatics. Zenodo. <https://doi.org/10.5281/zenodo.13356359>