norc.org | info@norc.org

April 2026

Fraudulent respondents and bots in nonprobability surveys: A literature review

Introduction

The recent exponential explosion of automated bots is a danger to the integrity of online survey data, specifically in nonprobability surveys. According to large-scale traffic analyses, automated bots now account for nearly half of all global internet traffic (Imperva, 2024). While bots can be useful in assisting with repetitive tasks, they can also be used by bad actors to commit fraud. It is estimated that fraud rates are at 15–30 percent across the market research industry, reaching as high as 45 percent on some survey platforms with spikes during the pandemic and continued growth fueled by the accessibility of large language models, bots, and click farms (Caven et al., 2025; Maimarides, 2025; McCarthy, 2022). In the context of survey responses, these bots pose real challenges to researchers by providing non-human technological responses. A marked increase in AI-generated responses was observed following public releases of new generative AI programs (Zhang, Katoh, & Pei, 2025). A couple of case studies also suggest that survey fraud has dramatically undermined data quality, with usable responses dropping from roughly 75% to just 10% in recent years (Pinzón et al., 2024; see also Peterson, 2025). The problem is particularly acute when surveys are accompanied by financial incentives (Teitcher et al., 2015; Bowen et al., 2008). Beyond monetary gain, bad actors may also be motivated by the desire to disrupt research, by compromising data validity, or to test and challenge their own technical skills (Buchanan & Scofield, 2018).

How bots impact data quality by recruitment source

Online surveys have been becoming increasingly vulnerable to fraudulent responses and bot activity. Early evidence from Konstan et al. (2005) showed that 11 percent of responses in a sexual behavior risk survey were fraudulent. Removing these responses altered statistical significance. More recent studies reveal even higher rates: Bell and Gift (2023) detected identity falsification in over 81 percent of respondents, while Bonnamy et al. (2025) abandoned a Qualtrics survey after determining that 95 percent of responses were likely bot-generated.

This challenge was echoed by Krawczyk and Siek (2024), who received 981 responses within four hours, of which only three were authentic. Goodrich et al. (2023) similarly reported that in a U.S. beekeeping survey, only 4 percent of 2,622 responses were legitimate. Taken together, these findings constitute a striking pattern that reflects broader shifts in data collection.

Social media recruitment has proven equally susceptible to fraud. Pozzar et al. (2020) found that 94.5 percent of responses in a health research study were fraudulent, while Imes et al. (2024) reported that over 95 percent of Twitter-recruited responses were bot-generated. Detection is also getting more difficult. Irish and Saba (2023) warned that bot-generated data can closely mimic human responses. Overly conservative exclusion criteria may also distort findings, highlighted by Carballo-Diguez and Strecher (2012). This study noted that flagged cases sometimes contribute to sample diversity and observed relationships. Collectively, these studies underscore the growing challenge of maintaining data integrity in online and social media-based research.

Evidence that it's difficult to verify respondents' identities

Verifying respondents' identities in online surveys has become increasingly difficult. Bell and Gift (2023) uncovered serious irregularities in their dataset, including evidence that respondents deliberately detected and circumvented screening measures and repeatedly submitted surveys under different identities. Bots can now leverage large language models, such as ChatGPT, to interpret survey questions and generate human-like responses, making fraudulent entries nearly indistinguishable from authentic ones (Bhayana, 2024). Westwood (2025) specifically designed an autonomous LLM-based synthetic respondent capable of passing 99.8 percent of standard data quality checks and was able to generate coherent, persona-consistent responses that were indistinguishable from human responses. These developments highlight the growing sophistication of bad actors and the limitations of traditional verification strategies.

Why these issues are hard to resolve

Resolving survey fraud is inherently complex due to methodological and technological constraints. Lawlor et al. (2021) emphasized that individuals can use bots to conceal their identity and submit multiple responses, amplifying one "voice" over genuine participants, and noted that no single method works universally across surveys. Traditional strategies such as domain-knowledge checks and open-ended questions have become ineffective against LLM-powered bots (Yee, 2025). Teitcher et al. (2015) cautioned that aggressive screening risks excluding hard-to-reach populations. Vulnerable populations and those with limited digital access may be inaccurately flagged as fraudulent, underscoring the challenge of balancing rigorous detection with equitable participation (Stemmer et al., 2026).

Ethical considerations often conflict with fraud prevention strategies. Carey and Babicheva (2024) suggested delaying disclosure of survey context until after screening questions to deter bots, but this approach may violate informed consent requirements. Gleibs and Albayrak-Aydemir (2021) noted concerns when researchers signal monitoring or deny compensation for suspicious responses, particularly given the economic vulnerability of some participants. These trade-offs highlight the tension between protecting data integrity and safeguarding participant rights.

Fraud detection is further complicated by behavioral confounds such as satisficing, where genuine respondents provide minimal effort that triggers fraud indicators. Hamby and Taylor (2016) described this phenomenon, and Donegan and Gillan (2021) along with Moeck et al. (2021) warned that overly strict exclusion criteria risk removing populations prone to attention lapses, potentially biasing results.

AI-assisted responding broadly contaminates survey measurement. The reliability of behavioral paradata like response times is blurred for a respondent using AI to assist in survey completion, making it difficult to distinguish persons from bots (Van der Stigchel et al., 2026; Westwood & Frederick, 2026). In one study of active online survey participants, 34 percent reported using LLMs to help them answer open-ended questions to better express themselves, casting doubt on whether researchers can assume all textual responses originate from the participant themselves (Zhang, Xu, & Alvero, 2025). LLM-generated responses are generally more homogeneous and emblematic of common or cultural touchstone responses which make it difficult to distinguish these responses from the opinions of an average member of the population, systematically suppressing natural variation (Martherus et al., 2025; Zhang, Xu, & Alvero, 2025).

Operational challenges persist across survey platforms. Burdick (2025) pointed to gaps in fraud prevention caused by multiple vendors and subcontractors, while Wardrop (2025) reported that strict fraud checks like IP fingerprinting can introduce bias and harm respondent experience. One study observed that eliminating financial incentives does not deter all fraud, as some actors are motivated by ideological disruption rather than monetary gain (Johnson et al., 2024). Wu et al. (2022) highlighted the trade-off between adding validity checks and risking lower response rates, underscoring the complexity of balancing security and usability.

Additionally, hybridization strategies further complicate fraud detection in online surveys. Hense and Ahmed (2025) reported that fraudsters often begin surveys manually to bypass initial screening checks, then deploy automated bots to complete the remaining questions and secure incentives. This tactic blends human and machine input, making it increasingly difficult for researchers to identify fraudulent responses using traditional validation methods.

Existing solutions and why they often fail

Existing solutions to combat survey fraud often fall short against increasingly sophisticated tactics (see Panizza et al., 2026). Pozzar et al. (2020) found that even after implementing measures to prevent misrepresentation and automated enrollment, hundreds of fraudulent respondents still managed to bypass eligibility checkpoints. Building on this, Gutierrez (2023) noted that while honeypot questions can help identify bots, more sophisticated bots are now designed to detect and avoid these hidden items (see also Martherus et al. 2025). Similarly, Booth (2024) highlighted that fraudsters increasingly rely on tools such as Browser Automation Studio, VPNs, and headless browsers to evade detection, while bots mimic human response timing and produce polished, AI-generated answers to open-ended questions. Another study observed that these bots exploit CAPTCHA-solving services, rotate IP addresses, and leverage AI-generated text to circumvent existing safeguards (Gutierrez, 2023). Even vendor solutions, such as Qualtrics fraud flags, fail against sophisticated attacks, leaving researchers to contend with persistent contamination despite screening efforts (Hitches et al., 2025). Martherus et al. (2025) found that agentic AI tools could be detected using “reverse shibboleth” questions, where we expect the target audience to answer incorrectly (in contrast to shibboleth questions where the target audience is expected to answer correctly), but the viability of that approach long-term is only limited by the time it takes for AI agents to adapt. Manual review remains one of the most effective approaches, but it is prohibitively costly and time-intensive (Griffin et al., 2022).

Some researchers are using machine learning to detect machine learning. While AI detector tools broadly have been highly touted, the evidence for their efficacy is lacking (e.g., Perkins et al., 2024), especially with a one-size-fits-all approach (e.g., Liang et al., 2023). Claassen and colleagues (2026) show that transformer-based models can identify bot-generated narrative text with up to 100 percent accuracy when tuned to a specific question context, but cross-question prediction was poor. Stemmer et al. (2026) proposed a multistage structured framework for mitigating fraudulent completes: the Configure, Assess, Triage, Corroborate, and Hone (CATCH) framework. The CATCH framework emphasizes the importance of a layered approach from pre-study configuration through post-data processing, as well as a combination of systematic and human-led methods. However, even with a multipronged approach, fraudulent responses are still likely to go undetected. amet eu lacus.”

Conclusion

BOTTOM LINE:

Bot- and AI assisted contamination is no longer an outlier case for surveys. It is a structural risk for nonprobability online surveys. With bots constituting roughly half of web activity, and credible industry estimates showing double digit fraud rates in market and academic research contexts, the usual “status quo” defenses are insufficient (e.g., single honeypots, generic CAPTCHAs, or static fraud flags). The literature shows that sophisticated actors now hybridize tactics (beginning surveys manually, then automating), evade platform-level checks, and leverage LLMs to pass many conventional quality gates, often while preserving persona coherence. Accordingly, researchers can take the following practical steps to reduce the risk of fraudulent responses:

- Layer defenses rather than rely on a single fix. Multi-stage detection approaches (such as those emphasizing configuration, corroboration, and targeted human review) are more resilient than any single detection method.

- Generic AI-detectors and static rule-based tools often fail and detection improves when reviews are tailored to study specific instruments and populations.
- Overly aggressive screening risks excluding and alienating legitimate respondents, especially those with low digital literacy, unstable broadband, or atypical response patterns.
- Collect paradata, document rules in advance, and apply exclusion criteria transparently.

THE CENTER'S PERSPECTIVE:

We recommend investing in designs that are structurally less vulnerable. While no approach is immune to fraud, probability-based surveys and panels are substantially less susceptible to the types of automated attacks, identity falsification, and volume-based bot infiltration that plague open-access, link-based nonprobability recruitment. Unlike convenience and opt-in routes (such as social media ads, survey routers, or open web links) probability-based panels rely on sampling frames, verified contact information, established recruitment protocols, and protecting ecosystems (e.g. unique PIN-based survey entry), all of which significantly constrain bot entry points and make impersonation far more difficult. These design features do not eliminate fraud risk, but they meaningfully reduce exposure and provide researchers with more robust tools for validating identities, tracking participation, and preserving sample integrity.

IN SHORT:

Researchers should combine layered fraud mitigation with a strategic shift toward probability-based survey designs, which by their very structure offer a more defensible path to trustworthy data in an era of rapidly escalating automated threats.

References

Bell, A. M., & Gift, T. (2023). Fraud in online surveys: Evidence from a nonprobability, subpopulation sample. Journal of Experimental Political Science, 10(1), 148–153. https://doi.org/10.1017/XPS.2022.8

Bhayana, R. (2024). Chatbots and large language models in radiology: A practical primer for clinical and research applications. Radiology, 310(1), e232756. https://doi.org/10.1148/radiol.232756

Bonnamy, J., Carr, B., Lazarus, M. D., & Connell, C. (2025). Survey sabotage: Insights into reducing the risk of fraudulent responses in online surveys. Anatomical Sciences Education. https://anatomypubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/ase.70015

Booth, M. (2024). Ensuring quality survey responses in an age of uncertainty. Quirk's Media. Retrieved from https://www.quirks.com/articles/ensuring-quality-survey-responses-in-an-age-of-uncertainty

Bowen, A. M., Daniel, C. M., Williams, M. L., & Baird, G. L. (2008). Identifying multiple submissions in Internet research: Preserving data integrity. AIDS and Behavior, 12(6), 964–973. https://pmc.ncbi.nlm.nih.gov/articles/PMC2615134/pdf/nihms85343.pdf

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. Behavior Research Methods, 50, 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Burdick, T. (2025, May 5). Data fraud: A threat to market research. Quirk's Media. Retrieved from https://www.quirks.com/articles/data-fraud-a-threat-to-market-research

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://pmc.ncbi.nlm.nih.gov/articles/PMC9697945/pdf/nihms-1710257.pdf

Carey, S. M., & Babicheva, V. (n.d.). Detecting and preventing BOT and fraudulent survey responses: A comprehensive overview. Retrieved from https://health.uconn.edu/clinical-research-center/wp-ontent/uploads/sites/50/2024/05/Avoiding-and-Detecting-Bots-and-Fraud_REDCapSurveys.pdf

Carballo-Diguez, A., & Strecher, V. (2012). Data quality in web-based HIV/AIDS research: Handling invalid and suspicious data. Field Methods, 24(3), 272–291. https://pmc.ncbi.nlm.nih.gov/articles/PMC3505140/pdf/nihms310913.pdf

Caven I, Yang Z, & Okrainec K. (2025) It's raining bots: how easier access to internet surveys has created the perfect storm. BMJ Open Quality. 2025;14:e003208. https://doi.org/10.1136/bmjoq-2024-003208

Claassen, J., Höhne, J. K., Bach, R., & Haensch, A. C. (2026). Identifying bots through LLM-Generated text in open narrative responses: A proof-of-concept study. Social Science Computer Review. https://doi.org/10.1177/08944393251408022

Donegan, K. R., & Gillan, C. M. (2022). New principles and new paths needed for online research in mental health: Commentary on Burnette et al. (2021). International Journal of Eating Disorders, 55(2), 278–281. https://doi.org/10.1002/eat.23670

Gleibs, I. H., & Albayrak-Aydemir, N. (2022). Ethical concerns arising from recruiting workers from Amazon's Mechanical Turk as research participants: Commentary on Burnette et al. (2021). International Journal of Eating Disorders, 55(2), 276–277. https://doi.org/10.1002/eat.23658

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://link.springer.com/content/pdf/10.1007/s11135-021-01252-1.pdf

Gutierrez, S. (2023, December 5). How can you outsmart a survey bot respondent? Here's what we learned from building one ourselves. SurveyMonkey. Retrieved from https://www.surveymonkey.com/curiosity/outsmart-bot-survey-respondent/

Hamby, T., & Taylor, W. (2016). Survey satisficing inflates reliability and validity measures: An experimental comparison of college and Amazon Mechanical Turk samples. Educational and Psychological Measurement, 76(6), 912–932. https://doi.org/10.1177/0013164415627349

Hense, M., & Ahmed, Y. (2025, October 9). Toluna virtual session: Stop bots, not people: Survey design in the age of online fraud. Quirk's Media. Retrieved from https://www.quirks.com/articles/toluna-virtual-session-stop-bots-not-people-survey-design-in-the-age-of-online-fraud

Hitches, E., Dudley, D., Johnstone, M., & Woodcock, S. (2025). Bots and baddies: Supporting the integrity of online survey research in the face of a growing challenge. Quality & Quantity, 59(2), 1481–1506. https://link.springer.com/article/10.1007/s11135-024-02001-w

Imes, C. C., Baniak, L. M., Luyster, F. S., Morris, J. L., & Orbell, S. L. (2024). Abstract P460: Attack of the (survey) bots: How we determined that our anonymous survey was full of false data. Circulation, 149(Suppl_1), AP460. https://doi.org/10.1161/circ.149.suppl_1.P460

Imperva. (2024) Bad Bot Report. 2024. Available from: https://www.imperva.com/resources/resource-library/reports/2024-bad-bot-report/

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, 213, 112289. https://doi.org/10.1016/j.paid.2023.112289

Johnson, M. S., Adams, V. M., & Byrne, J. (2024). Addressing fraudulent responses in online surveys: Insights from a web-based participatory mapping study. People and Nature, 6(1), 147–164. https://doi.org/10.1002/pan3.10557

Konstan, J. A., Simon Rosser, B. R., Ross, M. W., Stanton, J., & Edwards, W. M. (2005). The story of subject naught: A cautionary but optimistic tale of Internet survey research. Journal of Computer-Mediated Communication, 10(2). https://onlinelibrary.wiley.com/doi/full/10.1111/j.1083-6101.2005.tb00248.x

Krawczyk, M., & Siek, K. A. (2024). When research becomes all about the bots: A case study on fraud prevention and participant validation in the context of abortion storytelling. Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA. https://doi.org/10.1145/3613905.3637109

Lawlor, J., Thomas, C., Guhin, A. T., Kenyon, K., Lerner, M. D., Consortium, U., et al. (2021). Suspicious and fraudulent online survey participation: Introducing the REAL framework. Methodological Innovations, 14(3), 20597991211050467. https://doi.org/10.1177/20597991211050467

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7). https://doi.org/10.1016/j.patter.2023.100779

Maimarides, A. (2025, August 5). When bots take surveys: AI's impact on the integrity of market research. Lab42. Retrieved from https://www.lab42.com/blog/when-bots-take-surveys-ais-impact-on-the-integrity-of-market-research

Martherus, J., Podkul, A., Cook, E., & Liebowitz, R. (2025). How to detect AI-assisted interviews in online surveys. Survey Practice, 18. https://doi.org/10.29115/SP-2025-0016

McCarthy, T. (2022, May 31). How to stop professional survey cheaters. Quirk's Media. Retrieved from https://www.quirks.com/articles/how-to-stop-professional-survey-cheaters

Moeck, E. K., Bridgland, V. M., & Takarangi, M. K. (2022). Food for thought: Commentary on Burnette et al. (2021) "Concerns and recommendations for using Amazon MTurk for eating disorder research". International Journal of Eating Disorders, 55(2), 282–284. https://doi.org/10.1002/eat.23671

Panizza, F., Kyrychenko, Y., & Roozenbeek, J. (2026). How to deal with the survey-taking AI agents that threaten to upend social science. Nature, 650, 293-295. https://doi.org/10.1038/d41586-026-00386-2

Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughran, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. International Journal of Educational Technology in Higher Education, 21(1), 53. https://doi.org/10.1186/s41239-024-00487-w

Peterson, T. (2025). The impact of AI-generated responses on environmental survey data from MTurk. Journal of Applied Statistics: Environmental Statistics and Data Science, 1-16. https://doi.org/10.1080/29984688.2025.2545754

Pinzón, N., Koundinya, V., Galt, R. E., Dowling, W. O. R., Baukloh, M., Taku-Forchu, N. C., ... & Pathak, T. B. (2024). AI-powered fraud and the erosion of online survey integrity: An analysis of 31 fraud detection strategies. Frontiers in Research Metrics and Analytics, 9, 1432774. https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2024.1432774/pdf

Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulsky, J. A., Hong, F., ... & Berry, D. L. (2020). Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. Journal of Medical Internet Research, 22(10), e23021. https://www.jmir.org/2020/10/e23021/

Russ-Smith, J., & Lazarus, M. (2024). The AI (R)evolution: Valuing country, culture and community in a world of algorithms. https://doi.org/10.26180/27274938.v3

Stemmer, M., Tauscher, J., Buck, B., Wedgeworth, P., Bear Don't Walk, O. J., Cohen, T., & Ben-Zeev, D. (2026). Practical guidance for mitigating fraud in online research: The Configure, Assess, Triage, Corroborate and Hone (HATCH) framework. Digital Health, 12, 1-6. https://doi.org/10.1177/20552076261418807

Teitcher, J. E., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to "fraudsters" in internet research: Ethics and tradeoffs. Journal of Law, Medicine & Ethics, 43(1), 116–133. https://doi.org/10.1111/jlme.12200

Van der Stigchel, S., Strauch, C., de Zwart, B., & Van Maanen, L. (2026). Will online behavioral research follow the fate of online survey research? Proceedings of the National Academy of Sciences, 123(8), e2535585123. https://doi.org/10.1073/pnas.2535585123

Wardrop, B. (2025, March 17). Tackling survey fraud: Insights from CloudResearch & Toluna at Quirks LA 2025. CloudResearch. Retrieved from https://www.cloudresearch.com/resources/blog/survey-fraud-toluna-cloudresearch-quirks-la/

Westwood, S. J. (2025). The potential existential threat of large language models to online survey research. Proceedings of the National Academy of Sciences, 123(8), e2518075122. https://doi.org/10.1073/pnas.2518075122

Westwood, S. J. & Frederick, S. (2026). Reply to Van der Stigchel et al.: Empirical evidence that AI survey contamination is real and substantial. Proceedings of the National Academy of Sciences, 122(47), e2537420123. https://doi.org/10.1073/pnas.2537420123

Wu, M.-J., Zhao, K., & Fils-Aime, F. (2022). Response rates of online surveys in published research: A meta-analysis. Computers in Human Behavior Reports, 7, 100206. https://doi.org/10.1016/j.chbr.2022.100206

Ye, G., Tang, Z., Fang, D., Zhu, Z., Feng, Y., Xu, P., et al. (2018). Yet another text captcha solver: A generative adversarial network based approach. Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, Toronto, Canada. https://doi.org/10.1145/3243734.3243754

Yee, N. (2025, June 17). Fabricated data at scale: The state of survey panels & AI data. Quantic Foundry. Retrieved from https://quanticfoundry.com/2025/06/17/panel-data-ai/

Zhang, D., Katoh, M., & Pei, W. (2025) Detecting the use of generative AI in crowdsourced surveys: Implications for data integrity. arXiv preprint. https://arxiv.org/html/2510.24594v2

Zhang, S., Xu, J., & Alvero, AJ. (2025). Generative AI meets open-ended survey responses: Research participant use of AI and homogenization. Sociological Methods & Research, 54(3), X-X. https://doi.org/10.1177/00491241251327130

Appendix. University guidance on dealing with bots in human subjects' research

- BOT ATTACKS and Human Subjects Research - UNC Research
- Guidance on Preventing Bots and Fraudulent Responses in Online Surveys - Office of Research Compliance - University of Maine
- Columbia University, Teachers College: Bots and Online Human Subjects Research (external link)
- University of California, Los Angeles: Tip Sheet: Online Survey Protection Considerations (PDF) (external link)
- University of Kentucky Research: Survey Research (external link)

ABOUT NORC

NORC at the University of Chicago conducts research and analysis that decision-makers trust. As a nonpartisan research organization and a pioneer in measuring and understanding the world, we have studied almost every aspect of the human experience and every major news event for more than eight decades. Today, we partner with government, corporate, and nonprofit clients around the world to provide the objectivity and expertise necessary to inform the critical decisions facing society.