More Parameters Than Populations: A Systematic Review of Large Language Models in Survey Research

Trent D Buskirk
School of Data Science
Old Dominion University
Norfolk, VA 23529, USA
tbuskirk@odu.edu

Florian Keusch
School of Social Sciences
University of Mannheim
Mannheim, Germany
f.keusch@uni-mannheim.de

Leah von der Heyde
Department of Statistics
LMU Munich
Munich, Germany
l.heyde@lmu.de

Adam Eck
Computer Science Department
Oberlin College
Oberlin, OH 44074, USA
aeck@oberlin.edu

Abstract

Survey research has a long-standing history of being a human-powered field, but one that embraces various technologies for the collection, processing, and analysis of various behavioral, political, and social outcomes of interest, among others. At the same time, Large Language Models (LLMs) bring new technological challenges and prerequisites in order to fully harness their potential. In this paper, we report work-in-progress on a systematic literature review based on keyword searches from multiple large-scale databases as well as citation networks that assesses how LLMs are currently being applied within the survey research process. We synthesize and organize our findings according to the survey research process to include examples of LLM usage across three broad phases: pre-data collection, data collection, and post-data collection. We discuss selected examples of potential use cases for LLMs as well as its pitfalls based on examples from existing literature. Considering survey research has rich experience and history regarding data quality, we discuss some opportunities and describe future outlooks for survey research to contribute to the continued development and refinement of LLMs.

1 Introduction

Large language models (LLMs) are rapidly transforming many professional domains, including survey research. Eloundou et al. (2024) rank survey research among the most highly exposed occupations to LLM-driven automation, raising both opportunities and challenges for practitioners. While survey research has a rich tradition of adopting technological tools for tasks like data collection, analysis, and instrument design (e.g., Waksberg, 1978; Baker, 1992; Couper, 2000; 2013), the unique affordances and risks associated with LLMs call for a structured examination. Jansen et al. (2023) provide largely a conceptual overview of the potential uses and considerations for incorporating LLMs within the survey research context. Since their work was published, the field is transitioning from an ideation phase to an implementation phase. This paper aims at expanding our understanding of the potential and concerns of this technology within the survey research context by presenting work-in-progress on a systematic literature review of empirical and theoretical work at the intersection of LLMs and survey research. Specifically, we synthesize examples of how LLMs are being applied across three broad phases of the survey research pipeline: pre-data collection, data collection, and post-data collection.

2 Systematic Literature Review

Our systematic literature review process identified 1,766 candidate publications retrieved from keyword searches using 10 LLM- and 62 survey-related keywords (see Table A1 in the Appendix) across arXiv, Semantic Scholar, and Web of Science databases and citation networks (see Figure 1). After deduplication across these three databases, we retained 1,099 unique papers. Of those, 134 were not written in English or appeared to be outside of our 6-year publication window spanning from January 2019 through January 2025 and were thus excluded. The remaining 965 papers underwent abstract review by members of our team and produced a total of 161 papers that were deemed eligible for inclusion. In addition, we identified another set of 28 papers as citing or cited papers not already included in the collection of 161 eligible papers. One of the authors performed a review of the total of 189 papers that discuss applications of LLMs within the survey research process.

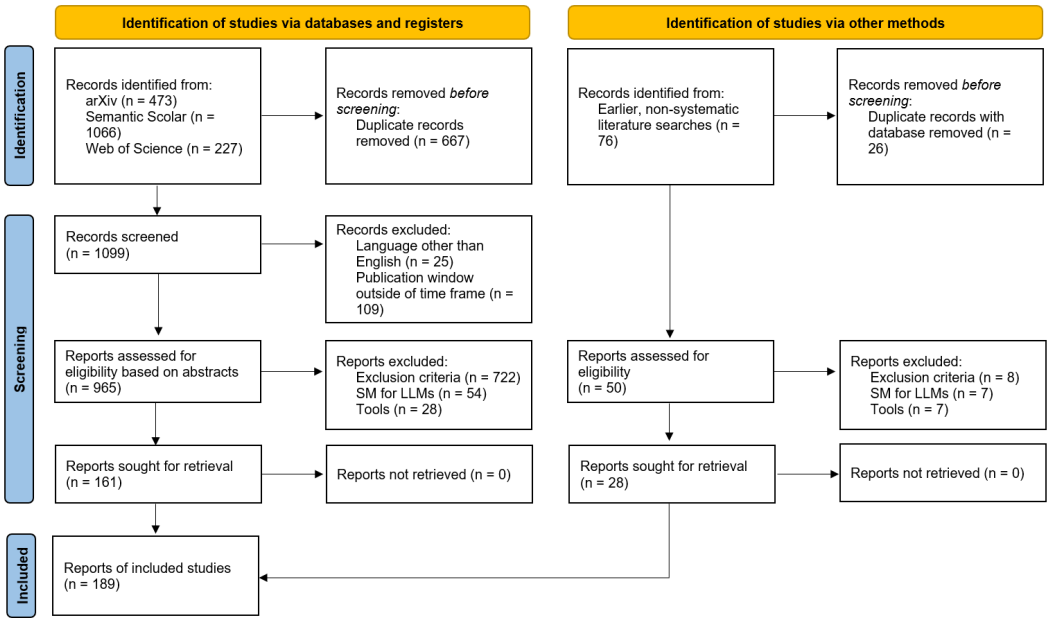


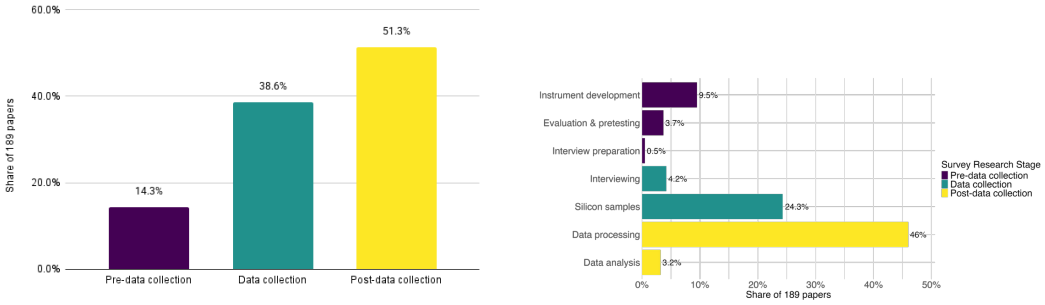
Figure 1: PRISMA 2020 flow diagram for systematic review (Page et al., 2021)

3 Application of LLMs in Three Phases of the Survey Process

We classified the papers in our corpus based on in which of three phases of the survey process an LLM is applied. The *pre-data collection phase* considers tasks such as questionnaire writing, item generation, translation, sampling designs, and recruitment materials and efforts. The *data collection phase* encompasses AI-assisted interviews, silicon sampling, and similar tasks that result in produced data. Finally, the *post-data collection phase* includes data processing tasks, weighting, imputation, summarization, report generation, and dataset curation.

Initial findings indicate that empirical applications of LLMs concentrate in three areas (see Figure 2): instrument development, synthetic respondent modeling, and automated text classification. For instance, Adhikari et al. (2025) demonstrate how LLMs can simulate cognitive interviews, enabling automated pretesting of cross-cultural questionnaires. Their method adapts survey items from U.S. contexts to new populations (e.g., South Africa), revealing both the promise and pitfalls of LLM personas in usability assessment. For generating synthetic data, Cho et al. (2024) introduce LLM-based "doppelgänger" models, which fine-tune LLMs on individual-level conversational data to forecast participant responses across a survey, raising questions about generalizability, personalization, and

bias. LLMs have also been deployed for open-text data classification, such as the work of Li et al. (2024) on accessibility sentiment extracted from Google Maps reviews to inform urban design. Meanwhile, Bisbee et al. (2024) and Mellon et al. (2024) issue cautionary notes, highlighting the risk of LLMs hallucinating responses or misclassifying nuanced political opinion data—thereby raising concerns about representational accuracy and construct validity in "silicon samples."



(a) Distribution by phase

(b) Distribution by application

Figure 2: Distribution of papers by phase and application

4 Potential New Tools

We also saw a smaller number of papers that we identified as tools or methods that are being developed with LLMs that have the potential to be applied in the survey research process but haven't yet been. For example, work is being done to identify LLM output (Nathanson et al., 2024) that could be used to assist in fraudulent respondent detection. There are also new methods and frameworks being developed for synthetic samples and data generation using LLMs (e.g., Balog et al., 2024) that could enhance related synthetic sampling approaches that are currently being explored within survey research.

5 Survey Research Improving LLMs

While we focused on papers in which LLMs were used within the survey research process, there were a fair number of excluded papers that pointed to ways in which survey research could also improve LLMs. A few themes that emerged among these papers were the curation and creation of training or benchmark data for improving or evaluating LLM output (Xie et al., 2024) as well as the use of surveys to understand perceptions and use cases of LLMs in various industries (Gasparini et al., 2024; Johnson et al., 2024).

6 Conclusion

Methodologically, the current state of our review identifies an uneven distribution of LLM applications across the survey pipeline. While pre-data collection stages (e.g., item writing, translation) are well explored, core practices such as live interviewing, recruitment, and cross-lingual adaptation remain under-investigated. This stands in contrast to LLMs' key features as multi-modal, multi-lingual processors and producers of natural language, offering room for further important research. Additionally, few studies assess LLMs systematically across multiple populations, languages, or survey topics, leaving questions about the generalizability of successful applications.

Our work highlights not just the breadth of current use cases, but also the methodological and ethical considerations that must accompany them. Given survey research's long-standing expertise in assessing error sources and ensuring data quality, it is well-positioned to guide responsible and effective LLM integration into scientific workflows.

7 Next Steps

In a next step, we will continue the work on the review of the 189 papers in our corpus. We will implement double-coding by augmenting the manual coding performed by one of the authors with an AI-based approach. In addition to coding in what specific phase of the survey process and for what particular application the LLMs are used, we will also classify whether papers are theoretical or empirical, which LLM(s) are discussed, which populations are covered or languages are applied, the substantive domain and topic, and under what conditions the application is deemed successful. Going forward, our review will reveal how LLM training data, model architecture, as well as research designs (see, e.g., von der Heyde, 2025) pose challenges to augmenting survey research with LLMs.

Author Contributions

Trent Buskirk: Conceptualization, Methodology, Writing - Original Draft. **Florian Keusch:** Conceptualization, Methodology, Writing - Review & Editing. **Leah von der Heyde:** Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing, Visualization. **Adam Eck:** Conceptualization, Methodology, Investigation, Data Curation, Writing - Review & Editing.

Acknowledgments

We would like to acknowledge our student research assistants who assisted with our literature review: Elinor Frost and Reefayat Bin Shahjahan at Oberlin College and Ella Häcker and Leonard Bek at the University of Mannheim.

References

Divya Mani Adhikari, Vikram Kamath Cannanure, Alexander Hartland, and Ingmar Weber. Exploring llms for automated pre-testing of cross-cultural surveys, 2025. URL https://arxiv.org/abs/2501.05985.

Reginald P. Baker. New technology in survey research: Computer-assisted personal interviewing (capi). *Social Science Computer Review*, 10(2):145–157, 1992. doi: 10.1177/089443939201000202. URL https://doi.org/10.1177/089443939201000202.

Krisztian Balog, John Palowitch, Barbara Ikica, Filip Radlinski, Hamidreza Alvari, and Mehdi Manshadi. Towards realistic synthetic user-generated content: A scaffolding approach to generating online discussions, 2024. URL https://arxiv.org/abs/2408.08379.

James Bisbee, Joshua D Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M Larson. Synthetic replacements for human survey data? the perils of large language models. *Political Analysis*, 32(4):401–416, 2024.

Suhyun Cho, Jaeyun Kim, and Jang Hyun Kim. Llm-based doppelgänger models: Leveraging synthetic data for human-like responses in survey simulations. *IEEE Access*, 12: 178917–178927, 2024. doi: 10.1109/ACCESS.2024.3502219.

Mick P. Couper. Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64(4):464–494, 2000. doi: 10.1086/318641.

Mick P. Couper. Is the sky falling? new technology, changing media, and the future of surveys. *Survey Research Methods*, 7(3):145–156, 2013. doi: 10.18148/srm/2013.v7i3.5751.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. Gpts are gpts: Labor market impact potential of llms. *Science*, 384(6702):1306–1308, 2024. doi: 10.1126/science.adj0998. URL https://www.science.org/doi/abs/10.1126/science.adj0998.

Loretta Gasparini, Nitya Phillipson, Daniel Capurro, Revital Rosenberg, Jim Buttery, Jayne Howley, Sarath Ranganathan, Catherine Quinlan, Niloufer Selvadurai, Michael Wildenauer, Michael South, and Gerardo Luis Dimaguila. A survey of large language model use in a hospital, research, and teaching campus. *medRxiv*, 2024. doi: 10.1101/2024.09.11.24313512. URL https://www.medrxiv.org/content/early/2024/09/12/2024.09.11.24313512.

Bernard J Jansen, Soon-gyo Jung, and Joni Salminen. Employing large language models in survey research. *Natural Language Processing Journal*, 4:100020, 2023.

Nicole Johnson, Jeff Seaman, and Julia Seaman. The anticipated impact of artificial intelligence on us higher education: A national study. *Online Learning*, 28(3):9–33, 2024. doi: 10.24059/olj.v28i3.4646. URL https://olj.onlinelearningconsortium.org/index.php/olj/article/view/4646.

Cheng Li, Jindong Wang, Yixuan Zhang, Kaijie Zhu, Wenxin Hou, Jianxun Lian, Fang Luo, Qiang Yang, and Xing Xie. Large language models understand and can be enhanced by emotional stimuli. *arXiv preprint arXiv:2307.11760*, 2023.

Lingyao Li, Songhua Hu, Yinpei Dai, Min Deng, Parisa Momeni, Gabriel Laverghetta, Lizhou Fan, Zihui Ma, Xi Wang, Siyuan Ma, Jay Ligatti, and Libby Hemphill. Toward satisfactory public accessibility: A crowdsourcing approach through online reviews to inclusive urban design, 2024. URL https://arxiv.org/abs/2409.08459.

Jonathan Mellon, Jack Bailey, Ralph Scott, James Breckwoldt, Marta Miori, and Phillip Schmedeman. Do ais know what the most important issue is? using language models to code open-text social survey responses at scale. *Research & Politics*, 11(1): 20531680241231468, 2024. doi: 10.1177/20531680241231468. URL https://doi.org/10.1177/20531680241231468.

Samuel Nathanson, Yungjun Yoo, David Na, Yinzhi Cao, and Lanier Watkins. A step towards modern disinformation detection: Novel methods for detecting llm-generated text. In *MILCOM 2024 - 2024 IEEE Military Communications Conference (MILCOM)*, pp. 615–620, 2024. doi: 10.1109/MILCOM61039.2024.10773838.

Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. 372, 2021. doi: 10.1136/bmj.n71. URL https://www.bmj.com/content/372/bmj.n71. Publisher: BMJ Publishing Group Ltd _eprint: https://www.bmj.com/content/372/bmj.n71.full.pdf.

Leah von der Heyde. Who counts? the potentials and pitfalls of using llms in survey research. In *Proceedings of the First Workshop on Bridging NLP and Public Opinion Research (NLPOR @ COLM 2025)*, 2025. Forthcoming.

Joseph Waksberg. Sampling methods for random digit dialing. *Journal of the American Statistical Association*, 73:40–46, 1978. doi: 10.1080/01621459.1978.10479995.

Zhentao Xie, Jiabao Zhao, Yilei Wang, Jinxin Shi, Yanhong Bai, Xingjiao Wu, and Liang He. Mindscope: Exploring cognitive biases in large language models through multi-agent systems, 2024. URL https://arxiv.org/abs/2410.04452.

A Appendix

Table A1: List of keywords used in the search strategy.

Group	Keyword
AI	Gemini
AI	GenAI
AI	Generative AI
AI	GPT
AI	large language model
AI	LLaMA
AI	LLM
AI	Mistral
AI	pre-trained transformer
AI	prompt engineering
survey	attitude measurement
survey	bot respondent
survey	chatbot survey
survey	cognitive interview
survey	construct validity
survey	coverage error
survey	cross-cultural survey
survey	data representativity
survey	fraudulent respondent
survey	hard to survey
survey	human response
survey	human sample
survey	item development
survey	longitudinal survey
survey	measurement error
survey	multi-lingual survey
survey	nonresponse
survey	open ended answer
survey	open ended question
survey	open ended response
survey	opinion mining
survey	opinion prediction
survey	panel survey
survey	polling
survey	pretesting
survey	public opinion
survey	public poll
survey	question development
survey	question wording
survey	questionnaire design
survey	questionnaire development
survey	response bias
survey	sampling design
survey	sampling error
survey	sampling frame
survey	scale development
survey	silicon sample
survey	silicon sampling
survey	social desirability
survey	survey data analysis
survey	survey data collection
survey	survey data imputation

Group	Keyword
survey	survey data quality
survey	survey design
survey	survey development
survey	survey error
survey	survey experiment
survey	survey interview
survey	survey invitation
survey	survey item
survey	survey methodology
survey	survey methods
survey	survey question
survey	survey questionnaire
survey	survey recruitment
survey	survey research
survey	survey respondent
survey	survey response
survey	synthetic sample
survey	synthetic sampling
survey	synthetic survey data
survey	text answer

Inclusion criteria

- **Screening**
  - **English** language
  - **Published between 2019 and before search date** (Feb 4, 2025)
- **Eligibility**
  - Access to **full paper**
  - Paper about use of **LLMs in survey research or public opinion research**
  - **Theoretical discussion** or **empirical application**
  - Use of LLMs in...
    - **pre-data collection or generation** including questionnaire development (e.g., item generation, translation), evaluation, pretesting (e.g., cognitive interviewing, focus groups), sample design (e.g., frame construction, sampling plan), recruitment preparation (e.g., drafting of invitation letters, interviewer training), and keyword selection;
    - **data collection or generation**, including sample management (e.g., call scheduling), data generation (e.g., silicon samples), and AI interviewing (LLMs embedded for augmentation or used for autonomous interviewing);
    - **post-data collection or generation**, including data processing (e.g., the extraction or coding of open-ended questions, social media data, or other digital trace data indicative of public opinion or behavior, such as comments on news articles; weighting, imputation); data analysis, summarization, and visualization; and dissemination (e.g., report writing, codebook creation, creation of public use files)
  - Data that represents **information about human attitudes and behaviors** to include survey data, administrative data, social media, and digital trace data (e.g., user posted comments or blog posts) and sensor data.
  - **LLMs** include everything starting with BERT (i.e., "pre-trained transformers")

Exclusion criteria

- **Screening**

  - **Non-English** language
  - **Published before 2019 or after search date** (Feb 4, 2025)
- **Eligibility**
  - **Only abstract** available
  - Papers about **"surveys" of LLM literature** (i.e., literature reviews)
  - Papers about **opinions/attitudes about LLMs/AI or their use**
  - Papers using LLMs in the context of **(Q&A) testing/evaluation and diagnostics** (e.g., education, medicine, law)
  - Papers **conducting lab or field experiments** with LLMs as the unit of analysis [Survey Research helping LLMs category]
  - Papers exclusively using **news content**
  - Papers that develop **methods that are not used in context of surveys and public opinion research yet** (but could be in the future) [TOOL category]
  - Papers that **use survey data to assess quality of LLMs** [SM helping LLMs category]
  - Papers using ML/NLP methods that are **not LLMs or predate BERT LLMs**