Empowering Surveys with Generative AI

Johanna Laiho-Kauranne[1]

[1] DAIN Studios, Finland
e-mail: johanna.laiho-kauranne@dainstudios.com

Abstract

While the AI has been used widely for few years across industries, popularity of generative AI has boomed since the end of 2022 in many daily applications and digital services, utilizing its feature to generate novel content rather than building logic upon existing data, (García-Peñalvo and Váquez-Ingelmo, 2023). Generative AI has the potential to revolutionize many industries, e.g advertising, entertainment, and education (Gozalo-Brizuela and Garrido-Merchan, 2023). This study focuses on its usability for survey industry. Encouraging results by Beck, Dumbert and Feuerhake (2018) show that already in 2018 majority of OECD countries had some applications of machine learning in official statistics, thus it is expected that also new methods such as generative AI will raise interest if ensured the responsible and ethical usage in the field of official statistics and surveys.

The usage of generative AI has potential but also caveats as the facts and fiction can be indistinguishable for service users, and if not used responsibly can deteriorate public trust. In addition, data protection of generative AI solutions cause concerns amongst the public, and experts of data protection professionals. Significant risks related to *bad data* generate by AI are inaccurate decision making, spreading misinformation, privacy violations, legal liabilities, damage to trust (Tang et al., 2023), and may further impact trust to democracy (Arguedas and Simon, 2023). Taking the challenging risks into account, the advantages for improving surveys utilizing the potential of generative AI is the main incentive to explore responsible and ethical use of generative AI for survey research.

There is a global unmet challenge for surveys that are battling with the long prevailing trend of reduction in response rates, reaching levels of severe questioning of the reliability of the estimates. This has been studied for decades, but instead of finding effective strategies for promising initiatives, the survey organizations are increasingly struggling with low or negligible participation to surveys impacting the reliability and accuracy of the survey estimates. Generative AI can be used as a tool survey design in multiple steps of survey process. This is demonstrated in the context of the GSBPM model. Potential use case is defined as using generative AI especially in the design and data collection stages of the GSBPM model, e.g. in improving survey design, respondent approaches, contact strategies and conversion strategies for soft refusals.

Considering the relation to the fundamental aim of surveys providing reliable facts about society or population, the use of generative AI may be argued to be in contradiction with the ultimate purpose of generating reliable information. Thus, it is emphasized that the focus is not on supplementing and creating information arbitrary information content, instead the purpose is to explore usage of generative AI to improve the quality of surveys and the way their processes are managed. In other words, the AI-generative content (AIGC) has been exclude from this study. The data quality components in focus are survey accuracy, and relevance, reviewing also potential impact to survey comparability, timeliness, and accessibility.

The study explores the potential use of generative AI for response conversion strategies. The traditional differentiative motivation strategies base on ad hoc strategies or tailored data analytics can be further enriched using large language models (LLM). These LLMs can be developed so that it is

Indistinguishable from human-generate content (Arguedas & Simon, 2023). The use case is narrowed to reviewing potential applications that require quick adaptation of strategies to build motivation and raise interest of the selected units to be surveyed. Application potential may also reach to support tool for interviewers, or for contact persons to help them to build the best motivation strategies for retrieval of requested information. New technology is fast developing, and promising new solutions are using generative AI text-to-speech diffusion model applications, which features also provide incorporated controllable emotional models (Zhang et al. 2023).

To conclude, AI can be used to explore the validity and relevance of the underlying assumptions of the survey in fast developing societies and survey phenomena. The purpose of this study is to evaluate suitability of generative AI solutions to traditional population surveys. As the usage of generative AI is likely to raise contradiction and successful surveys are based on the trust of data providers, we review the acceptance for information production and survey industry in the AI landscape. As the EU is currently working on the world's first AI law, it is important to pave the way for utilizing future potential and understand the scalability, limitations, and risks of AI in an objective manner. Thus, statistical ethics and code of practice, merely in the European context are reviewed to examine the fit for purpose and preparedness of the legislative grounds.

As the legislative groundings are still evolving, and the generative AI is rather recent, this study is limited to conceptual and theoretical level, and the purpose is to be a starting point for further pilot cases.

Keywords: generative AI, surveys, GSBPM, survey design, respondent conversion.

References

Arguedas, A. & Simon, F. (2023). *Automating democracy: Generative AI, journalism, and the future of democracy.* Institute for Ethics in AI.

Beck, M., Dumbert, F., and Feuerhake, J. (2018). *Machine Learning in Official Statistics*. Preprint.

Gozalo-Brizuela, R. and Garrido-Merchan, E. (2023) *A survey of Generative AI Applications*. Preprint.

García-Peñalvo, F. and Váquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. *International Journal of Interactive Multimedia and Artificial Intelligence.*

Zhang, C., et al. (2023) *A Survey on Audio Diffusion Models: Text To Speech Synthesis and Enhancement in Generative AI*. Kyung Hee University, South Korea. Manuscript submitted to ACM.