

High Agreement, Different Stories: How LLM Classifiers Reshape Demographic Patterns in Survey Data

Word Count: 6,047 (excluding figures, tables, references, and appendices)

Author:

Chris Soria

Department of Demography, University of California, Berkeley, CA, USA

Corresponding Author Contact Information:

Chris Soria

Department of Demography

University of California, Berkeley

Social Sciences Building, Floor 3

Berkeley, CA 94720

Email: ChrisSoria@Berkeley.edu

Telephone: (831) 585-9507

ORCID iD: <https://orcid.org/0009-0007-3558-5896>

Author Biography:

Chris Soria is a PhD candidate in Demography at the University of California, Berkeley. His research focuses on computational social science methods, social networks, and the application of large language models to qualitative data analysis. His work examines how emerging AI technologies can augment—and where they fall short of—traditional qualitative research methods in understanding social phenomena.

Acknowledgments:

The author thanks Professor Claude Fischer for his collaboration in developing the mixed inductive-deductive codebook used in this study and his thoughtful feedback and editing. The author also thanks Paige Park, Dennis Feehan, and Juana Montoya for their careful feedback on the manuscript.

Funding:

This work was supported by the Greater Good Science Center Libby Fee Fellowship; the Bashir Ahmed Graduate Fellowship; a Mentored Research Award from the University of California, Berkeley; the National Institute of Child Health and Human Development [P2CHD073964]; and the National Institute on Aging [P30AG012839].

Declaration of Conflicting Interests:

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