

## **Challenges of Data Collection in the Age of AI**

1) Strategies to Prevent Fraudulent Responders Infiltrating Online Survey Research

2) Prevalence of Fraudulent Responders Affecting Online Survey Research

Panelists (1 & 2):

- Shubham Agrawal, Assistant Professor, Department of Psychology
- Shelia R. Cotten, Associate Vice President for Research Development, Provost's Distinguished Professor, Department of Sociology, Anthropology, and Criminal Justice, Department of Communication
- Amy M. Schuster, Research Assistant Professor, Department of Psychology
- Paige Watson, Assistant Professor, Department of Psychological Sciences, Auburn University

3) Responsible Data Collection Practices for Human-Subjects Research

Panelists:

- Kelly Caine, Professor, School of Computing
- Bart Knijnenburg, Associate Professor, School of Computing
- Emily A. Sidnam-Mauch, Research Scientist, School of Computing

4) Social Media Data Collection in the Age of AI

Panelist: Long Cheng, Assistant Professor, School of Computing

Online surveys are a commonly used data collection method due to their cost-effectiveness, convenience, and efficiency. With the increase of human and bot fraudulent responders, data quality arises as a critical concern. This panel presents researchers' experience and knowledge of fraudulent responders/spammers infiltrating online surveys, along with new methods to improve data collection. First, we will discuss the WEAVE (Preparing the Future Workforce for the Era of Automated Vehicles) project strategies enacted to prevent fraudulent responders/spammers and approaches to online survey data cleaning. Next, we will expand on the extent of fraudulent responders/spammers infiltrating an online survey by presenting results from a study examining researchers' (N=682) knowledge and encounters with fraudulent responders/scammers in online surveys.

We will discuss research on improving the data collection practices of AI researchers. Finally, we will present results on a data collection tool that uses large language models to assist in data annotation.