September 24: Indira Sen

Is AI a Friend of Foe to Survey-based Data Collection?

AI Systems, including Large Language Models (LLMs) like ChatGPT or Claude, have disrupted many aspects of our lives, including how we learn about people's attitudes and behaviors. LLMs, due to their fluent natural language generation, can be used for many human-centred tasks, e.g., designing survey questions or processing free-text data such as social media or survey responses to surface novel insights. In a more radical approach, because they have been trained on human-generated behavioral data, LLMs might even mimic human behavior. This potential mimicry offers both promise and peril --- LLM simulations of people can catalyze the field of computer-driven social simulations, allowing scientists to study and test behavior in cases where real human data is hard to get, e.g., experiments that might cause harm. However, it can also lead to 'LLM pollution' in social data collection where seemingly real human data is tainted by non-human entities. To further complicate matters, an LLM's ability to simulate people faithfully is far from established. In this talk, I will discuss how LLMs, and other AI systems, can both benefit from and provide benefit to survey methodology. I will focus on (1) how we can use existing survey instruments to evaluate the behavior of LLMs, (2) how we can establish the limits of LLM simulations and respresentativeness, and (3) what it actually means to ask an LLM a question designed for a human.

October 1: Tobias Rettig

AI Tools in Applied Survey Research and Survey Management: Current Developments in the German Internet Panel

The GIP is the longest-running probability online panel of the German population. Operating since 2012 with currently over 4,000 active respondents from four different probability samples who answer six surveys per year, the GIP has a wealth of existing survey data and tight schedules for new panel waves. To leverage the advantages of the GIP, support both researchers working with GIP data and the GIP team in conducting the fieldwork, as well as exploring new innovations in the field of survey methodology, we are currently working on several applications for AI support tools in the survey lifecycle: (1) make GIP data more findable, accessible, and replicable using an interactive agent trained on all questions from over 80 questionnaires over 14 years that can answer questions such as: Has this topic every been surveyed? When and how often? What were the exact questions? (2) create a tool for automated coding of open text responses and respondent feedback that can identify points of criticism, identify which comments require a response, and draft a response to them. (3) train a model on the wealth of existing GIP data to create artificial “in-silico” respondents based on the real GIP respondents, for example as a more comprehensive tool for imputation of missing data or to lighten response burden by only asking some of the questions on a given topic and inferring the responses to the rest.

October 8: Lea Cohausz, Daan Brückner-Collet

Rethinking Study Orientation: Combining Educational Psychology and Machine Learning

In Germany, roughly 25% of bachelor's students drop out of their programs, resulting in high societal and personal costs. Although multiple factors contribute to dropping out, research has shown that the main reasons are a mismatch between the study program's content and personal interests, as well as misjudging the program’s demands.
Interest tests for prospective higher education students that link their interests to study programs exist and are even mandatory in Baden-Württemberg in an effort to reduce drop-out, but have known weaknesses. As part of a project with the state of Baden-Württemberg, we

develop a novel adaptive interest test that automatically extracts items from module catalogs using large language models and that selects items based on previous answers using an active learning strategy. This procedure allows the interest test to start broad and general and to get more specific and personalized as more information about the student is provided. In our lecture, we will describe the motivation and the requirements for the project, explain the methodology, and provide an outlook on how we monitor and investigate the interest test. The project represents a continued collaboration between the chairs of Educational Psychology and Artificial Intelligence and highlights the value of such interdisciplinary collaborations.

October 15: Heiner Stuckenschmidt

tbd

October 22: Florian Keusch

AI in the Loop?! Use of Large Language Models in Survey and Public Opinion Research

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 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 for survey and public opinion research. In this lecture, I will discuss potential applications of LLMs across three broad phases of the research process: pre-data collection, data collection, and post-data collection. I will present the results of a quantitative and qualitative assessment of 136 studies, synthesizing at which tasks, contexts, and research designs LLMs succeed and fail and which applications are underexplored. Based on the findings, I will assess the potentials, concerns, and gaps of LLM use in survey and public opinion research, providing a starting point for future innovations.

October 29: Leah von der Heyde

Silicon Surveys: Applications, Implementations, and Error Sources

The proliferation of AI and the continued challenges with recruiting participants for surveys have led to a rapidly increasing use of so-called silicon samples – survey data generated by large language models (LLMs). The quality of this data in terms of representation and validity, however, is highly contested. In this lecture, I will introduce different ways of generating silicon survey data, how both LLM-inherent and design-driven errors affect how well this data represents and measures public opinion, and how we can evaluate this. I will also highlight applications in political polling and how errors manifest in these applications.

November 5: Davud Rostam-Afschar

Beliefs about Bots: How Employers Plan for AI in White-Collar Work

We provide experimental evidence on how employers adjust expectations to automation risk in high-skill, white-collar work. Using a randomized information intervention among tax advisors in Germany, we show that firms systematically underestimate automatability. Information provision raises risk perceptions, especially for routine-intensive roles. Yet, it leaves short-run hiring plans unchanged. Instead, updated beliefs increase productivity and financial expectations with minor wage adjustments, implying within-firm inequality like limited rent-sharing. Employers also anticipate new tasks in legal tech, compliance, and AI interaction, and report higher training and adoption intentions.

November 12: Alexander Wenz

Questionnaire Development with Large Language Models

Developing good survey questions is often described as both an art and a science and is an area where survey researchers can particularly benefit from large language models (LLMs). In this lecture, I will present use cases of how LLMs can facilitate various aspects of questionnaire development. Trained on large amounts of human-generated text data, LLMs might serve as effective, yet easy-to-implement and inexpensive tools for designing and evaluating survey questionnaires as well as creating engaging survey invitations that are adapted to different target groups. I will showcase applications from survey practice, reflecting upon the benefits and limitations of LLM-based approaches.

November 19: Matthias Mehl

Towards the Wearable-based Measurement of Social Connection

How much time do we spend alone every day? How many minutes do we converse with others around us? And how many of our daily interactions are deeper exchanges versus just small talk? These seemingly trivial questions are important because they illustrate how little we know about our everyday social interactions. Over the last 25 years, I have developed and validated the Electronically Activated Recorder or EAR as an ambulatory assessment tool to track people's naturally occurring (acoustic) social lives. The EAR is a digital audio recorder that intermittently samples snippets of ambient sounds while participants go about their normal lives. The EAR allows studying how subtle yet objective aspects of people's daily social interactions matter psychologically. While much of my past research with the method relied on human annotation, recent advances in audio signal processing now begin to allow exploring the objective tracking of everyday social connection in automatic and privacy preserving ways. In that regard, we are developing the SocialBit as a wearable-based measure of minute-to-minute conversational activity. With the SocialBit, we can pursue the vision of "sensing" daily social life at scale and study the implications of everyday human social connection for mental and physical health.

November 26: Darshit Pandya, Dario Leanza

Appetite for Innovation: The Potential of Sensing Meals to Better Understand Eating in Social Contexts

Social eating plays a central role in food-related decision-making, yet it has traditionally been studied primarily in laboratory settings or through self-report measures. This feasibility study investigates whether smartphone sensors can be used to objectively and unobtrusively capture social processes during shared meals in real-world environments, based on the Dyadic Health Influence Model.
A total of 45 meal recordings were collected and analyzed using smartphone-based audio sensing. Features such as shared mealtime duration and participants' speaking shares were extracted as indicators of social engagement and interpersonal influence during meals. The results showed a 75.5% accuracy in detecting meal start and end times. Average deviations were relatively small, corresponding to 4.75% for meal start detection and -1.95% for meal end detection relative to the average meal duration of approximately 15.5 minutes.
Overall, the findings suggest that smartphone sensors are a promising tool for studying social meal dynamics in naturalistic settings. This approach enables non-intrusive monitoring of eating behavior and offers new opportunities to better understand how social interactions shape food-related behaviors in everyday life.

December 4: Ruben Bach

tbd