Contribution ID: **279**   Type: **Oral Presentation**

Addressing Measurement Error in Machine Learning-Assisted Social Science Modeling

Thursday 24 July 2025 10:40 (15 minutes)

With the advent of machine learning tools and large language models (LLMs), the collection of measurements related to social science constructs (e.g., personality traits, political attitudes, human values) has become easier, faster and more affordable. These measurements are subsequently used for modelling of societal and group processes that social scientists typically engage in, where inferences from samples to populations are also made. Valid modelling and inferences, however, requires high-quality measurements or at the very least, methods to deal with the presence of measurement error. Just like traditional questionnaire-based measurements, machine learning- and LLM-based measurements have been shown to suffer from validity and reliability issues.

While there is an abundance of research literature in dealing with measurement error, they focus on questionnaire-based measurement error. It is unclear yet how measurement issues arising from machine learning tools and LLMs should be handled in social science modelling research.

This study has two primary objectives. First, we review existing literature to identify practices for addressing machine learning- and LLM-related measurement error, both in computer science and in social sciences. Second, we synthesise these findings with existing measurement modelling literature to propose a practical framework for making valid inferences using machine learning- and LLM-based measurements in social sciences. By bridging the gap between modern machine prediction capabilities and social science inference requirements, our framework aims to enhance the reliability and validity of social science research outcomes in the era of machine learning and LLMs.

Primary authors: VAN KESTEREN, Erik-Jan; GARCÍA BERNARDO, Javier; FANG, Qixiang (Utrecht University)

Presenters: VAN KESTEREN, Erik-Jan; GARCÍA BERNARDO, Javier; FANG, Qixiang (Utrecht University)

Session Classification: Symposium : "Measurement and Machine Learning"