

Beyond 'Accuracy': AI vs. Humans as Raters

Thursday 24 July 2025 09:45 (15 minutes)

Oral presentation

Beyond 'Accuracy': AI vs. Humans as Raters

Author

Aaron Petrasch

Affiliation

University of Munich (LMU)

Abstract

Artificial intelligence is becoming increasingly prevalent in social science research, raising critical questions about its role as a complement or substitute for human raters or judges. While AI-based judgments offer new possibilities, their validity and comparability to human judgments still need to undergo careful examination. This talk presents a framework for evaluating AI as raters or judges, emphasizing the need for assessments that go beyond simple 'accuracy' (often measured as correlation with some criterion). First, I will introduce several psychometric methods to compare AI and human judgments systematically. Second, I will present an extension of the Brunswikian Lens model that enables the examination of which textual or visual cues (units of information) are used by humans versus AI in forming judgments. Drawing on empirical examples of text- and image-based evaluations, I will demonstrate how these methods reveal meaningful differences in how judgments are made. Ultimately, I argue that integrating AI into social science research requires at least the same level of methodological rigor as human-based evaluations, ensuring that AI-driven assessments are both valid and reliable.

Keywords

AI; judgment; lens model; cues

Primary author: PETRASCH, Aaron (University of Munich (LMU))

Session Classification: Session 14 : "Dynamic and temporal models in psychology"

Track Classification: Measurement: Measurement