Contribution ID: **238** Type: **Paper presentation**

Reliability of Generative AI in Argument Classification: Identifying Pro and Anti-Vaccination Stances on Reddit

Friday 19 September 2025 10:30 (15 minutes)

In this study, we evaluate the reliability and methodological implications of generative artificial intelligence (genAI) models, specifically ChatGPT, in classifying nuanced vaccination stances expressed in online debates. We analyzed 295 comments from three Reddit threads discussing vaccination using multiple measurements with different ChatGPT model versions (4o and 4.5). Each comment was contextually paired with its parent comment to capture argumentative stance, and analysis was enhanced by including a preliminary qualitative thematic evaluation. Our findings show moderate-to-strong consistency across measurements, with correlation coefficients ranging from 0.64 to 0.77. However, there is variability in interpreting nuanced arguments, such as attributing disease reduction solely to sanitation rather than vaccination, or distinguishing between rhetorical style and genuine argumentative extremism. These variations point to deeper interpretive ambiguities that current generative AI models handle inconsistently. We discuss both the opportunities and challenges of using generative AI for qualitative and quantitative content analysis, emphasizing the importance of careful prompt design and methodological transparency. Our findings contribute to advancing AI-assisted research methods in sociology and computational social science, showing pathways to enhance the reliability and interpretive precision of generative thematic analysis.

Authors: Dr RUGHINIS, Cosima (University of Bucharest); Dr VULPE, Simona-Nicoleta (University of Bucharest)

Co-author: FLAHERTY, Michael G. (Eckerd College)

Presenter: Dr VULPE, Simona-Nicoleta (University of Bucharest)

Session Classification: Social Aspects of Networking Environment Today

Track Classification: Social Aspects of Networking Environment Today