

Evaluating AI Tools for Mental Health and Crisis Support

Participants:

Kelly Zuromski (Chair), Gregory Kiar (Presenter), Robbie Torney (Presenter), Kelly Zuromski (Presenter), Peter Franz (Presenter), Matthew Nock (Moderator)

Symposium Synopsis:

Despite growing use of AI in mental health care, the field lacks standardized methods and accessible data for evaluating these tools. This symposium brings together four presentations to address this gap, with a particular focus on suicide risk, given the increasing use of AI chatbots for crisis support. The first presentation introduces a method for generating synthetic mental health datasets that preserve ecological validity while protecting patient privacy, helping to address data-sharing barriers that have limited AI testing. The second presents an empirical evaluation of how well large language models predict the outcomes of high-risk crisis counseling conversations compared to human clinicians, using real-world crisis line data. The third presents a qualitative, expert-driven evaluation framework using clinical red-teaming to assess suicide and self-harm safety at the product level in chatbots used by youth. The fourth translates these evaluation efforts into clinical guidance, describing a structured consensus process involving suicide prevention and AI experts to establish recommendations for chatbot responses to suicide risk disclosures. Together, these talks address the methodological infrastructure, empirical testing, qualitative assessment, and clinical guidance needed to thoroughly evaluate AI tools in psychiatric care.