

Title: Conversational AI in cancer genetic education: A critical review of assessment metrics and methodologies

Authors & Affiliations: Jennifer Scalia, PhDc, MS (Clemson University), Jessica Laprise, MS, CGC (Ambry Genetics), Jason Thrift, PhD, RN, CHSE (Clemson University), Christopher Farrell, PhD (Clemson University), Sara Sarasua, PhD, MSPH (Clemson University)

Background: Recent advances in artificial intelligence (AI) are transforming healthcare by integrating AI-powered chatbots into clinical practice and education. Chatbots, particularly in hereditary cancer risk assessment and genetic counseling, are establishing new standards for personalized patient care and educational support. However, current evaluation methods exhibit heterogeneity and lack systematic rigor, revealing a critical gap that needs attention to ensure the safe and effective delivery of these novel AI clinical tools.

Objective: This systematic review critically evaluates the metrics used to assess AI-driven chatbot platforms in oncogenetic risk evaluation and education. The study objective is to identify limitations in current methodologies and propose a cohesive framework that leverages AI's distinct capabilities to enhance healthcare education and patient outcomes.

Methods: A literature search was performed utilizing PubMed, Web of Science, and Engineering Village, following the PRISMA guidelines. Selected studies were analyzed for study design, chatbot functionality, and evaluation metrics. The extracted metrics were mapped onto the RE-AIM framework (Reach, Effectiveness, Adoption, Implementation, and Maintenance) to highlight evaluative gaps and provide a multifaceted perspective on the effectiveness of AI integration in healthcare education.

Results: The literature search yielded 684 citations, with 11 studies meeting the inclusion criteria. These studies, conducted in varied clinical settings with diverse methodologies, identified 104 metrics categorized into 16 categories (median=8 per study) across five domains: user experience, knowledge acquisition, outcomes and behaviors, emotional response, and technical performance. While user experience measures predominated, evaluation of knowledge acquisition was limited. The RE-AIM analysis also revealed challenges in data accuracy, algorithmic transparency, and user privacy protection.

Conclusion: The variability observed in evaluation metrics highlights the need for a standardized framework that integrates AI-specific performance indicators with traditional healthcare outcome measures. This integration is essential for ensuring the effectiveness and ethical soundness of AI-assisted clinical practice. Given the shortage of genetic counselors and high demand for real-time results, a systematic approach is vital to optimize resource use, improve patient outcomes, and advance precision medicine.