IHI Lucian Leape Institute | Expert Panel Report

Patient Safety and Artificial Intelligence

Considerations for Key Groups

Researchers

The entry of generative artificial intelligence (genAI) into health care creates significant concerns regarding validity and effectiveness. Researchers have the opportunity to build a research base on genAI in health care as well as bridge the gap between research and practice. The IHI Lucian Leape Institute offers the following recommendations for researchers:

- **Build and ensure equitable functionality:** A concern raised about genAI is bias in the dataset, which impacts how AI-based tools perform. Research into mitigating the problem of biased datasets and inaccurate outputs of AI-based tools for underserved and racialized patients needs to be a central theme in genAI research efforts. Research in the equitable distribution of and access to quality AI-based tools also needs to be prioritized.
- **Harness validated evidence to build trust and confidence:** Inaccuracies, including hallucinations (situations in which genAI fabricates results), diminish the trustworthiness of AI tools. Researchers can help improve the trustworthiness of AI-based tools and systems by ensuring that data and outcomes, including recommendations on diagnosis or treatment, are accurate and based on the latest evidence-based data. A fruitful line of research is to develop ways of conveying levels of confidence for genAI outputs, such that users could ascertain at a glance (e.g., with color coding) how confident an AI-based tool is in specific statements and recommendations. In addition, researchers can help test and validate each tool's ability to handle conflicting information and the credibility of evidence, and how well AI-human dyads work in real-life practice settings.

IHI Lucian Leape Institute Expert Panel Report on Patient Safety and AI

In January 2024, the IHI Lucian Leape Institute convened an expert panel to further explore the promise of generative artificial intelligence (genAI) and its potential risks for patient safety.

The panel reviewed the literature on AI and patient safety and engaged in a robust discussion that focused on three likely use cases for genAI in health care: documentation support, clinical decision support, and patient-facing chatbots.

The panel also reviewed considerations for key groups and provided specific recommendations and mitigation strategies for these audiences.

Visit ihi.org/LLISafetyAI

- **Prioritize people:** Identifying ways to safely and effectively deliver AI-based tool decision support for patients and clinicians is another area that needs substantial study, using human factors expertise and human-centered design. For instance, would a clinician want an AI-driven alert to pop up automatically with a diagnostic or management recommendation, or would they want to ask for help before genAI provides a recommendation? Human factors issues related to patient preference, alert fatigue, and the impact of technology on the patient-clinician relationships also deserve further study.

- **Advise on the guardrails:** Another critical area of research is developing approaches for ensuring effective human oversight of AI-based tools. For instance, if an AI-based tool like a chatbot is generating responses to patient queries and the process is designed to ensure that clinicians review the AI output to identify inaccuracies and edit prior to responding to the patient, then strategies will be needed to ensure that this human review actually (and meaningfully) occurs. Researchers can also help identify strategies to avoid clinical overreliance or dependence on AI tools and resultant deskilling, as well as study potential AI biases and ways to overcome them.