Incorporating AI into Healthcare

Position Statement

Artificial intelligence (AI) is actively transforming healthcare. The potential to improve clinical outcomes, reduce costs, and minimize healthcare inequities is immense; but so too is the potential for preventable harm. ECRI advises healthcare organizations to define goals with AI systems, proactively assess risk, and monitor AI performance. Ultimately, AI is an advanced tool to assist clinicians and healthcare staff, but human decision-making remains essential.

Summary

AI innovation is rapidly increasing, but AI systems are only as good as the model generated from the data on which they are trained, and the working clinical environment—or total system—in which they are implemented. Shortcomings in any area can lead to an inappropriate AI response that degrades (rather than improves) patient care or that exacerbates (rather than reduces) health inequities, which can lead to patient harm.

So how should healthcare organizations proceed?

— First, recognize that AI is not infallible. AI functionality should be systematically assessed before implementation. This assessment should consider the performance of the AI solution, the ways that the solution will impact other aspects of the healthcare organization, and the human and system factors associated with its use.

— Second, establish an AI implementation plan. This includes identifying the desired outcomes from implementing an AI solution, the patient population in which the AI solution will function, and the risk associated with implementing AI.

— Then, continuously monitor AI performance. This includes periodic assessment of current patient population relative to population at implementation, regular vendor updates per predetermined model control plans, and adverse event reporting and investigation associated with AI technologies.

Discussion

AI: What is it, and why now?

Artificial intelligence, or AI, broadly refers to the ability of computers to perform the type of generative analysis that is typically associated with a rational human being. Put in another way, AI is a machine that generates predictions based on internal calculations. AI has been present in healthcare for years, mostly in image-processing capabilities incorporated in x-ray, magnetic resonance, and other imaging technologies. Such applications boast a considerable clinical history of steady performance.

Recent advancements in computing power and software tools, however, have opened the door to an ever-expanding array of AI applications. These advancements are enabling rapid innovation in areas such as patient experience and management, clinical decision support, diagnostics, cardiovascular monitoring, oncology, medication delivery, and many more administrative and procedural applications.

Terms such as machine learning (ML), deep learning (DL), and foundational models are often used interchangeably with the term AI but refer to specific processes or capabilities. However, they all can be thought of as falling under the "AI" umbrella. "Augmented intelligence" is another term that is sometimes used. This term offers an alternative conceptualization that focuses on AI's assistive role, emphasizing the fact that AI design enhances, rather than replaces, human intelligence.

How is AI being used in healthcare?

Healthcare applications for which AI is currently being used include:

— Imaging applications—such as image reconstruction, de-noising, and segmentation, and labeling the region of interest

— Clinical decision-making support, often embedded within an electronic medical record (EMR)

— Interpretive functionality incorporated into diagnostic and therapeutic medical devices (e.g., point-of-care ultrasound, estimated coronary flow reserve, risk-based assessments using ECG, sepsis monitoring, medication adherence and outcome tracking)

— Process optimization—for example, orchestration systems across the whole organization

— Chatbots—used, for example, for prescription refills or triage

— Scheduling—for example, interacting with the patient to find the best time for a visit or to remind them of appointments; future applications might even arrange transportation

— Medical notes generation

— Applications in cardiovascular, neurology, hematology, gastroenterology, urology, and other specialties.

Some healthcare AI applications have reached a mature stage of development, and some are still in the early stages of use—but many more are just now being implemented or are on the horizon. The American Medical Association (AMA) has identified key AI use cases being used by physicians today, as well as use cases that are likely to increase in scale and sophistication in the future[1].

What are the risks, and how can they be managed?

AI-generated results derive from the model used and the data on which it was trained. Shortcomings in either of these can cause an AI solution to perform worse than advertised, to provide misleading or inappropriate results, or to create a false sense of security in clinical decision support and other clinical applications. In such circumstances, rather than improving patient care, AI could instead contribute to harm or exacerbate health inequities.

For instance, certain AI models have been reported to demonstrate bias regarding demographics such as race, socioeconomic status, or geographic location. As noted, biased data will result in biased models. Other risks, outlined in AMA's Future of Health report, include limited explainability, "hallucinations" (i.e., AI responses that are false or misleading), and privacy and security concerns.

To help manage risk, clinical teams and users should consider the following before implementing AI in clinical practice:

— **Whether the AI system has regulatory clearance or approval.** Recognize, however, that regulatory clearance does not ensure that models are generalizable to all clinical practices. Furthermore, some AI applications that don't require regulatory clearance (i.e., the applications aren't defined as medical devices) can nevertheless impact patient care.

— **Explainability.** The AI application should provide appropriate and clear information about a device (that could impact risks and patient outcomes) that can be communicated to all stakeholders.

— **The AI application's risk profile.** The International Medical Device Regulators Forum (IMDRF) risk framework[2] considers two major factors to provide a description of the intended use: (1) the significance to the healthcare decision of the information provided (i.e., to treat or diagnose, to drive clinical management, or to inform clinical management), and (2) the state of the healthcare situation or condition, which identifies the intended user, disease, or condition (i.e., critical; serious; or non-serious healthcare situations).

Four risk categories, from lowest (I) to highest (IV), reflect the risk associated with the clinical situation and device use. This risk profile can help guide decisions regarding the level of independent evaluation appropriate for the AI solution.

[1] American Medical Association. Future of Health: The Emerging Landscape of Augmented Intelligence in Health Care. Feb 2024.

[2] International Medical Device Risk Framework (IMDRF) Software as a Medical Device (SaMD) framework.

State of healthcare situation or condition	Significance of information provided by SaMD to healthcare decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

Figure 1. SaMD IMDRF risk categorization

Doesn't regulatory clearance provide assurance of AI safety?

In their pursuit of medical device safety, regulatory bodies across the globe are developing processes to assess the safety of AI applications in healthcare.

— In the US, FDA has cleared several types of devices that use AI. Radiology applications are the largest category, accounting for 76% (671 out of 882) of AI-enabled cleared devices as of May 2024. Other categories of cleared devices include cardiovascular (10%), neurology (3%), hematology (1.9%), gastroenterology-urology (1.5%), and anesthesiology (1%).

— In the UK, the Medicines and Healthcare products Regulatory Agency (MHRA) is moving forward its AI-Airlock—a "regulatory sandbox" that will provide a regulator-monitored virtual area for developers to generate robust evidence for their advanced technologies.

— Canada has developed a draft pre-market guidance for machine learning-enabled medical devices.

However, regulating AI is a daunting challenge; and even if this challenge is met, limitations intrinsic to the approval process mean that clearance does not necessarily translate into safety. As noted, regulatory clearance does not ensure that AI models are generalizable to all clinical practices. Furthermore:

— The regulatory premarket testing required for clearance can vary depending on the nature of the AI application. In some cases, such testing can be very limited.

— After clearance, the effectiveness of an AI solution should be monitored by regulatory post-market surveillance. However, this has variable effectiveness, in that AI's role in any adverse event may not be apparent and known incidents may not be reported and investigated reliably.

— AI has the potential to continuously learn, which could lead to the rapid update of existing AI solutions or rapid development of new solutions. In other words: AI is evolving at a faster pace than has occurred with traditional medical devices, and regulatory processes need to evolve to address that reality. Premarket approval looks at the current state of AI, but it will not capture ongoing model learning. To address that need, global regulatory authorities defined a new framework for online-learning devices. A Post-Market Change Control Plan (PCCP) is intended to characterize a device and its bounds,

the intended changes to the ML system, the protocol for change management, and the change impacts.

— Looking into the future, foundational models that can be used in several clinical applications will add even more complexity to the clearance process.

Perhaps more significantly, many AI applications that can impact patient care do not strictly meet the definition of a medical device, and thus would not require regulatory clearance. For example, an AI-based orchestrator solution would not require clearance as a medical device since the solution assists with nonclinical processes. But imagine a case where that solution filters out certain patients, either due to a programming error or for nonclinical (e.g., demographic) reasons. The result could be patients not receiving needed care.

How can AI functionality be assessed?

AI-related decisions don't necessarily revolve around the question "Should we incorporate AI?" More commonly, organizations will need to assess the appropriateness or value of purchasing a device or other solution that relies on AI.

For such decisions, ECRI recommends a "total systems safety" approach. This approach evaluates AI solutions in the context of the complexity of the total system—a construct that includes people, tools and technology, tasks and processes, the physical environment, organizational structure and policies, and the external environment. The goal should be to understand the ways that different parts of the system interact to facilitate safe work—or the ways they interact to create barriers to safety.

Ways that ECRI pursues a total systems approach include:

— Applying human factors engineering methods to assess the interactions between users, the use environment, and the AI-based tools to enable efficient and safe use of these tools.

— Testing devices that incorporate AI capabilities to assess how well the product performs its intended function, and how safe the product is when used in its intended environment.

— Conducting systematic clinical evidence assessments (CEAs) of AI-based applications with a focus on clinical, patient-oriented outcomes.

— Engineering and clinical analyses of adverse events reported to ECRI.

How can AI performance be monitored?

The performance of an AI solution can change over time; thus healthcare organizations need to establish policies—and define a process—for monitoring the AI solution after it has been implemented.

In general, current technologies are closed; they have been trained by the developer and will not "learn" over time. In contrast, continuous learning models could encounter novel

inputs that influence the model to generate predictions slightly different from the defined target.

When incorporating an AI model, it is necessary to monitor:

— The model's "brittleness"—that is, how it responds over time to conditions different from the original training data (e.g., different patient populations, changes in clinical practices)

— If the model is consistently predicting the defined target

— If the performance of the model is affected by:

  ▪ Data drift—that is, changes to the input data, including changes in patient demographics

  ▪ Drift from the intended manner use—similar to "off-label" use of medications

  ▪ Model drift—that is, changes in performance in the absence adaptive learning. Performance will decline if the model cannot "learn" from small variations or novel inputs.

Additionally, organizations should regularly assess safety and clinical outcomes related to practices impacted by AI, any impact of the AI solution on healthcare disparities, and the user experience for both the patient and the healthcare worker.

Finally, AI safety would benefit from a robust post-market surveillance program and event reporting system. Event reporting can help healthcare organizations—and the healthcare community at large—detect safety gaps in a timely manner so that harm can be prevented.

ECRI Recommendations

For organizations that decide to adopt AI solutions, a key early step will be to define policies for the oversight of any AI solutions. Policies should address issues such as how to manage risks, how to monitor performance over time, and how to ensure that the system is being implemented ethically, considering the five pillars of AI ethics: Fairness, Robustness, Explainability, Transparency, and Privacy.[3]

ECRI recommends the following steps for AI implementation and governance:

1. Define the scope. This will involve assessing the process that requires improvement, and defining what it is that the AI solution is intended to predict. Ask: Do we want AI for this process? (It could be that AI is not required or even desirable.) And ask: What is the goal? Then define metrics that will assess progress toward the goal(s).

2. Educate management and users on the AI solution's capabilities, as well as its limitations.

3. Maintain stakeholder alignment. Constant communication between management and everybody involved in the AI implementation is crucial to maintain stakeholder alignment.

4. Assess data preparation needs.

   a. Don't expect AI solutions to be "plug-and-play." Considerable up-front effort may be required to get your data into a format that can be used by the AI application.

   b. Disparate data formatting within an organization is a major challenge when implementing a tool that uses the organization's data. Clinical informatics oversight may need to be implemented. At a minimum, data oversight should include an audit to ensure local data is mapped into the AI data structure in a compatible and congruent manner across all domains.

5. Establish a governance program.

   a. Create a registry of the AI solutions integrated into the institution's information system.

   b. Define the level of autonomy of the model; this will define the level of monitoring that is needed: the more autonomy, the more monitoring that is needed.

   c. Establish an oversight committee that will perform governance processes.

6. Assess AI integration considerations. Some implementations are seamless, with the AI functionality running behind the scenes with little to no interaction from the user. Others will require reformatting of the information that is input into the model. Processes will need to be established for managing updates and patches, which can change the model performance, and for assessing model performance over time.

7. Define a process for reporting potential adverse events. Organizations should:

   a. Implement a robust reporting system for AI-related medical incidents, errors, and adverse events. An institution's current reporting system may not be set up to capture AI-associated concerns.

   b. Educate users about how to identify incidents that could be attributable to AI functionality, and encourage them to report any suspected incidents, errors, or adverse events.

   c. Regularly monitor event reports and assess trends to ensure timely interventions.

Learn more: **www.ecri.org**

[3] IBM SkillsBuild. Artificial Intelligence Fundamentals.