AI-assessment: Towards Assessment As a Sociotechnical System for Learning
Alison Lentz, J. Oliver Siy, PhD, and Carol Carraccio, MD, MA

Abstract

Two decades ago, the advent of competency-based medical education (CBME) marked a paradigm shift in assessment. Now, medical education is on the cusp of another transformation driven by advances in the field of artificial intelligence (AI). In this article, the authors explore the potential value of AI in advancing CBME and entrustable professional activities by shifting the focus of education from assessment of learning to assessment for learning. The thoughtful integration of AI technologies in observation is proposed to aid in restructuring our current system around the goal of assessment for learning by creating continuous, tight feedback loops that were not before possible. The authors argued that this personalized and less judgmental relationship between learner and machine could shift today’s dominating mindset on grades and performance to one of growth and mastery learning that leads to expertise. However, because AI is neither objective nor value free, the authors stress the need for continuous co-production and evaluation of the technology with geographically and culturally diverse stakeholders to define desired behavior of the machine and assess its performance.

Our society is on the cusp of an enormous shift, both inside and outside of medicine, driven by advances in the field of artificial intelligence (AI). AI can be broadly used to describe any machine that is able to perceive, learn from, and make predictions within a given context. It is clear that AI will transform most industries and professions, and it is now our responsibility to engage with this technology to ensure a positive change in medical education, assessment, and ultimately patient outcomes. Schuwirth and Van der Vleuten 1 recently discussed 3 eras of assessment over the past 5 decades as being focused on: (1) measurement, (2) judgment, and (3) system. We use this opportunity to present an evolution of the third era to focus on assessment as a sociotechnical system where “the social and the technical are brought together and treated as interdependent aspects of a work system.” Using this lens, we consider the role of AI as a collaborator amongst learners, faculty, patients, and the clinical learning environment.

Since competency-based medical education (CBME) was introduced at the cusp of the 21st century, assessment has been and continues to be our greatest challenge. The introduction of entrustable professional activities (EPAs), the important and routine health care tasks that define a profession, is shifting the focus to what really matters in assessment: ensuring learners deliver safe and effective care to patients. 2 In this context, AI will impact both what is assessed—the activities that define a profession—and how those activities are assessed—through testing knowledge and observing workplace performance. Regarding what is assessed, the impact of AI on the role of the physician has been discussed elsewhere, 3 although it is important to briefly acknowledge the need for increasingly aligning curricular elements with the “art of medicine.” Uniquely human traits, such as empathy, and the ability to creatively curate and synthesize knowledge, such as translating machine diagnoses and prognoses into patient care plans, will become increasingly important over the next decade. 4

We believe the impact of AI on how activities are assessed presents more meaningful, immediate opportunities to aid in advancing CBME and EPAs by shifting the focus of education from assessment of learning to assessment for learning. 5 The former emphasizes summative assessment—that is, the grade without any feedback or opportunity to improve. The latter prioritizes directly observing learners in the workplace to provide ongoing constructive feedback and coaching with the goal of improvement. Although assessment for learning is the goal, faculty’s responsibilities tend to compete with and generally lose to an ever-increasing patient volume, the academic currency of research grants, and other obligations. The result is what we are currently experiencing in medical education: a system trapped in its past structure, despite the evolution of the world around it.

The thoughtful integration of AI technologies in observation can aid in restructuring our current system around the goal of assessment for learning—creating continuous, tight feedback loops that weren’t previously possible due to fixed constraints of time, money, and human bandwidth. At the forefront of the latest developments in AI are real-time multimodal perceptual capabilities (e.g., continuous processing of raw, audiovisual sensor data to understand human interactions and activity) that are able to run efficiently and locally on devices such as smartphones and wearables. 6–8 This approach ensures patient and learner privacy because the machine can recognize behavioral patterns of the learner in a secure, ephemeral loop, thus eliminating the need to store raw data or send raw data elsewhere to be processed.

For example, a machine can be trained to understand patterns in speech and interpersonal communication, such as the
time spent talking, interruptions, changes in tone of voice, or number of questions asked without sharing or retaining any logs of the raw conversation. A trainee could receive specific and nuanced feedback along these dimensions directly after seeing a patient, or even during an interaction through the use of subtle cues like haptic feedback. Perhaps a trainee has not paused to ask the patient for questions, and this is an area they seek to improve; a discrete vibration from a smartwatch could bring this to their attention in the moment. A system such as this would promote learner agency, as the trainee could decide when practice yields sufficient improvement to call in a faculty assessor for further feedback and guidance. Limited faculty time would be used wisely and more efficiently in assessing learners’ overall skill levels and judging their ability to advance. This system both supports formative assessment for learning and provides better evidence of a learner’s true capabilities for ultimate summative decision making.

Validation will be a critical step in leveraging AI to enhance learner assessment. Using the above example, would an AI capability that provides real-time feedback on how often trainees interrupt their patients decrease future interruptions? Furthermore, would decreases in trainee interruptions improve care and the patient experience? Answering these questions would involve tracking and providing feedback to trainees to analyze whether interruptions decrease over time and how patient interviews to understand the impact of lessened interruptions on their experience. This example represents just one of the many signals that could be used in combination to improve holistic communication skills. To close the loop, faculty must also learn from trainees about the types of feedback that contributed most to their improvement. Implementing and scaling AI in the clinical environment should only be considered if and when programs can provide enough evidence to support the efficacy of the application.

We hypothesize that this personalized and less judgmental relationship between learner and machine could shift today’s dominating mindset on grades and performance to one of growth and mastery learning that leads to expertise. This shift is important for practitioners across the entire continuum. Beyond the period of training, preservation of competence requires ongoing practice, and advancing from competent to expert requires deliberate practice. In medicine, this practice involves repetition, reflection, and refinement of tasks over time with the goal of improving some aspect of care. Feedback and perseverance are critical to deliberate practice, and the latter is critical to mastery learning.

Although AI holds much promise for shifting to a model of assessment for learning, some caution is warranted. AI can be trained to understand behaviors like speech patterns; however, the judgment the machine places upon a particular behavior is neither objective nor value free, but rather a reflection of the creators’ choices and values. As a result, it is imperative we engage in continuous co-production and evaluation of the technology with and by end-users. Despite this, some caution is warranted. AI is a former pediatrician, clinician educator, program director, and researcher with a focus on medical education.

References