

Artificial Intelligence: Beginning to Understand Implications for Education and Assessment

With recent advances in generative artificial intelligence (AI), conversations about the impact of new AI technologies on various industries has proliferated in both public and private spaces (print and broadcast journalism, opinion pages, social media, and industry conferences, as well as boardrooms and organizational strategy meetings). Within education, many commentators predict AI will have significant impacts on teaching and learning broadly and assessments specifically. Some recent examples of media exploring the potential for AI to impact teaching and learning include articles in [Forbes](#), [The Minnesota Daily](#), and [The New York Times](#)¹. Outside education, a paper in [McKinsey](#) suggests information all CEOs should know about generative AI, and [Open Mind BBVA](#) questions what is at stake for workers due to AI advancements.

While it's too early to gauge with certainty the impact that new generative AI technology and large language models (LLM) may have on large-scale assessment, it is not too soon for the Governing Board to learn more about these emerging technologies, understand how they are currently being deployed in the field of assessment, and consider longer-term impacts they may have for the future of NAEP.

During the May 2023 Quarterly Board meeting multiple Board members expressed interest in starting this exploration, and the 2023 August Quarterly Board meeting will provide a first opportunity to do so, with two separate sessions planned on this topic:

- 1) **Wednesday, August 2nd**: A keynote presentation will be held from 4:30 – 6:00 pm. Kristen DiCerbo, the Chief Learning Officer from Khan Academy, will share exciting innovations for student learning using generative AI. The objective for this session is for Board members to engage with a leader in the field of education using AI in innovate ways to enhance teaching and learning.
- 2) **Friday, August 4th**: A plenary session will be held from 12:45 – 2:15 pm with a more specific focus on understanding how AI is currently being used in various large-scale assessments and on potential new uses and impacts in the near-term. It will include the following presentations:
 - Peggy Carr and Emmanuel Sikali from NCES will share plans for using AI for various aspects of NAEP. Some of these activities are ongoing, and others are in early stages of exploration.
 - Susan Lottridge of Cambium Assessment will describe the current state of the field regarding AI use for automated scoring, including implications for NAEP.
 - Matthias Von Davier of the TIMSS & PIRLS International Study Center at Boston College, will share plans for using AI in the TIMSS and PIRLS international assessments.

¹ New York Times articles may require a subscription. Please let Board staff know if you are unable to access.

Biographies for the speakers are included at the end of this attachment.

What is Artificial Intelligence?

Before jumping into these discussions, it is important for all members to have a shared foundational understanding of key terms. Each of these definitions include links to websites where you can learn more, if desired.

Artificial intelligence (AI) is defined by the [Encyclopedia Britannica](#) is the “ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.”

Generative AI is defined by the [Government Accountability Office](#), in their brief report describing opportunities, risks, and policy considerations of AI, as “a technology that can create content, including text, images, audio, or video, when prompted by a user.”

Another common term present in AI discussions is natural language processing.

Natural language processing (NLP) is described by [IBM](#) as “the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.”

Though AI has been a hot topic in recent months, it is not a brand-new phenomenon. In fact, an online search for “history of AI” results in various publications that span the past couple decades and refer to processes that have been used since the onset of computers. In [The History of Artificial Intelligence](#), published by the University of Washington in 2006, it is noted that the term Artificial Intelligence was first coined in 1956. What has changed in recent years and months is the pace of development and the increased interest in using AI in varied ways, including in education and assessment.

The public conversation has been accelerated in recent months due the introduction of Chat GPT (Generative Pre-Trained Transformer) released by OpenAI in November 2022. Chat GPT is a chat box, a software application that aims to mimic human conversation through text or voice interactions, typically online. In less than a year, ChatGPT has had more than 100 million users and been updated multiple times. The system currently offers a free tier available to all users as well as a more enhanced version for a fee. Chat GPT has attracted attention due to its ability to create resumes, write essays, explain complex topics, summarize long documents, create poetry and various other compositions. and have human-like conversations.

At the same time, concerns have been raised about copyright, privacy, misuse, bias and transparency associated with ChatGPT specifically and AI generally. Though ChatGPT is a powerful tool that undoubtedly has immense potential, there are limitations to it and

other generative AI technologies that merit serious consideration, including that they may not always provide correct answers, may have limited knowledge of certain events, and may perpetuate bias and encourage plagiarism. An article published by [Forbes](#) (March, 2023) highlights different categories of bias that may be introduced by using ChatGPT and other similar chat boxes. As reported by [Brookings](#) (May, 2023), there is bipartisan agreement for the need to regulate AI given its potential widespread use across sectors, and the U.S. Senate has held recent hearings on the topic, where concerns were expressed about its possible use for mass manipulation and making high-stakes decisions through processes not fully transparent or understood.

AI and Assessment

There are various aspects of assessment that are currently impacted or may be impacted by advances in AI in the near-term. Particularly, AI may allow procedures currently completed by humans to be partially or fully completed by computers. Some examples of the processes that may be impacted include:

- Item/Content Generation
- Test Assembly
- Scoring of Constructed-response Items
- Reporting and Analysis

You will hear more about how NAEP and international assessments are approaching these processes during the Friday plenary session.

Another area for which AI tools have the potential to impact educational assessment is their incorporation into teaching and learning. For example, generative AI tools have the potential to change the way students engage with writing. They may also allow for increased real time personalized feedback on student work. These changes will need to be monitored to ensure proper adjustments to NAEP frameworks as necessary over time, if/when these changes fundamentally impact what and how students are learning and what constructs are most important for assessments to measure.

More Information

There have been many reports and articles exploring the potential impacts of AI in education published in recent months. We encourage Board members to review the following resources to learn more about the current landscape:

- The Center for Assessment released the blog, [The Opportunities and Risks of Artificial Intelligence in K-12 Assessment and Accountability](#), to highlight potential implications of AI in education assessment.
- The New York Times provides an interactive tool to illustrate AI used for writing: [Watch an AI Learn to Write using Nothing but...](#)

- The U.S. Department of Education Office of Educational Technology's new policy report on how AI may impact teaching and learning: [Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations.](#)
- Robin Lake of the Center on Reinventing Public Education (CRPE) shares insights on AI's potential to impact teaching and learning in her article: ['This changes everything:' AI is about to upend teaching and learning – Center on Reinventing Public Education.](#)

Invited Speakers on Artificial Intelligence in Education

August 2 Keynote: Practical Implications of AI for Teaching and Learning



Dr. Kristen DiCerbo is the Chief Learning Officer at Khan Academy, a nonprofit dedicated to providing a free world class education to anyone, anywhere. In this role, she is responsible for developing and implementing a research-based teaching and learning strategy for Khan Academy's offerings in order to improve student and teacher engagement and outcomes. She leads the content and product management teams. Dr. DiCerbo's work has consistently been focused on embedding what we know from education research about how people learn into digital learning experiences. Prior to her role at Khan Academy, she was Vice-President of Learning Research and Design at Pearson, served as a research scientist supporting teaching and learning in the Cisco Networking Academies, and worked as a school psychologist in an Arizona school district. She publishes and speaks to a variety of audiences about educational technology, learning, and learning science research. Kristen received her Bachelor's Degree from Hamilton College and Master's Degree and Ph.D. in Educational Psychology at Arizona State University.

August 4 Plenary: Opportunities and Challenge for Artificial Intelligence and Large-Scale Assessment



Dr. Susan (Sue) Lottridge is a Chief Scientist of Natural Language Applications at Cambium Assessment, Inc. She has a Ph.D. in Assessment and Measurement from James Madison University and Masters' degrees in Mathematics and Computer Science from the University of Wisconsin – Madison. In this role, she leads CAI's machine learning and scoring team on the research, development, and operation of CAI's automated scoring and feedback software. Dr. Lottridge has worked in automated scoring for fifteen years and has contributed to the design, research, and use of multiple automated scoring engines including equation scoring, essay scoring, short answer scoring, speech scoring, crisis alert detection, and essay feedback.



Matthias von Davier is the Executive Director at TIMSS & PIRLS International Study Center and Monan Professor in Education at Boston College. Dr. von Davier's research focuses on developing psychometric models for analyzing data from complex item and respondent samples and integrating diagnostic procedures into these methods. His areas of expertise include topics such as item response theory, latent class analysis, classification and mixture distribution models, diagnostic models, computational statistics, person-fit, item-fit, and model checking, as well as hierarchical extensions of models for categorical data analysis, and the analytical methodologies used in large scale educational surveys, as well as methods for assessment automation and text and sequence data analysis using machine learning, natural language processing, and artificial intelligence.