

University of Florida's AI in Education Workshop: A Perspective from Academia and Learning Industry

Seyedahmad Rahimi, Salah Esmaeiligoujar

December 2024



Rapid Community Reports

Workshop Outcomes

University of Florida's AI in Education Workshop: A Perspective from Academia and Learning Industry

How will artificial intelligence (AI) transform our learning environments and address the challenges we face in education?

Authors

Seyedahmad Rahimi, University of Florida, srahimi@ufl.edu

Salah Esmaeiligoujar, University of Florida, s.esmaeiligoujar@ufl.edu

Abstract

The University of Florida's AI in Education workshop, conducted in April 2023, brought together three NSF-funded AI institutes, three learning companies, and scholars in the fields of Computer Science, Learning Sciences, Instructional Design, and Computational Psychometrics. This online workshop attracted about 750 participants from 42 countries.

Keywords

AI, Artificial Intelligence, AI in Education, Learning, Learning Sciences, Assessment

Overview

Artificial intelligence in education dates back to the 1950s with the introduction of Intelligent Tutoring Systems (ITSs)—computer programs designed to provide personalized instruction tailored to individual learners' needs and responses (Buchanan, 2005; Chassignol et al., 2018; Chiu et al., 2023). Advances in AI technologies beyond academia (e.g., ChatGPT by Open AI) are rapidly enabling AI innovations in education. The emergence of generative AI, such as ChatGPT, presents both opportunities and challenges to conventional education systems. Therefore, we need rigorous studies to evaluate the efficacy of various AI tools and applications across diverse learning contexts and populations. It is also essential to investigate best practices for integrating AI in pedagogically

sound ways and fostering necessary competencies in both students and teachers to effectively leverage these technologies.

The growing importance of AI in education has led organizations such as the National Science Foundation (NSF) to fund multi-site AI research institutes focused on cutting-edge research and development in educational settings. Moreover, major learning companies have been making progress in using advanced AI in recent years. The launch of Chat GPT in particular has sparked global interest and discussion around AI. It was around the advent of these new AI technologies in the fall 2022 that the idea for a workshop started to take shape, with the goal of exploring the potential impacts of such technologies on education. The University of Florida's AI in Education workshop was conducted in April 2023 and brought together three NSF-funded AI institutes, three learning companies, and numerous AI scholars in the fields of Computer Science/AI, Learning Sciences and Instructional Design, and Computational Psychometrics to discuss the latest developments in the field of artificial intelligence.

Workshop Attendees



The NSF-funded AI research institutes that were invited to this workshop included:

- The AI Institute for Student-AI Teaming (iSAT), an interdisciplinary research community dedicated to transforming classrooms into more effective, engaging, and equitable learning environments focusing on collaboration;
- The EngageAI Institute, an institute that supports and extends the capabilities of teachers and students with AI. This institute's mission is to produce transformative advances in STEM teaching and learning with AI-driven narrative-centered learning environments;
- The AI Institute for Adult Learning and Online Education (AI-ALOE), which develops an AI-based transformative model for online adult learning.

Each of these institutes received approximately \$20M from the U.S. National Science Foundation.

The three well-known learning companies that use AI in their products included:

- Duolingo, a language learning company;
- Khan Academy, a personalized learning company;
- ASSISTments, a non-profit learning company which creates responsive online technology that is teacher-paced and evidence-based for subjects such as mathematics.

Finally, the workshop invited 10 other top scientists whose work has been nationally and internationally recognized to discuss important topics around AI in education. This workshop was conducted online on April 6, 2023 and attracted about 750 participants from 42 countries.

UF's AI in Education Workshop

Number of Participants
1 387

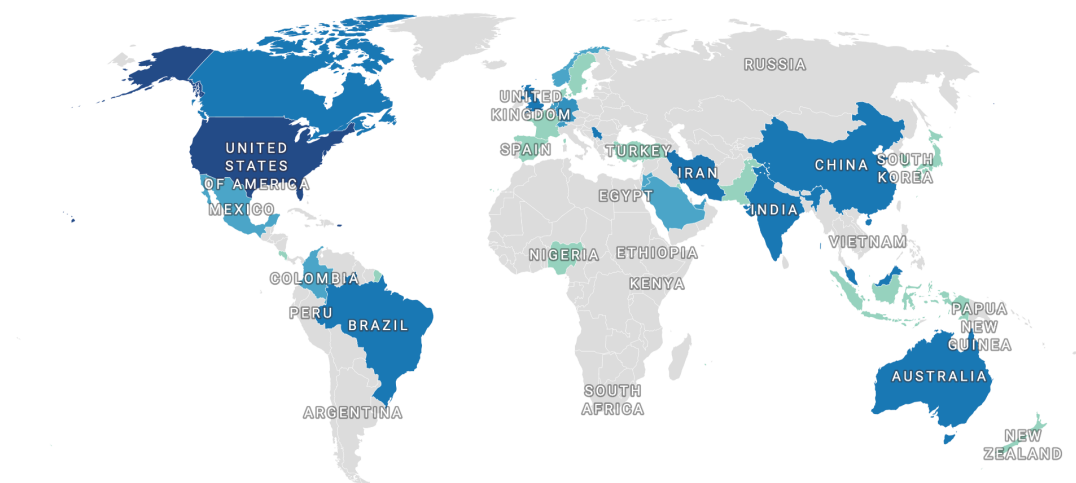


Figure 1. The distribution of participants participating in the workshop

The presenters were Chris Dede, James Lester, Neile Heffernan, Kristen DiCerbo, Peter Foltz, Alina A von Davier; the contributors were Sidney D'Mello, Valerie Shute, Danielle S. McNamara, Ryan Baker, George Siemens, Diego Zapata-Rivera, Kristy Boyer, Victor Lee, Lin Lin Lipsmeyer, Richard Golden, Russell Almond, and Rene Kizilcec. These highly respected scholars come from diverse backgrounds including AI in education, learning sciences, computer science, instructional design, computational psychometrics, and AI education.

The workshop had two main parts: (1) presentations, where each AI institute and learning company presented their latest work and discussed their findings, and (2) six virtual panel discussions, where scholars discussed specific topics in AI in education. To elicit the important themes for the panels, we asked the invited scholars to list the 10 most important issues they think we need to discuss in the workshop panels. Then, we analyzed the qualitative responses. The following five themes emerged:

1. Partnership of Humans and AI / Intelligence Augmented (IA)
2. AI Education from K-12 to Adult Learning
3. Fundamental Research, Development, and Theory Building
4. AI in Assessment and Improving Learning
5. Policy and Challenges related to AI in Education

Workshop Structure



In this section, we first briefly describe the six presentations by the three AI institutes and the three learning companies.

Institute for Student-AI Teaming (iSAT)

The [Institute for Student-AI Teaming \(iSAT\)](#), represented by co-PI Peter Foltz and PI Sidney D'Melloe at the workshop, presented on the topic of “Envisioning Next Generation Human-AI Partnership at iSAT” (see Figure 2). iSAT brings together researchers, graduate students, and postdocs from 14 research areas, from nine major universities in the United States (e.g., University of Colorado Boulder, Colorado State University, University of California, Berkeley). This AI institute also partners with various K-12 and industry network affiliates. The key question that the iSAT team is investigating is how to promote deep conceptual learning via rich socio-collaborative learning experiences for all students. iSAT systematically focuses on several key areas to provide measurable learning outcomes. These include domain knowledge acquisition, fostering 21st century skills such as collaborative problem solving and critical thinking, and enhancing motivation and affect by building students' interest, curiosity, and self-efficacy in AI, AI education, and STEM. The institute also focuses on teaching methods that include encouraging curiosity-driven exploration and meaningful classroom discussions. It also addresses the ethical and societal dimensions of AI, helping students understand the nature, behavior, and impact of AI systems in classroom settings. Finally, it aims to empower students to actively contribute to an AI-driven workforce.



Figure 2. iSAT vision of AI in classrooms

Through these efforts, the team came up with new metaphors for how they refer to an AI partner: community builder (i.e., help students to develop trusting collaborative relationships with each other, their teacher, and the AI partner); augments (i.e., support teachers in classroom management using the data from each group); and co-pilot (i.e., support and facilitate students' work by providing tailored guidance). These AI metaphors have been created in iSAT and tested in various projects for different curriculum.

EngageAI Institute

James Lester, PI, presented the work he and his colleagues have been doing at the [Engage AI Institute](#), a multi-site institute involving North Carolina State University, University of North Carolina, Vanderbilt University, Indiana University, and Digital Promise.

The EngageAI Institute focuses on two major areas of work. The first is use-inspired AI research, which includes developing storytelling-based approaches, interactive virtual agents, and tools to analyze various types of student interactions, such as written responses and video observations. The second area focuses on essential AI research, including improving how computers process language, interpret images, and learn from data. Together, these areas aim to create engaging, ethical, and effective AI-driven learning environments. The institute focuses on creating AI-driven learning environments guided by ethical principles, improving AI methods to ensure fairness and transparency, and expanding the educational impact both in and outside schools while promoting equity and inclusion.

The AI-driven, narrative-centered, learning environments in EngageAI Institute include narrative-centered learning (i.e., any learning experience that can immerse students in stories), embodied conversational agents that can invite students to participate in the story, and multimodal learning analytics. These environments create learning experiences for students that are effective (targeting science and computer science) and engaging in formal (e.g., classrooms) and informal (e.g., museums) settings.

For example, the EngageAI Institute has developed methods using reinforcement learning—a type of machine learning that improves through human feedback—to create and refine narratives. These narratives are designed to adapt to students' needs by learning from strategies used to generate them (Figure 3). This system also includes AI tools that can create stories and learning environments, including their settings, plots, and characters. This AI-based system can generate a learning environment based on students' interest in an imaginary storyworld or in a hospital. The institute also used large language models (LLMs) for empowering the embodied conversational agents. Specifically, LLMs are used to support conversation between the students and the agents and allow the agents to take different roles (e.g., pedagogical, companion, or teacher assistant).

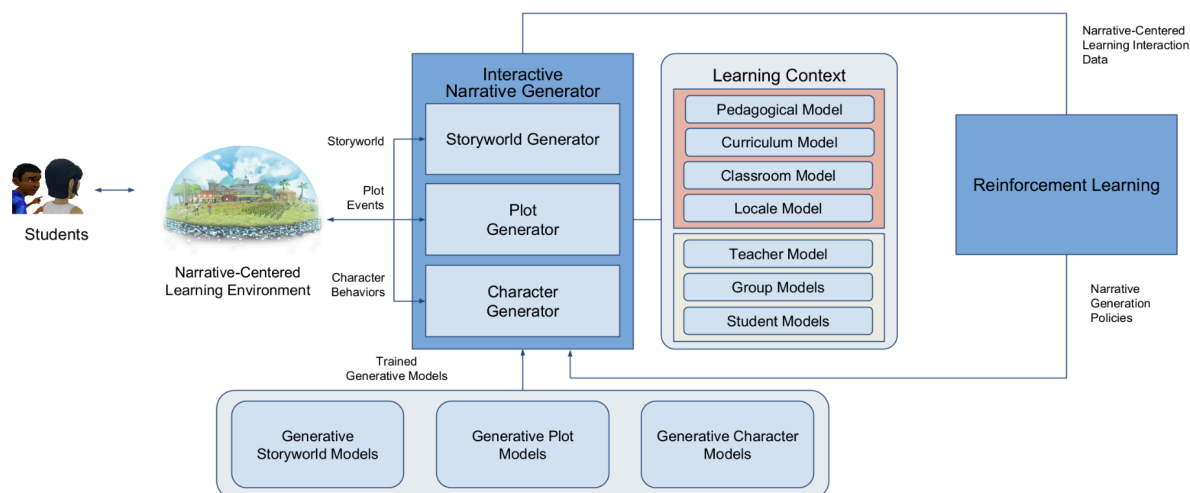


Figure 3. Reinforcement-Learning-based Narrative-Centered Learning in Engage AI Institute

The AI Institute for Adult Learning and Online Education (AI-ALOE)

Chris Dede, Co-PI, shared his insights about the initiatives at the [AI-ALOE \(Artificial Intelligence for Adult Learning and Online Education\) Institute](#). Dede pointed out that AI-ALOE differs from the other institutes and learning companies because of its audience—i.e., adults. AI-ALOE stands out in its approach to addressing the dynamic needs of adult learners in a rapidly changing world beyond formal education.

A critical aspect of the AI-ALOE's approach is recognizing that adult learners differ significantly from children. For instance, the diversity in developmental and experiential stages demands a unique pedagogical approach for adult learners, distinct from traditional pre-college and college-level education. Analyzing what AI agents we need for adult learners, Dede indicates that we need AI agents that can help with a learning episode, a course, a whole program, and eventually as a career coach in a life-long learning approach for a series of careers one might have in their life.

AI-ALOE is designing AI-powered cognitive assistants that are sensitive to age and experience, among other factors, to cater to the diverse needs of adult learners. The theory of action that the AI institute addresses the question of how to use online education for adult learning effectively. AI-ALOE hypothesizes that using AI assistants and personalizing learning for working adult learners can make online learning for adults effective. This hypothesis is supported by Benjamin Bloom's (1984) research on the "2 sigma problem," which found that students receiving one-on-one tutoring combined with mastery learning techniques performed two standard deviations (2 sigma) better than those in traditional classroom settings. This effect size means the average tutored

student outperformed 98% of students taught conventionally, highlighting the transformative potential of personalized learning approaches.

The second hypothesis focuses on leveraging AI assistants to address common challenges in online education, grounded in the Community of Inquiry (CoI) framework proposed by Garrison et al. (2000). The CoI framework emphasizes three critical elements for a successful learning experience: cognitive presence (students' ability to construct and confirm meaning through reflection and dialogue), social presence (participants' ability to connect with the community and communicate meaningfully), and teaching presence (the design, facilitation, and management of cognitive and social interactions for learning). By integrating AI within this framework, we can enhance personalization, engagement, and interaction in online education, bridging gaps that traditional approaches often leave unaddressed. These online problems are generally related to how to engage learners in better ways in online learning, specifically, how to enhance learners' cognitive presence, teacher presence, and social presence in online education. The Garrison et al. (2010) framework provides various components for how to address these issues; AI-ALOE aims to address these problems using AI-powered assistants.

AI-ALOE is grounded in the philosophy that our focus should shift from traditional education to the broader concept of learning, which centers on continuous capacity building and lifelong growth. This perspective calls for leveraging emerging technologies, such as AI, to support and enhance human learning in innovative ways and redefine how we approach growth and transformation.

Duolingo

Alina A. von Davier, Chief of Assessment at [Duolingo](#), presented her work on the use of generative AI for test item development. There are several problems that Duolingo has been addressing using generative AI regarding traditional test development (i.e., time-consuming, expensive, and prone to validity issues and security breaches). Specifically, traditional test item development required expert-written test items that needed thousands of pilot administrations and psychometrics item analyses which, due to the labor intensive nature of this process, would usually lead to limited pool of test items. We need a paradigm in assessment that can provide access to high quality (valid, reliable, and secure) tests to as many people as possible. Digital-first assessments can provide a solution to this problem (Langenfeld et al., 2022).

Digital-first assessments are born, maintained, and modified in a digital environment and are meant to be used in digital form. These assessments are designed with computational models and algorithms in mind. Digital-first assessments are theory-based and evidence-centered, accounting for the complexity of the nature of data, and they are valid and reliable. These assessments are integrated through a fluid infrastructure with automatic tools and databases. Most importantly, these assessments derive from and focus on test-takers' experience with technology. For instance, people today are used to the on demand, anytime, anywhere nature of technology. With digital-first assessments, the same user experience can be produced by online testing with remote proctoring from the comfort of the test takers' homes.

Generative AI can help advance the development of digital-first assessments. Generative AI involves sensing (through computer vision and audio processing), comprehending (through knowledge representation and natural language processing), and analyzing various data types (thorough machine learning and statistics techniques). In the context of test development and assessment, the analysis part is called computational psychometrics (von Davier et al., 2022)—that is, computational models from machine learning and statistics but put into the psychometrics framework of reliability, validity, and generalizability (Trognon et al., 2022). Generative AI, in general, deals with large computational models dealing with language, text, image, voice, and video. At Duolingo, generative AI is used to develop a large item bank to support a valid and secure digital-first assessment, administered on-demand and at scale. Generative AI will replace item writing by experts, pilot administrations, and the limited psychometric analysis steps in the traditional test development process.

This new, AI-based item generation process still involves humans in the loop. The process begins with human experts, such as those specializing in TESOL (Teaching English to Speakers of Other Languages) or applied linguistics, designing initial items using a framework that guides evidence-based development. These items are then examined to determine the key features that best demonstrate the targeted skills or knowledge. Next, AI systems use these features to generate additional items. A second group of trained human experts reviews and refines the AI-generated items to ensure they are accurate, unbiased, and fair, while also verifying facts and addressing any potential fairness concerns. Finally, the reviewers' feedback and edits are incorporated to improve the AI systems, enhancing their ability to produce high-quality items in the future. Following this loop, Duolingo's assessment team has used many AI models and technologies (e.g., GPT3, BERT, Chat GPT, Stable Diffusion) to automatically generate various English test item types (e.g., fill-in-the-blank vocabulary questions, multiple choice questions, free response comprehension questions, writing and reading passages).

Khan Academy

Kristen DiCerbo, the chief learning officer at [Khan Academy](#), presented the work she and her colleagues have been doing regarding use of Open AI's generative system (i.e., Chat GPT4) in Khan Academy's ecosystem from September 2022. Her presentation was entitled "harnessing large language models for empowered learning at Khan Academy." As of April 2023, Khan Academy has been experimenting with ChatGPT in four key areas through small-scale projects. These include developing a dialogue-based tutor, creating a student partner for language activities, serving as a teaching assistant, and providing support for content creators. In her presentation at the AI in Education workshop, DiCerbo focused primarily on the dialogue-based tutor and the teaching assistant roles. These experimental projects aim to collect diverse feedback and refine the AI-powered features of Khan Academy.

The first use case DiCerbo discussed was an AI-powered tutor. Through this effort, the Khan Academy team asked how they can use generative models to do things they have never been able to do at Khan Academy (e.g., asking students why they have answered a problem in a certain way, asking students to explain their solution, engaging students in a dialogue with the tutor agent). The tutor that the Khan Academy team has designed and have been testing is called Khanmigo, which can assist students with their math problems (Figure 4).

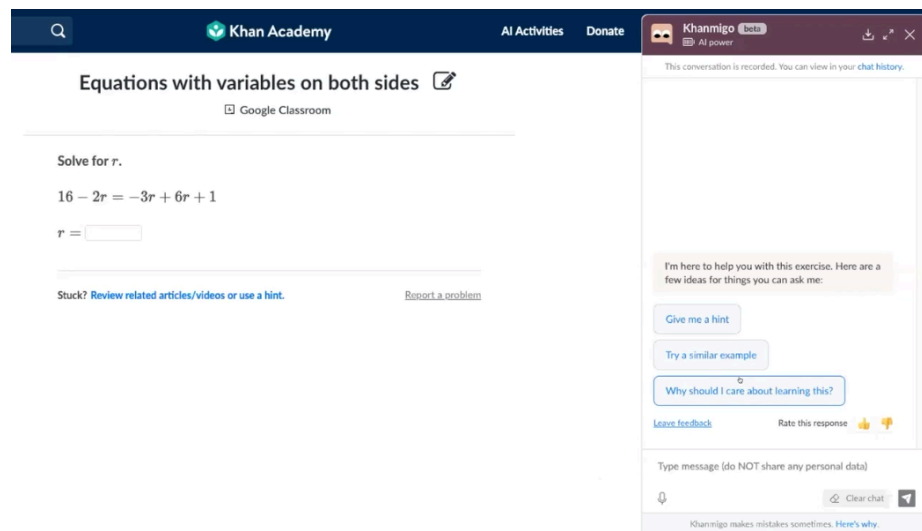


Figure 4. Khanmigo: the AI-powered math tutor at Khan Academy

Khanmigo initially prompts students with three existing commands of “Give me a hint,” “Try a similar example,” and “Why should I care about learning this?” The Khanmigo chat box reminds students not to share their personal data (e.g., their name) to protect students’ privacy. Khanmigo anonymously sends what students typed, along with a prompt explaining to Open AI’s Chat GPT how to respond (i.e., prompt engineering.. The only way Open AI could know who sent the message is if the students write their personal information in the chat box.

Additionally, Khan Academy is exploring a prompt playground empowered with AI to help and assist teachers. For example, this playground allows the teachers to interact with the AI system (connected to Chat GPT in the background) and create rubrics (e.g., illustrating different levels of performance when it comes to writing an essay). Such prompt-engineered AI assistants can have a lot of potential to augment teachers’ performance in the classroom.

ASSISTments

Niel Heffernan, professor of computer science at Worcester Polytechnic Institution and [ASSISTments](#)’s director, shared what he and his colleagues have been doing regarding the use of AI in their math tutoring platform. ASSISTments is a platform that teachers use to build their own content (i.e., math problems and feedback) or use what exists in ASSISTments’ library to assign to their students. Other than letting teachers curate the problems they want their students to work on, ASSISTments provides learning analytics that can quickly and formatively help teachers address any issues or misconceptions students may have in real time. Sometimes teachers realize that since a large number of students got a problem wrong, the instructions might have been confusing or insufficient. Thus, remediating actions will take place to prevent poor performance by students in the future.

The ASSISTments team has been paying close attention to how to help teachers with evaluating students’ responses to open-ended questions. So far in the ASSISTments platform, students have answered an open-ended question 45 million times, teachers have graded those responses 6 million times, and 1 million times a teacher gave a comment to those responses in ASSISTments. One solution to help teachers in responding to students’ responses is similar to Google smart responses (i.e., the teacher can choose a crafted response which is relevant to the student’s answer and quickly send it back to the student). Using reinforcement learning and large language models, ASSISTments provides three appropriate suggested messages teachers can choose from after they grade a student’s response. This is a safe use of LLMs with a human (teacher) in the loop vetting the messages created by LLMs. In a similar effort, the ASSISTments team used GPT-3 to generate explanations based on teachers’ explanations and then asked other teachers to rate those AI-generated explanations. Results showed the AI-generated explanations were unable to outperform teacher written explanations. These findings could be improved using the more updated GPT models.

Key Issues From The Panel Discussions



In this section, we summarize the discussions in each panel: (1) Partnership of Humans and AI / Intelligence Augmented (IA); (2) AI Education from K-12 to Adult Learning; (3) Fundamental Research, Development, and Theory Building; (4) AI in Assessment and Improving Learning; (5) Policy and Challenges related to AI in Education.

Partnership of AI and Humans: Augmented Intelligence

The first panel included Kristy Boyer from the University of Florida, Kristen DiCerbo from Khan Academy, and Neil Heffernan from ASSISTments and Worcester Polytechnic Institute. The following are the main takeaways from this panel discussion:

The definition of AI evolves with its use. Panelists began by defining AI, agreeing that its definition evolves based on its use to augment human capabilities. They noted that AI is perceived differently over time; for example, what was once considered intelligent, such as grammar correction, no longer is. One panelist argued that AI consists of predictive tricks rather than true intelligence, while another emphasized that AI's definition is fluid and context-dependent, with no clear boundaries. The consensus was that AI's value lies in how it enhances problem-solving and goal achievement.

Despite the promise of AI, there are challenges to consider, such as the need for human oversight and the potential risks of relying solely on AI in self-directed learning environments. The panelists highlighted that AI can make errors, and without a teacher to provide guidance, these mistakes can be problematic. Additionally, while both AI and human teachers can make mistakes, the ability to engage in dialogue and seek explanations from human teachers is crucial, a capability that AI currently lacks. Therefore, a cautious, human-in-the-loop, and research-driven approach to integrating AI in education is recommended.

Learner agency should be maintained and cultural sensitivity should be considered when implementing AI in education. Panelists discussed the potential and challenges of using AI tools such as Chat GPT in education. They emphasized the importance of evolving pedagogical paradigms alongside AI, empowering learners to provide feedback, and maintaining their agency. Panelists mentioned that historical practices often ignored learners' cultural backgrounds, a mistake to avoid with AI. The need for due diligence when using AI was stressed, and reinforcement learning with human oversight, as practiced by ASSISTments and Khan Academy, was highlighted as a method to improve accuracy.

AI tools should be evaluated to ensure they meet the diverse needs of learners, including those with special needs. Panelists discussed how AI can address individual learner needs and incorporate differences. The experts anticipated future AI tools that go beyond simple interactions to deeply integrate student profiles, including proficiency, reading levels, dispositions, and interests. This deeper connection would tailor AI interactions more effectively. Additionally, the panelists emphasized the importance of evaluating AI interventions, for instance those in ASSISTments, to ensure they benefit students with special needs and work well across diverse backgrounds and demographics.

AI Education

The AI Education panel included Lin Lin Lipsmeyer, Victor Lee, and George Siemens. The points of discussion in this panel revolved around several essential topics such as:

- Factors that influence people's trust in AI and how education can help address these factors
- Transparency in AI
- Addressing the public's misconceptions about AI and how it works
- AI education, competencies, and literacies that we as educators need to cultivate in our students with the dominance of AI
- The digital divide and how it can threaten education
- The following are the main takeaways from this panel discussion:

AI literacy is essential for demystifying misconceptions and preventing the formation of new ones. Panelists advocated for broadening the understanding of various AI technologies and highlighted the role of sense-making conversations in shaping our comprehension. The experts emphasized AI literacy is crucial for recognizing the potential for misinformation and disinformation spread by AI. Additionally, educators should equip learners with competencies for both short-term practical use and long-term adaptation to AI's evolving role. Short-term literacy includes understanding AI capabilities, limitations, and data literacy. Long-term literacy involves conceptualizing new roles for humans in an AI-driven world. Data literacy was also emphasized by the panelists, stressing the need to understand data sources and quality. Educators need to broaden their understanding beyond popular AI technologies and address misinformation and disinformation amplified by AI. AI literacy can help mitigate the spread of misinformation and disinformation by making people aware of AI's limitations and potential for generating false information.

Trust in AI depends on transparency and alignment with users' goals. Panelists discussed the issue of trust in AI, emphasizing the need for transparency and predictable performance to build trust. While complete transparency may not be necessary for end users, it is essential for technical populations. The panelists suggested that educating both developers and users about the capabilities and limitations of AI systems is necessary for building trust. Moreover, the panel suggested that AI developers should create systems worthy of trust and that users should be educated on what to expect from AI. Trust in AI should be seen as a series of subsystems, varying based on the AI's role, such as generating ideas versus handling social and emotional dynamics. An example was given where an AI-generated letter by the administrations of a large university following a shooting incident led to criticism, highlighting trust issues in emotionally sensitive contexts. It was also noted that trust in AI will evolve over time, and education on AI must adapt to these changes, emphasizing the importance of setting appropriate expectations and staying current with AI developments.

AI technologies may exacerbate the digital divide due to economic barriers, lack of support for proper usage, and biases in AI models. Panelists highlighted economic barriers, such as the cost of accessing advanced versions of AI tools such as Chat GPT, which could be mitigated by adopting business models similar to those used in social media. Additionally, the panelists emphasized the need for support in using AI tools properly, such as assistance with prompt engineering, particularly for those without adequate social support. Solutions include providing access to AI tools and supporting local productions of AI. Lastly, the experts in this panel discussed the inherent biases in AI models, which may favor certain perspectives and exclude others, potentially leading to further alienation for those not represented in the models.

Fundamental Research, Development, and Theory Building

James Lester, Peter Foltz, and Diego Zapata-Rivera formed this panel discussing the latest developments around fundamental research, development, and theory building. The discussions in this panel were mainly about the infrastructure needed for AI in education, the opportunities that LLMs have introduced, use of AI in fundamental educational research, and the interplay between AI models and theories of learning. The following are the main takeaways from this panel discussion:

A national infrastructure for AI in education could facilitate the adoption of new technologies in school districts. Panelists discussed the infrastructure needed to promote AI in education. The panel emphasized continuous evaluation of AI models and developing AI-powered learning environments to support learners. The panelists noted the importance of capturing rich interactions between students and teachers, beyond just computer activity, to include context such as gaze and communication. Additionally, the discussion highlighted the need for a national infrastructure that enables school districts to use new AI technologies. This infrastructure should include not only hardware and software but also skilled individuals who can implement the technology effectively. The AI institutes program at NSF was mentioned as a crucial initiative for developing this infrastructure rapidly.

While LLMs offer significant opportunities, researchers are dependent on large tech companies for infrastructure and computing power. Panelists discussed the dependency on large companies for AI infrastructure and computing power, especially with large language models such as Chat GPT. The panel noted that while foundational models from big tech companies are essential, many smaller models can be fine-tuned for specific purposes and run on less powerful devices. The importance of having skilled individuals in learning theories, cognition, and prompt engineering was highlighted. The panelists also emphasized incorporating human oversight to provide contextual information for these models and protect student privacy, ensuring that AI tools do not have overly deep knowledge of individual students.

AI technologies have the potential to revolutionize educational research by enabling qualitative analyses, multimodal learning analytics, and faster assessment creation. Panelists discussed AI's role in advancing educational research. The panel highlighted how AI enables qualitative analysis of learning interactions more efficiently, allowing researchers to address old questions and explore new ones. AI technologies such as natural language processing (NLP) and multimodal learning analytics can automate labor-intensive processes, such as coding qualitative data. For example, AI can accelerate conversation-based assessments, which previously required extensive human training. Furthermore, the panelists expressed excitement about using AI to test theories related to higher-level thinking skills, analyzing students' discourse, and exploring the impact on knowledge gain and affective responses. In addition, AI's ability to handle large datasets allows for new generalizations and the testing of longstanding theories.

Researchers can now ask new questions and test hypotheses that were previously challenging or impossible to explore. During the discussion, panelists addressed the criticism that AI models often lack guidance from learning theories and operate as black boxes. The panelists acknowledged that while AI can predict learning outcomes using vast amounts of data independently of theories, integrating learning constructs into these models remains challenging. The experts in this panel discussed the potential of hybrid approaches that combine top-down (theory-driven) and bottom-up (data-driven) methods to enhance AI models. Despite recent advancements in applying AI to self-regulated learning and emotion regulation, explainable AI (XAI) is still developing. The panelists expressed hope that as industry advances model capabilities, it will also improve model explainability, benefiting educational applications.

AI in Assessment and Improving Learning

In this workshop, we had two separate panels on the topic of AI in Assessment and Improving Learning. Panelists included Valerie Shute, Ryan Baker, and Russell Almond on one panel and Sidney D'Mello, Alina von Davier, and Richard Golden on another panel. The panelists in both rooms talked about issues around validity, reliability, fairness, and equity of AI-driven models in advanced learning environments. The following are the main takeaways from this panel discussion:

Validity and reliability should be carefully considered in AI-driven assessment models. The panelists stressed the critical importance of ensuring the validity and reliability of AI-driven assessment models through methods such as stealth assessment and evidence-centered design. [Stealth assessment](#) (Shute & Rahimi, 2022) involves integrating evaluations directly into learning tools, like educational games, so students' progress can be measured seamlessly. These assessment models are validated through correlations with external measures and reliability checks such as confirmatory factor analysis. Additionally, the analogy between machine learning approaches and classical statistical methods was discussed, emphasizing the necessity of expert labeling in both supervised and unsupervised learning to maintain validity. Calculating weights of evidence was also highlighted as crucial for understanding the impact of each data point on the overall assessment's accuracy to enhance transparency and reliability.

Researchers should address bias and equity issues in AI-driven assessments using techniques such as expert review, fact-checking, and diverse representation in data to mitigate bias and ensure fairness in assessments. The panelists emphasized the dangers of algorithmic bias, stressing the importance of collecting diverse data and ensuring AI models do not disadvantage specific groups. Positive outcomes from work on stealth assessments showed that equitable learning outcomes can be achieved through careful design, with different genders and ethnicities starting at different levels but ending up with similar learning outcomes. Fact-checking and human review were discussed as necessary steps to mitigate biases in AI-generated items. Designing equitable learning environments that respect diverse communication forms and addressing implicit biases were also emphasized as crucial for fair and effective AI-driven assessments.

Assessing hard-to-measure constructs such as creativity, collaboration, and persistence is challenging—AI can enhance our capabilities in this area.

It is important to consider multiple indicators and contextual factors in assessing such constructs accurately. The panelists highlighted difficulties of using speech recognition for assessing collaboration in classrooms. Multimodal data, including gestures, is needed to be integrated in classroom-based data collection to accurately assess hard-to-measure constructs. Moreover, transparent AI, which engages students in validating assessments, was advocated for, as it involves students in the assessment process. Finally, the complexities of measuring persistence were elaborated on, noting that it should be understood contextually with different strategies and outcomes significantly affecting assessments. Such a multifaceted approach allows for more accurate evaluation of such complex constructs.

We need to have a human-centered approach to AI in education rather than an AI-centered approach. The panelists advocated for a human-centered approach to AI in education, where AI supports human interaction rather than replacing it. One of the panelists emphasized that tasks best handled by AI, such as classroom management and grading, should be distinguished from those requiring human care and motivation. AI could free up teachers' time, allowing them to focus on nurturing and motivating students. By automating administrative tasks, AI allows teachers to focus on providing personalized, empathetic education, ensuring that the human element remains central in the learning process. This perspective underscores the unique value of human interaction in education, promoting the idea that AI should enhance the educational experience by supporting teachers, rather than being the central component of the learning environment.

AI Policy Challenges and Scale

The last panel included Chris Dede from Harvard University, Danielle McNamara from Arizona State University, and Rene Kizilcec from Cornell University. The panel started by investigating practical ways to ensure students' learning without violating their privacy. The following are the main takeaways from this panel discussion:

AI in education can learn from AI in health (e.g., scale and anonymization to address privacy issues effectively). Drawing parallels between AI applications in education and health, the panelists explored how techniques used in healthcare, such as data anonymization and large-scale analysis, could be applied to educational contexts. The panelists pointed out that in healthcare, electronic health records have revolutionized patient care and research, suggesting that similar approaches could benefit education. By clearly distinguishing between research-focused and commercial use cases of educational data, stakeholders can better manage privacy concerns and leverage AI to enhance learning outcomes while safeguarding learners' privacy.

The panelists emphasized the need for measures such as training and memorandums of understanding (MOUs) to ensure compliance with privacy regulations. The panel highlighted the complexity of navigating legal constraints such as the Family Educational Rights and Privacy Act (FERPA) and stressed the importance of creating frameworks that allow for the analysis of aggregate data without exposing individual private data. These measures would help institutions maintain compliance with privacy laws while enabling the use of data to improve educational practices and outcomes.

While recognizing AI's potential to enhance learning outcomes, AI can be misused, therefore ethical and responsible use AI should be promoted. The panelists emphasized the need for ethical and responsible AI (e.g., Arrieta et al., 2020) deployment in education, advocating for a thoughtful approach that considers the nature of educational tasks and the implications of AI applications. The discussion included the importance of developing guidelines to differentiate between beneficial and harmful uses of AI, ensuring that AI technologies are used to genuinely enhance educational experiences rather than just streamline processes.

There are essential skills and competencies for students and educators in the age of AI that we need to identify and establish. The panelists highlighted the need to identify and establish essential skills and competencies for both students and educators in the age of AI. The discussion also highlighted the idea of creating new roles, such as “learning engineers,” who would work alongside teachers to personalize learning experiences. This collaborative approach would help integrate AI into education more effectively, ensuring that educators are not expected to become technologists but can focus on their primary role of engaging with learners.

Addressing the use of AI chat apps on college campuses, the panelists advocated for establishing clear guidelines rather than outright bans. This panel also highlighted that guidelines would promote responsible usage and help users understand how to effectively and ethically leverage such technology. By setting these standards, institutions can mitigate potential risks while embracing the benefits of AI chat apps in enhancing communication and learning experiences.

AI technologies can be transformative in education, however, thoughtful consideration of its implications are necessary by all stakeholders. A shift in focus from merely improving processes to creating better educational outcomes is needed, particularly addressing equity and enrichment. In their final thoughts, the panelists called for a shift in focus from improving processes to creating better educational outcomes. The experts stressed the importance of addressing the achievement gap by concentrating on enrichment rather than remediation. The panelists conveyed optimism about the future of AI in education, emphasizing the need for transformative approaches that prioritize equity and thoughtful consideration of AI’s broader implications on learning and society.

Recommendations for Future Work



The AI in Education Workshop, held online in April 2023 at the University of Florida, gathered scholars to explore AI-driven education advancements and challenges. The workshop facilitated discussions on AI applications in education, ranging from research to practical implementations and policy. Key accomplishments included identifying themes such as human-AI partnerships, AI literacy, infrastructure needs, and ethical implications. Additionally, discussions emphasized human oversight, cultural sensitivity, and addressing the digital divide. The workshop featured examples of successful human-AI collaboration models, such as those demonstrated by ASSISTments and Khan Academy.

Since April 2023, researchers have developed guidelines for generative AI technologies. Now that these guidelines are established, it is time to refine them and implement them effectively. However, AI in education scholars must remain vigilant to rapid advancements, particularly in industry. Looking ahead, the evolution of AI technologies, including advancements in quantum computing and computational power, along with the integration of vast datasets, promises to further revolutionize the educational landscape. These technological leaps could lead to more

sophisticated AI models capable of providing deeper insights into learning processes and student engagement, thereby refining educational practices and outcomes at all levels.

With quantum computing-powered AI on the horizon, educators face several open questions: What educational problems can be addressed by such powerful AI technologies? What challenges may arise that we need to be prepared for?

Several next steps emerged from the workshop:

- Development of AI literacy and trust: Create AI literacy programs for students and educators, focusing on practical skills, ethical considerations (e.g., Akgun et al., 2022), and understanding AI's limitations and biases.
- Creation of national infrastructure for AI in education: Establish a robust national infrastructure, including hardware, software, and training programs, to support AI integration in school districts.
- Addressing the digital divide: Develop business models and smaller, accessible AI models to lower economic barriers and ensure equitable access to AI technologies.
- Integration of learning theories in AI models: Focus research on incorporating learning theories into AI models through hybrid approaches, enhancing their educational value and explainability.
- Ensuring ethical and responsible AI use: Develop guidelines to promote ethical AI use in education, ensuring privacy, addressing bias, and considering societal implications.
- Ongoing evaluation and adaptation: Continuously evaluate and adapt AI tools to evolving educational needs, incorporating new findings and technologies.

References

- Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2(3), 431-440.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58, 82-115.
- Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. *Ai Magazine*, 26(4), 53-53.
- Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science*, 136, 16-24.
- Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118.
- Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The internet and higher education*, 13(1-2), 5-9.

- Langenfeld, T., Burstein, J., & von Davier, A. A. (2022, May). Digital-First learning and assessment systems for the 21st century. In *Frontiers in Education* (Vol. 7, p. 857604). Frontiers.
- Shute, V. J. & Rahimi, S. (2022). Stealth assessment: A Primer. *Rapid Community Report Series. Digital Promise and the International Society of the Learning Sciences*. <https://repository.isls.org/handle/1/7671>, pp. 1-11.
- Trognon, A., Cherifi, Y. I., Habibi, I., Demange, L., & Prudent, C. (2022). Using machine-learning strategies to solve psychometric problems. *Scientific Reports*, 12(1), 18922.
- von Davier, A. A., Mislevy, R. J., & Hao, J. (Eds.). (2022). *Computational psychometrics: New methodologies for a new generation of digital learning and assessment: With examples in R and Python*. Springer Nature.

Resources

- UF's AI in Education Workshop: <https://ialt.education.ufl.edu/ai-in-education-workshop/>
- iSAT AI Institute: <https://www.colorado.edu/research/ai-institute/>
- Engage AI Institute: <https://sites.google.com/ncsu.edu/ai-engage/home>
- AI-ALOE Institute: <https://aialoe.org/>

Acknowledgments

We would like to thank the [Institute for Advanced Learning Technologies \(IALT\)](#), and the [E-Learning, Technology and Communications](#) teams at the University of Florida who helped in various stages of conducting this workshop. We also would like to thank everyone who helped in planning, conducting, and moderating the panels in this workshop: Tom Dana, Pasha Antonenko, Jason Dean Arnold, Albert Ritzhoup, Corrine Huggins-Manley, Walter Leite, Anthony Botelho, Jinnie Shin, Wanli Xing, Robe Moore, Kevin Coulson, Alexis Moore, Moxy Moczygemba, Hailey Kuang, Yukyeong Song, Andrea Ramirez-Salgado, Brian Abramowitz, Hyunju Oh, Maedeh Agharazi, Hanxiang Du, Woorin Hwang, Ray Opoku, Lissette Tolentino, Ziyang Li, Amber Hatch, and Priya Ganapathy.

Rapid Community Reports are supported by the National Science Foundation under grant #2021159. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.