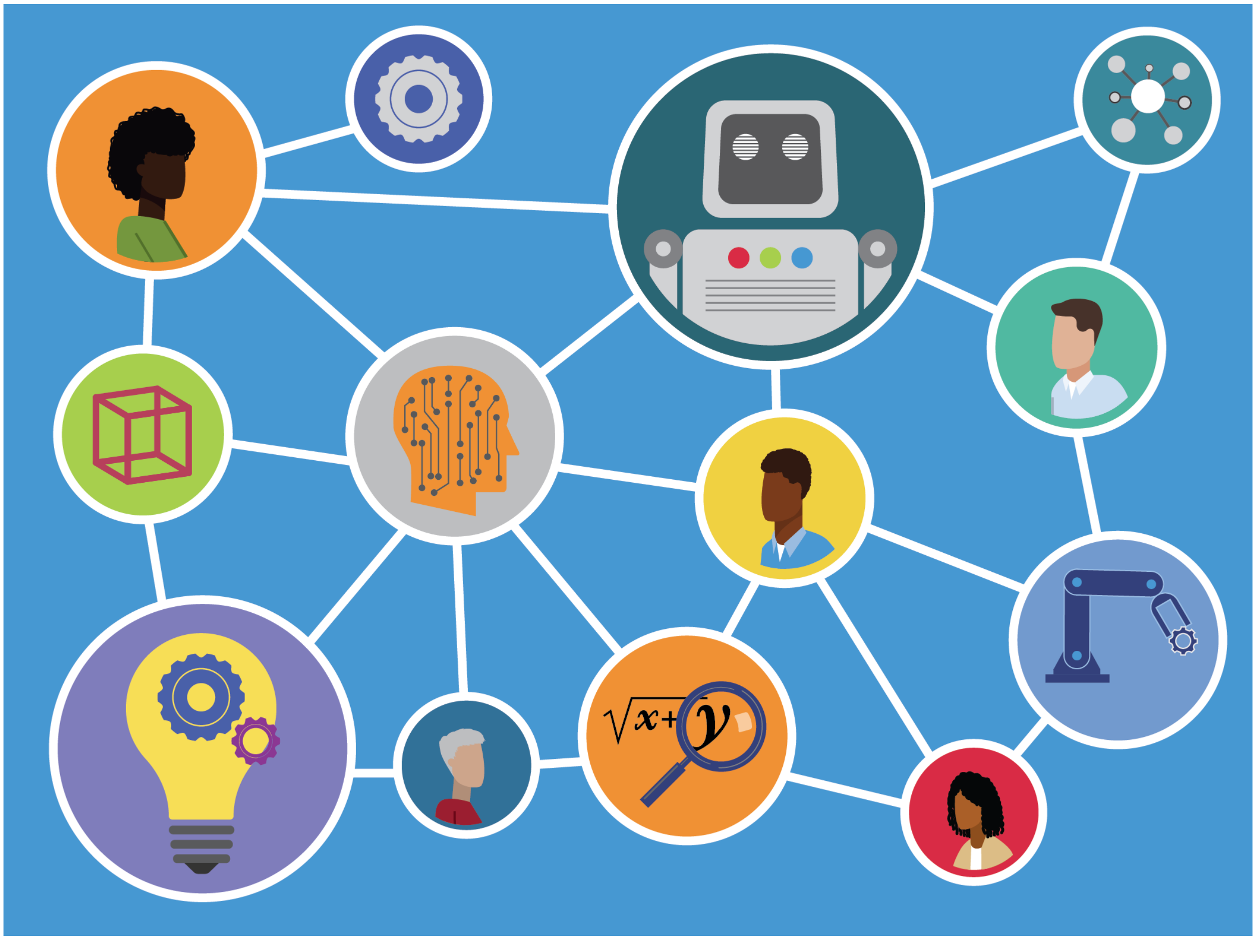
AI and the Future of Learning: Expert Panel Report

Jeremy Roschelle, James Lester, and Judi Fusco (editors)
November 2020



CIRCLS
Center for Integrative Research in Computing and Learning Sciences
University of Pittsburgh
SRI Education
NC STATE UNIVERSITY
EDC Learning transforms lives.
Digital Promise
Accelerating Innovation in Education

Suggested Citation

Roschelle, J., Lester, J. & Fusco, J. (Eds.) (2020). *AI and the future of learning: Expert panel report* [Report]. Digital Promise. https://circls.org/reports/ai-report.

Acknowledgements

We are grateful to the 22 experts who participated in the panel: Dr. Russell Almond, Dr. Ryan Baker, Dr. Avron Barr, Dr. Gautam Biswas, Dr. Justine Cassell, John Cherniavsky, Dr. Sherice Clarke, Dr. Chris Dede, Dr. Sidney D'Mello, Dr. Janice Gobert, Dr. Cindy Hmelo-Silver, Dr. Susanne Lajoie, Dr. Diane Litman, Dr. Rose Lucklin, Dr. Maja Mataric, Dr. Danielle McNamara, Dr. Jaclyn Ocumpaugh, Dr. Amy Ogan, Dr. Zach Pardos, Dr. Brian Smith, Dr. Kurt Van Lehn, and Dr. Marcelo Worsley.

In addition, we thank the leaders from the National Science Foundation, Karen Marrongelle, Amy Baylor, and Tanya Korelsky; and from the U.S. Department of Education, Jake Steel, Bernadette Adams, along with participants Adam Safir, Kevin Johnstun, Sara Trettin, Christina Chhin, and Edward Metz; and our Digital Promise colleagues, Karen Cator, Barbara Means, Melissa Bellin, and Kasey Van Ostrand.

Finally, a special thank you to Erin Walker for helping to improve this report.

This material is based upon work supported by the National Science Foundation under grants 2021159 (CIRCLS) and 1837463 (CIRCL). 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.

Table of Contents

Executive Summary

Artificial intelligence (AI), machine learning, and related computational techniques have the potential to make powerful impacts on the future of learning. Technology's impact on education is often to amplify impacts, regardless of whether the impacts are intended. Due to the accelerating pace of integration of technology in learning environments, the knob on the amplifier is rapidly going from low to high. Impacts on learning, whether positive or negative, could soon have consequences for many more students. Now is the time to begin planning for how to best develop and use AI in education in ways that are equitable, ethical, and effective and to mitigate weaknesses, risks, and potential harm.

We convened a panel of 22 experts in AI and in learning to address these issues. They met online for seven hours over two days in a facilitated process with different topics and breakout formats. The experts considered two broad questions:

1. What will educational leaders need to know about AI in support of student learning in order to have a stronger voice in the future of learning, to plan for the future, and to make informed decisions?
2. What do researchers need to tackle beyond the ordinary to generate the knowledge and information necessary for shaping AI in learning for the good?

This report introduces three layers that can frame the meaning of AI for educators. First, AI can be seen as "computational intelligence" and capability can be brought to bear on educational challenges as an additional resource to an educator's abilities and strengths. Second, AI brings specific, exciting new capabilities to computing, including sensing, recognizing patterns, representing knowledge, making and acting on plans, and supporting naturalistic interactions with people. These specific capabilities can be engineered into solutions to support learners with varied strengths and needs, such as allowing students to use handwriting, gestures, or speech as input in addition to more traditional keyboard and pointer input. Third, AI can be used as a toolkit to enable us to imagine, study, and discuss futures for learning that don't exist today. Experts voiced the opinion that the most impactful uses of AI in education have not yet been invented. The report enumerates important strengths and weaknesses of AI, as well as the respective opportunities and barriers to applying AI to learning.

Through discussions among experts about these layers, we observed new design concepts for using AI in learning. Experts discussed how AI could support learning in terms of orchestrating complex learning activities with multiple people and resources, augmenting human abilities in learning contexts, expanding naturalistic interactions among learners and with artificial agents, broadening the competencies that can be assessed, and revealing learning connections that are not easily visible. These approaches go beyond familiar design concepts for individualized, personalized, or adaptive learning. To bring these approaches to life, experts suggested two broad scenarios.

The **learning environment scenario** notably featured social learning. In this scenario, AI supports orchestration of the multiple types of activities, learning partners, and interaction patterns that can enrich a classroom. This is different from many older images of AI that focus on an isolated individual who interacts only through and with a single device. An AI agent could provide support to a group of students as they work on a project or assignment together. This could include support for students to work as team members (e.g., noticing, listening to, and building on each other's contributions) as well as task supports that help them organize, manage, and connect their contributions to an overall group goal. It could adapt to what groups need to work well together, as well as how groups transition between individual work, small group work, and discussions with the whole classroom. This kind of AI system is socially aware, and can use social interaction with students as a way of bootstrapping their academic performance.

The **assessment scenario** envisioned going beyond what today's assessments can measure. Instead of just grading an essay, an AI agent could help a teacher build a portrait of a student's competencies. For example, AI can assess their writing, including students' strengths and weaknesses, and provide suggestions for improvement that considers all the contexts in which the student does the writing. Teachers could, in turn, showcase other features of students' writing experiences and accomplishments and build on these in instruction. This is fundamentally different from assessment metaphors today that focus on automated grading of a particular performance or diagnosing specific student errors. With added information about the student, the AI system can provide advice on how to support the learning process of that complete individual, based not only on their narrow test and class performance for a given class, but on their holistic background and trajectory.

The two scenarios connect. In a few years, it might become possible to connect learning with assessment as students collaborate to learn and learn to collaborate. A range of social, emotional, and cognitive skills might be better supported, going beyond the academic content typically measured today. Further, a focus on collaborative learning is just one way in which AI could enable powerful learning that aligns to how our society and its work are evolving.

We also noticed that, in terms of risks, the experts were obviously concerned with well-known risks related to data—such as privacy, security, bias, transparency, and fairness. But they also went beyond these expected concerns to talk about design risks and how poor design practices could unintentionally harm classes of users. In addition, they foresaw a major risk in not informing and involving educational policy makers and practitioners early and deeply enough.

The panel made seven recommendations for research priorities:

1. Investigate AI Designs for an Expanded Range of Learning Scenarios
2. Develop AI Systems that Assist Teachers and Improve Teaching
3. Intensify and Expand Research on AI for Assessment of Learning
4. Accelerate Development of Human-Centered or Responsible AI
5. Develop Stronger Policies for Ethics and Equity
6. Inform and Involve Educational Policy Makers and Practitioners
7. Strengthen the Overall AI and Education Ecosystem

The expert panel was well aware that this meeting had limited involvement of other stakeholders and wanted to make clear that this was an initial discussion. They noted that the involvement of practitioners, policymakers, innovators, and industry in further discussions on these issues is imperative and that the experts would gladly participate with a broader set of stakeholders.

Introduction: Why Now?

Artificial intelligence (AI), machine learning, educational robotics, and related technologies will have powerful impacts on the future of learning. We do not yet know all of the uses and applications of AI that will emerge; new innovations are appearing regularly and the most consequential applications of AI to education are likely not even invented yet. Amidst the rapid expansion, we know there are both potential benefits and considerable risks. Although the greatest, scalable impacts may still be many years into the future, educational planning needs a long horizon to be effective.

Technology Amplifies Impacts of Design Tradeoffs

To explain why these issues are now urgent, we begin with a metaphor: technology can be an amplifier (Toyama, 2015).

When we say amplify, what do we mean? By amplification, we mean that learning technology can take an aspect of a learning process and emphasize it, refine it, intensify it, and scale it widely. This can be good or bad; undesirable or desirable effects on learning can scale with equal ease. Whether people design innovative ways to teach and learn or use technology to scale existing best practices, tradeoffs always occur. For example, a student may spend more time on assignments that exactly match their level, but less time learning important social skills. They may get more feedback on those aspects of learning that technology can easily measure and less feedback on equally or more important aspects that are hard to measure. When technology amplifies an approach to teaching and learning, the consequences of each decision affects more learners with greater intensity.

Further, due to the accelerating pace of integration of technology in learning environments (U.S. Department of Education, 2017), the knob on the amplifier is rapidly going from low to high. For example, in today's COVID-19 pandemic environment, schools, teachers, and learners have had to rapidly make a lot of tradeoffs in how they use technology for learning. We can all observe that suboptimal choices about how to use learning technology can quickly have widespread effects, including magnifying learning loss for some students while others continue to grow apace. As in planning for pandemics, there is always a tendency to invest in what is possible or exciting and to underinvest in analysis of inequitable impacts and mitigation of risks. Given that we anticipate AI will come to greatly impact teaching and learning dramatically in the coming years, now is the time to intensify our society's planning around how to use these powerful capabilities for good.

The Accelerating Intensity and Impacts of AI in Education

Applying AI in education is not new. The history of intertwined research and development of AI, learning research, and educational applications goes back over 50 years. For example, as one of the founders of AI, Marvin Minsky was exploring machines and the nature of mind in the late 1960s and early 1970s, the seminal learning theorist Seymour Papert was inventing the educational programming language Logo (Papert, 1980)—a language based on the AI programming language, LISP. Together they wrote an influential early book closely related to machine learning (Minsky & Papert, 1968).

Defining Artificial Intelligence

AI doesn't have a single accepted definition. For this report, we conceptualize three aspects of AI—as an ambitious leading edge of computing, as a set of specific capabilities that are rapidly advancing, and as a toolkit for synthesizing and exploring possible futures for learning and teaching.

As a leading edge of computing, AI's ambition is clear: to create computational machines that examine data, make inferences, and then act by themselves. This can be seen in a historical set of definitions of AI going back to the 1950s, such as "AI is the study of how to make computers do things at which, for the moment, people are better" or "the investigation and construction of intelligent agents that perceive and act in order to maximize their chances of success" or "the theory and development of computer systems able to perform tasks normally requiring human intelligence" (Richter et al., 2019). Relative to today's advances in machine learning, the ambition can be expanded to include machines that are able to optimize outcomes given a set of data, constraints, and preferences. These capabilities evoke what humans can do. The sense that an AI-based computer program can be self-contained and reason and act on goals without the direct supervision of a human gives rise to thinking of these computer programs as AI agents.

AI is also set of specific capabilities, which are advancing rapidly today. In a report to school technology leaders, Holland (2020) described these as:

- **Perception**, via multiple sensors and ability to recognize complex sets of features (e.g., use of cameras and motion detectors to recognize particular faces entering a building)
- **Representation and Reasoning**, building models of people and their behaviors and making inferences based on those models about what might happen next
- **Learning**, discovering meaningful patterns in large amounts of data
- **Natural interaction** (e.g., interacting through speech or gesture)
- **Societal impact**, leveraging infrastructures to do all the above at a massive scale and in ways that directly affect people's lives

In a third possible definition, AI also empowers a "science of the artificial" (Simon, 1969/1996) where innovators can create new learning environment configurations in order to study what the future could bring. Some examples include the possibility of students learning collaboratively with an artificial agent that facilitates their social interactions, for example, preschool-age children learning science with a social robot who motivates them and supports their inquiries (Kim et al., 2018), or differently-abled learners getting personalized support from a near-peer socially assistive robot buddy (Clabaugh et al., 2019). AI can also be a toolkit for building innovative approaches to assess students' competencies (Paquette, et al., 2014; Mislevy, et al., 2020) and the results can be used to support further learning.

Thus, we define AI and the future of learning as including all three of these layers. First, the layer where computational intelligence can be brought to bear on educational challenges as an additional resource. Second, the layer where specific emerging capabilities can be engineered into solutions for specific education problems. Third, a layer where this toolkit can enable us to imagine new futures for learning, teaching, and assessment.

Although the intertwined history of AI and education is long, for most of the time, impacts have been small scale. With limited exceptions, the uses of AI in learning have been in research projects. A classic early AI and education paper proposed extending computer-aided instruction systems with question-answering capabilities based on a representation of knowledge (Carbonnel, 1970). This concept led to Intelligent Tutoring Systems (ITS), interactive technologies that provide guidance and feedback to learners based on models of student and expert knowledge. One well-known example is the Cognitive Tutor for Algebra I, which has been successfully commercialized and tested at scale (Pane et al., 2014). More generally, meta-analyses of ITS approaches have found positive impacts on student learning (Van Lehn, 2011; Ma et al., 2014; Kulik & Fletcher, 2016). Even though this strand has developed a cogent body of useful knowledge over decades, use of ITS in everyday educational situations greatly lags behind the capabilities that have been demonstrated in research projects.

Experts see AI as accelerating rapidly now, and more intense and widespread impacts will soon become prevalent. One set of factors driving the acceleration is not specific to education. For example, AI has become a core part of our cell phone technology and home assistants, allowing us to talk to phones and to use them as personal assistants. Machine learning, neural networks, and deep learning algorithms are ever-increasing in their prevalence in products to support image processing and speech recognition (Richter et al., 2019). Further, as industry creates and refines interfaces, such as voice assistants, that support more naturalistic interactions between AI and learners, incorporating mobile devices more deeply into learning becomes more appealing to teachers and students.

Another set of factors is more specific to education, where research is expanding rapidly. For example, the AI-based fields of learning analytics (e.g., Krumm et al., 2018) and educational data mining (e.g., Fischer et al., 2020; Slater et al., 2017) are engaging many more scholars each year, resulting in a wealth of research findings. In addition, developers are producing applications such as early warning systems (Krumm et al., 2014). These systems detect when a student's behavior may indicate an increased chance of an undesirable later event, such as dropping out of a course. The capabilities, however, are going beyond observing what students type on a computer or how they answer questions. Newer research-based systems can listen to recordings or watch videos of classrooms, finding events that are significant for learning outcomes (Suresh et al., 2019; Aung et al., 2018). Automated essay scoring is another long-standing application (Page, 2003), which is now rapidly expanding to include assistive systems for peer grading, student collaboration, and other educational applications. More generally, researchers are using AI in ambitious mashups that combine AI technologies with other emerging technologies to produce learning innovations (CIRCL, 2020). These go beyond the most common AI-in-education scenarios to include rigorous performance assessment, virtual reality, voice-based systems, gesture-based systems, social and educational robots, collaborative learning, mobile learning, and more.

In addition, the COVID-19 pandemic has made many people realize that technology will forever be a much bigger part of teaching and learning than it was in the past. Whereas in the recent past, learning technology could have been considered a “nice to have” addition to teaching and learning, now it has become a “must have.” In a future where technology is ubiquitous in education, AI will also become pervasive in learning, teaching, and assessment. Now is the time to begin responding to the novel capabilities and challenges this will bring.

Organizing the Expert Panel

To further investigate AI and the future or learning, we invited 22 expert researchers to a facilitated, online meeting. We sought to address two questions over seven hours of conversation:

- What will educational leaders need to know about AI in support of student learning in order to have a stronger voice in the future of learning, to plan for the future and to make informed decisions?
- What do researchers need to tackle beyond the ordinary to generate the knowledge and information necessary for shaping AI in learning for the good?

In this report, we discuss how experts see the strengths and weaknesses of AI, as well as the opportunities and barriers. We share several scenarios for applying AI to learning that differ from the most common applications and may portend new applications of the future. Additionally, we discuss the recommendations of the experts regarding what research topics need more emphasis in the future.

The expert panel we report was part of our work as the Center for Innovative Research in Cyberlearning (CIRCL). CIRCL hosted the convening in coordination with colleagues at Digital Promise who were working to support policy needs of the U.S. Department of Education, specifically with issues around AI. CIRCL and its successor, the Center for Integrative Research in Computing and Learning Sciences (CIRCLS), are National Science Foundation (NSF)-funded projects that serves as a community center for a cluster of independent NSF-funded projects in the Cyberlearning program. More than 400 of these projects look five to 10 years into the future and apply concepts from computer science and the learning sciences to investigate future learning scenarios. Through CIRCL, we have seen an increasing number of projects that explore "ambitious mashups" of AI capabilities with other resources, technologies, approaches, and capabilities (CIRCL, 2020). At a fall 2019 CIRCL convening of approximately 200 Cyberlearning researchers and investigators, the attendees indicated that challenging issues around ethics and equity of AI applications is a very important area for the field's attention. We've seen our colleagues in other countries organize around some of the issues (e.g., Learning Analytics Community Europe, http://www.laceproject.eu) and the issues around AI and learning with ethics and equity are recurrent at the recent conferences that cyberlearning investigators attend. Further, NSF recently awarded a first-of-its-kind $20 million center on issues of AI in education (Strain, 2020), with a sense that this investment is not the end but rather the beginning of a much more significant emphasis on these issues at NSF. Overall, we are experiencing surging awareness that responsible researchers need and want to start doing more to tackle issues relating to AI and education.

At the opening of our meeting, two invited speakers shared how this expert panel could relate to needs in their respective federal agencies. Jake Steel, deputy director of the Office of Educational Technology at the U.S. Department of Education, said:

> As Secretary Betsy DeVos has stated, "We want to ensure that nothing limits students from being prepared for what comes next." We need to look into the power of AI in education to see how we can empower teachers in their daily job. How can teachers be better at what they do to create stronger learning and stronger assessments? The ultimate goal is to make sure that all students

> everywhere have equal opportunities to learn and engage. We need to make sure that at any time in a learner's life they have that capability.

Karen Marrongelle, assistant director of education and human resources at the National Science Foundation, shared:

> We need to accelerate the pace at which we understand how advances in AI and related technologies can change the landscape of education and conversely how education must change in order to prepare for a world more deeply infused with AI and technology. As an agency that funds basic research on education, NSF realizes that the nature of research on mechanisms for teaching and learning science, technology, engineering, and mathematics (STEM) is rapidly changing. As such a large knowledge base will be needed to use AI safely, equitably, and effectively. This makes having thoughtful conversations about inequities and discussing plans for the future of AI all the more important.

The 22 assembled experts were selected on the basis of recommendations from several sources—our CIRCL team, co-chair James Lester, colleagues at Digital Promise, leaders at the U.S. Department of Education, and NSF. The experts met on a Zoom conference call for three and a half hour-long sessions on two consecutive days. We used the tool Mural.co to provide an online board onto which all participants could post notes; images of these boards appear in this report (see Figures 1 and 2). We worked through an agenda (see Appendix) that asked the experts a series of questions and gave them opportunities to work in various groupings and breakout rooms with a whole panel discussion and reflection at the end of the convening. In addition, at the end of the meeting, experts were invited to write individual recommendations and to post them to a shared space. After the meeting concluded, we reviewed, reflected upon, and synthesized the copious traces of the meeting: voice recordings, transcripts, discussion boards, chat sessions, suggested research papers, and more. The result is this report.

Strengths, Weaknesses, Opportunities, and Barriers

We began our expert panel by asking the attendees to identify the most important strengths, weaknesses, opportunities, and barriers of AI—from the vantage point both of learning technology researchers and of educators. (See Figure 1 for the virtual white board and Post-it notes that the experts created during this discussion.)

With regard to strengths, experts viewed AI as augmenting human intelligence, like when an AI agent and a teacher work together to support a student's learning. The AI agent may be able to give consistent, timely, and nonjudgmental feedback to a student as they work on a complex task, while practicalities might make the teacher less available to do this. An AI agent can be patient, always available, and have access to a dataset of what helped students in similar learning situations; teachers do these things well, but often do not have as much time as it would take to do them for all students. AI agents can increasingly work with more than one person at a time and thus support small groups or a whole classroom. Indeed, an AI agent can be replicated to support all students, whereas teachers strain to spend time with all their learners. When an AI agent has access to data from contexts that a human may not (e.g., information about what another class may be doing), it can make connections across disparate contexts and data sources and can detect patterns that humans miss.

Experts also highlighted how interactions with AI agents are becoming more naturalistic, for example, via speech, gesture, and drawing. In addition, an AI agent can track not just a response to a learning task, but also a student's behavior as they are learning. AI agents can potentially recognize the student's work across different written, spoken, and drawn and enacted modalities. Sensors can also track a student's eyes and machine learning is becoming increasingly good at analyzing body posture from videos, which can detect gestures, motions, and stances which are important to analyzing learning. AI is rapidly improving in terms of speed, ability to be embedded into mobile devices, and amount of data that can be collected and processed. Economies worldwide are investing to accelerate progress. These emerging strengths of AI, coupled with the expanded scale of progress, contribute to the potential risks: Harm may result to specific people or populations, and students are vulnerable populations that demand our protection.

Figure 1: Chart of Strengths, Weaknesses, Opportunities, and Barriers. Notes were placed by individual experts in an introductory discussion. (Names have been obscured.)

The experts were also highly aware of limitations and weaknesses of today's AI. Some specific limitations that the experts called out include:

- Limitation and weaknesses of available datasets, which limit the resulting AI progress
- Presence of bias in data
- Lack of attention to equity and learner differences
- Tendency to fail non-gracefully
- Hard to integrate multiple AI capabilities

- Expense of building systems
- Limited, narrow, or insufficient interfaces to work together with people
- Lack of accessibility to learners with special needs, and lack of universal design in today's AI applications
- Limitations in what kinds of inferences can be made (e.g., AI is often better at correlation than cause and effect)
- Lack of transparency and violations of privacy
- Issues of fairness and accountability for harm
- Ethics not firmly established or adhered to
- Weak understanding of these limitations by the public (AI literacy)
- Humans are better at the task than AI—sometimes humans are needed to perform the task because of their ability to give a human touch and understanding

The experts cautioned against overestimating what is possible. They also warned against underestimating the potential errors in what an AI agent does given a slightly different context or input. A conventional distinction is between narrow or weak AI, which can undertake specific, well-defined tasks, and general or strong AI, which can perform in a context-sensitive manner and self-improve. Beyond general AI, there is "super AI"—artificial general intelligence, which would be better than a person at a wide range of unpredictable or novel tasks. Such artificial general intelligence is not within current reach of AI research and development.

It is tempting to mischaracterize today's narrow AI as super AI by not understanding the boundaries of what AI can do today. In one recent example, an educational technology (edtech) product graded students' short essays using what might appear to be strong AI, but a student discovered that the system was only looking for keywords and figured out how to always get a perfect score by adding lists of unrelated keywords to every essay (Chin, 2020). The assessment technology was so brittle that it might not even reach the bar for weak AI. There is no general or super AI yet (Fjelland, 2020). Experts are concerned about the tendency to overpromise what AI can do and to overgeneralize beyond today's limited capabilities.

Despite the risks and limitations of today's AI, the experts saw many reasons to continue work on opportunities to apply AI in education. Below we list some of the quick phrases the experts used to talk about opportunities; in the next section we elaborate on some of these more deeply, including:

- Offloading some of the cognitive load of teaching, helping teachers orchestrate classrooms, and extending what teachers can do;
- Analyzing learner performances in collaborative groups, simulations, and other rich contexts, recognizing additional forms of knowing;
- Adapting to learner variability in more ways and with more techniques than is now possible;
- Making invisible aspects of teaching and learning more visible, such as uncovering missed connections between different skills the student is learning in two different classes, to deepen support for learning across the related but separate contexts;
- Interacting with a student, privately, to provide as much individualized, guided practice as they need; and
- Supporting the long-term development of valuable expertise, beyond a single subject or context, such as expertise in writing across subjects.

When AI Has a Weak Grasp of Context: An Example of the Risks
A grasp of context is one way in which today's AI is less robust than it may seem. When used in the context for which it was intended, the AI agent may work well. When the context is expanded, errors may occur.

Today, learning technologies diagnose a missing element of student knowledge and direct them back to learn that missing piece of knowledge. Yet, a student may have a very strong math background, but poor English vocabulary. What if the AI misdiagnoses the student as having weak math concepts because they misunderstand aspects of the math word problems written in English? Or an AI might accelerate a student's pace because it sees rapid progress, but not know that an aide was working with the student during that particular session. What if the AI oscillates between speeding up and later offering remediation because it does not know the student only sometimes has an aide present?

Now imagine technology systematically applied to tens of thousands of similar students, all who are mistakenly held back or caught in an oscillating pattern due to a lack of understanding of context. When systematically applied, a flawed design might compound system problems or biases, such as issues that can unfairly impact dual language students. Although this example is only meant to be illustrative, it shows how a misunderstanding of context, amplified through the replication of a decision-making pattern across tens of thousands of students, could systematically harm a student group.

New Design Concepts for AI in Learning

Earlier, we introduced three ways to think about AI: as an ambitious leading edge of computing, as a set of specific capabilities that are rapidly advancing, and as a toolkit for synthesizing (and exploring) possible futures. We now discuss a fourth way to think about AI and the future of learning—as inspiring new design concepts. The new design concepts the experts discussed are not fully worked out today. Yet they are valuable to illuminate what may become possible and enable people who have different expertise to consider the possibilities together, thinking through both the opportunities and the risks.

The following design concepts were recurrent in our expert panel across scenarios and discussion. These design concepts expand beyond familiar ideas of technology supporting "personalized," "adaptive," or "blended" learning. The conventional metaphors may continue to be useful, but they also may limit how we envision futures of AI in learning. Here are five additional design concepts to consider.

The concept of **orchestrating** (e.g., Prieto et al., 2011) arose across both days of our conversations. Orchestrating is different from personalizing, adapting, or blending as it starts from a recognition of learning as a complex coordination of experiences that occur over time in a social community, where achieving learning goals relies on designing and modifying how people participate, how they move from activity to activity, and how they connect a flow of individual opportunities to learn into achieve more significant learning goal. Orchestrating sees AI as enabling students and teachers to link their participation in different groups and activities over time, towards a broader learning goal.

Experts described AI as **augmenting human intelligence.** This design concept has roots in the work of Douglas Englebart (1962), among others. It captures the quest for technologies that gracefully extend and strengthen human intelligence. In contrast to conventional computing where technology automates processes, presents information, or provides a tool, augmenting speaks to the possibility of machines that better understand a teacher or learner's goals, plans, intentions, and criteria for success and act in ways that help make people better at achieving their goals—more of a supportive partnership between people and machines.

Also, in the scenarios, experts described AI as **expanding naturalistic interactions.** This design concept captures our quest to escape the narrow confines of using a keyboard and pointing device to participate in learning. Learning has always been embodied and social, and now rapidly improving language, gesture, and other forms of AI-enabled recognition are enabling technology to become part of these conversations. Future technology may be able to respect more of what it means to be a human learner (e.g., Nathan et al., 2019). Likewise, for teachers, the design concept acknowledges how poor a fit technology has been to the performance art of leading a classroom in real time or how students express their emerging understandings.

A further feature in both scenarios is **broadening the competencies** that can be expressed and assessed. This captures the quest to go beyond what 20th century assessment technology measured well and also to feature learning experiences that involve collaborating on projects and other extended modalities of learning. For example, a recent workshop brought innovators together around the pressing need to map evolving competencies as students progress through higher education (Teasley & Kelly, 2020).

Another aspect of the scenarios is the possibility of **revealing connections and equivalencies**. There is a quest for AI to help us see important patterns that have eluded us so far. These could be connections in how a competency, like skill in writing, develops across many different domains, experiences, and time spans. For example, Pardos and Nam (2020) described how they used AI to discover non-obvious "equivalencies" across courses in a university course catalog, in terms of opportunities to learn similar concepts and skills in courses in different departments. Pardos et al. (2019) described how to empirically determine which courses at one institution (for example, a two-year college) could count as credit at another institution (for example, a four-year college)—a problem which is important to social mobility but intractable.

Due to circumstances, our convening brought together only experts in AI and research on learning. In the future, we and others need to bring educational policy makers, practitioners, innovators, and industry leaders into conversations as well. We notice that the available language for communication across sectors has too often been limited to the design concept of "personalized" learning and offer that design concepts of orchestrating, augmenting, expanding natural interactions, broadening competencies, and revealing connections may stimulate additional avenues of conversation with practitioners and policy makers.

Expanding Scenarios for AI

During the expert panel discussion, small groups developed scenarios to illustrate the new metaphors for AI and the attendant opportunities, risks, and barriers. We share two scenarios, which are synthesized from the small group discussions.

1) Classroom Orchestration

Marcelo Worsley of Northwestern University shared the following:

> Promote AI as a strategy that enables greater interactivity and embodiment across contexts. Learners are spending too much time alone, in front of computers, doing things that they don't care about. Moreover, much of the AI work is passively tracking students and providing no tractable benefits to them. AI should help us orchestrate learning experiences that are meaningful, exciting, and social. They should also help value multiple ways of knowing and being. Learners should be empowered … [and] should also be trained to design/construct custom AI systems that give them a useful lens for seeing and engaging with the world around them. Training students to develop AI systems will accelerate and diversify the ways that we think about using AI.

The experts viewed a "learning environment" as a physical and/or virtual space where interactions take place among learners, one or more teachers, various resources, materials, and technologies. Much of the public discussion of the technology has emphasized "personalized learning." In many scenarios, personalized learning can appear to be about a much simpler configuration. We observe that personalized learning sometimes emphasizes separating individual learners into their own learning experiences (their own "playlists" of learning activities). Personalized learning has been found to be promising, although with many caveats (Pane et al., 2015; Penuel & Johnson, 2016).

Blended learning (Means et al., 2013) refers to an overall learning experience that has a beneficial alternation between computer-based and teacher-led activities. Adaptive learning occurs when the pace, content, or sequence of learning opportunities is adjusted based on recent data about learners' performance (Brusilovsky & Peylo, 2003). Adaptation can occur within a learning task, in choosing next learning tasks, or by modifying the learning system more broadly (Aleven et al., 2016). Although these design concepts are important, they arose based on what has been possible with prior AI capabilities. As new AI capabilities come to the fore, the existing design concepts for personalized, blended, or adaptive learning will neither exhaust nor adequately describe the applications that become possible.

Experts in our workshop were highly attentive to the social aspects of learning and how learning often occurs across a set of different group sizes: individuals, pairs, small groups, and full classrooms. The concept of "orchestration" emerges from the need for a teacher to plan how learners will benefit from each group size and sub-activity, how these fit together to form a larger whole sequence of learning activities, and how to support the social transitions in and between the activities (Dillenbourg et al., 2013; Olsen et al., 2020; Van Lehn et al., 2016). In their conversations, key shifts in conceptualization of a learning environment included:

- From a sense of learning as a highly individualistic process to a focus on **social learning**;

- From a focus on meeting each learner’s need individually and in isolation to a focus on **learning communities that address equity and diversity**;
- From AI agents that react only to student keyboard and mouse input to a computer, to AI agents that **listen, observe, and interact naturalistically** with students, and that potentially take initiative to enhance small group dynamics; and
- From AI agents that are coordinated with a teacher only through a dashboard to **partnership between a teacher and an AI system** (e.g., enabling coordinated actions to support learning).

For the purpose of this illustrative example, consider a middle school lesson about the Renaissance period in which students can explore 16th century Venice with a virtual reality headset and an AI tour guide. As small groups of students together visit particular sites, they can ask the AI tour guide questions and it can prompt them to notice and discuss what they see.

Visiting Venice in small groups, however, is just a small part of the overall plan for a week of study. The teacher wants students to gather information during the tours and eventually write short essays and critique each other’s writing. She also wants the students to prepare a role to play in a historical debate relevant to 16th century Venice, and she hopes this reenactment will be a highlight of the week. On Sunday night before teaching this lesson, the teacher is excited about the possibilities but also nervous about how to keep individuals and groups moving through all the parts of the lesson plan to reach her ultimate goals. Not only that, she also wants to work toward more equal participation in classroom discussion (a few students tend to dominate) and more back-and-forth debate between students (currently, students tend to address her and not to react to what other students say).

The experts envisioned a future partnership between a teacher and an AI agent that helps with this lesson plan and with the teacher’s broader goals without taking control away from the teacher. They imagined how AI agents could provide many different types of help, and how the teacher might notice multiple benefits from the agent’s help, for example, saving time, feeling more aware of what all the students are doing and what they need, and also achieving ambitions for equal participation and in-depth conversations.

One type of help was support for forming groups. Who should go in each virtual tour group to enable the group members to ask good questions and make observations about old Venice? Perhaps the teacher, with knowledge of personalities, Individualized Education Plans, and recent behavior in class, might make some initial groups. Based on data from other group work sessions, an AI agent might help the teacher anticipate issues that could arise in particular groups. It’s possible that the teacher and agent could go back and forth with suggestions about how to rearrange groups or particular strategies that might be useful to help the groups succeed. It’s also possible that the teacher could ask the AI agent to monitor for particular behaviors or take particular actions in certain conditions (“let me know right away if Student K is talking too much” and “if Student S hasn’t asked spoken up, let’s consider pairing her with Student V and asking them to together come up with something to add”).

Another kind of help was nurturing better conversations among students. The experts discussed how hard it can be for a teacher to involve all students in a rich discussion during a lesson; more commonly, just a few students talk or students talk very little. The experts imagined how a future AI agent might listen to conversations in a classroom and offer gentle suggestions and nudges that help the teacher realize their goals for the conversation. For example, if students are not reacting to each other, the agent might suggest some sentence starters that could help, like "I like the idea of ____ but I wonder if you considered ____." The experts also imagined an AI agent that can review a classroom discussion with a teacher, perhaps noting the teaching goals that were and were not covered—and also who was and was not participating. This debrief might help the teacher make plans to cover additional material the next day or to reach out to particular students.

The experts also reimagined essay grading in a way that has more to do with organizing successful classroom flows. Rather than focusing on a final grade, an AI agent could suggest pairs of students that wrote about similar topics but each of whom missed some essential points. It could suggest students who could profitably work together to learn from each other's first drafts. Another kind of tracking that is hard for a teacher is to compare what the student mentioned in a small group discussion to what made it onto the written page; perhaps the student had some good ideas but needs additional prompting to write about them. The AI agent also might help the teacher prioritize the students they should spend time with, given the strengths and weaknesses of the initial essay drafts. The experts imagined the teacher informing the AI agent about the features of an essay that they would like to monitor—for example, perhaps a teacher is focused on how students structure comparisons in their writing. They imagined that the AI agent might learn the teacher's priorities and give more attention to these in early reviews of essays.

The experts also imagined how AI agents could help a teacher notice and respond to students' emotions. Today, learning analytics can pick up sequences of behavior that suggest a student may be frustrated, confused, or bored. In the Renaissance virtual reality simulation, a teacher may not be on hand to observe that a student tried to find out about what is in a particular home, but got frustrated when that was not possible in the simulation. Suggesting occasions to work with students around their emotions could be another way in which the AI agent helps.

Finally, the experts also recognized that the AI agent could help the teacher in ways that go beyond a particular class session or lesson plan. Agents might help a teacher know more about what is going on in a different class that is relevant to the Renaissance lesson—perhaps something from art class, for instance. Today, as many teachers work with coaches who help them with their teaching, an AI agent could suggest snippets of a classroom video that are especially worth reviewing and discussing with the coach.

Within five years, weak orchestration might be possible, involving offloading time-consuming and well-specified coordination tasks to an agent, such as forming groups, helping students work together to revise essays, and tracking the patterns in classroom conversations. Within 10 years, stronger orchestration might have a greater sense of partnership between a teacher and a supportive agent. Our experts noted that teachers want to save time and get help with burdensome tasks (like grading), but they also appreciate when their bigger goals (like more equal participation and richer conversations) are met. A stronger AI agent might both save time and subtly take action to help the teacher stay aware of student needs, to nudge or scaffold desired behaviors, to monitor progress towards bigger goals, and to help the teacher refine plans for the next day. Staying within the concept of orchestration, the AI agent's role might not be heavy-handed management of exactly how and when students learn, but rather a set of small actions that help teachers shape her classroom to fit her ambitions for what good teaching and learning looks like.

One of the concerns at the top of the list for the expert panelists is how to protect student privacy. In the orchestration scenario, the AI agent recorded conversations, interactions, and emotional data. How will the data be used? How will it be protected? Will it be saved? How long will the information be saved? Will it be a part of a student's record? Will records be shared with future teachers in the same school or will each teacher collect their own data? Will recommendations be made about a student's likelihood to succeed in college? As AI agents assess students, how do we ensure their privacy and that they are not used in unanticipated ways that may cause harm? How do students gain agency over their records?

In future classrooms, it could be common for AI agents to collect and use data associated with affective state through sensors such as pupil size, heart rate, and similar physiological data. Indeed, work on affect detection is well developed although challenging in terms of privacy and ethics (Greene, 2020). There remain important technical challenges and ethical policies to be determined; affect recognition may be controversial, and yet panel members note it is very helpful to know when a student is confused or frustrated.

One of the barriers to AI in classrooms is making it work for teachers. Panelists discussed how teachers go into teaching to work with students and not to work with technology. Panelists also discussed how they had collected data about what teachers want technology to do and not to do. They further discussed how different teachers may want different things: teachers are individuals just as students are. What might be important and useful to one teacher might be distracting and cumbersome to another, and an AI system should be able to support teacher preferences. Other questions remain: How much time will it take for a teacher to adjust all the parameters of AI systems? After routines are developed, will this save the teacher time or are they just getting more information to add to their already complicated job? How does a teacher decide what is important and useful?

In addition, the panelists who joined the breakout group focused on Teacher Professional Learning and AI discussed issues that might come up for the teachers. AI agents are likely to begin recommending professional development modules to teachers, but this can be problematic. How will a teacher feel about AI when it begins to tell them what they should do to improve their teaching? What if a teacher disagrees with what the AI agent suggests? Will the school administration know what the AI agent suggests for improvement? How will teachers' data be protected? There is FERPA for students, but we may need new policies to support teachers and to protect teacher data. How will a teacher learn to use the information from an AI agent in real time during a class? How long will it take to develop routines? As an intelligent assistant, it will need to provide information in a way that works and doesn't lead to cognitive overload. There is much research needed to understand how to incorporate an AI agent as a teacher assistant and the panelists discussed how important it was to have teachers involved in co-design. From the recordings an AI agent makes, it could highlight issues that a teacher needs to address in the classroom. The AI agent will reveal invisible patterns of who the teacher usually talks to and works with; this may reveal biases that the teacher could then address. They also noted that another use of AI for professional learning is to create a simulated teaching environment as a space to try out new pedagogical approaches (Cohen et al., 2020; Murphy, 2019; Peterson-Ahmad, et al., 2018).

Students were not part of our process in this expert panel, yet experts acknowledge the need for student voice as we design AI systems. Teachers report that students find technologies that advertise to them based on web searches as "creepy," so what are students going to think about AI agents that are with them every minute of the school day collecting their conversations in class and data related to their affect?

2) Transforming Assessment

The expert panelists raised a series of questions related to assessment: How could assessments help students learn more in real time? Could assessments be more relevant to real-world needs, like applying for a first job or to college? Could AI-based assessments reduce time spent in testing and free up more instructional time? Can we make AI-based assessments at least as fair as assessments are today? Will the results of AI-based assessment be explainable (or inscrutable) to students, teachers, parents and will they trust the scores?

Assessment can be viewed as a systematic, valid process of using evidence to support claims about a student's level of knowledge, skills and abilities based on a model or theory of competence and how people can acquire it (National Research Council, 2001). Conventionally, students experience assessments as tests on which they respond by solving problems or answering questions; the assessment is submitted and scored, and at some time later the student and/or their teacher gets a report that rates and/or ranks what the student knows and can do. Although this is less obvious to many test takers and score users, high-quality assessments have sophisticated design rationales that are grounded in an understanding of the skill or knowledge domain and how a student learns it.

The need for sophisticated and accurate theories or models of learning as a complex domain is a hinge point for the inclusion of AI, because AI technologies can be good at characterizing a complex array of features and factors that go into a complex decision. Although grading a simple word problem may seem easy, what worldwide mathematics educators would really like to know is often more akin to the complex judgement: "Does this student's mathematical reasoning enable them to formulate and use mathematical models to tackle real life problems?" (OECD, 2018). This is a much more complex judgement than whether a particular answer is right or wrong. The need for complex, sound judgements gives rise to possibilities for AI in assessment; the need for fairness in making such judgements gives rise to risks and barriers.

Importantly, experts in this forum did not see the role of AI in assessment as limited to what conventional assessments aim to do today; they did not see AI as merely making conventional assessment more accurate or efficient, but as supporting broader goals for assessment that are unmet by conventional tests. Indeed, the experts began by discussing powerful pressures on the assessment industry to change; in their view, although the overall paradigm for assessment has remained stable for a long time, pressure for change is now becoming more intense. Key contextual shifts away from conventional assessments discussed by the experts included:

1. From a focus on end-of-course outcomes to more rapid and useful **formative assessments** that can inform teaching and learning during the course.
2. From the activity of question-answering to capturing more realistic **performances or portfolios** of work, including over longer periods of time and across more settings.
3. From a narrower definition of academic achievement to a broader range of **competencies** that are valued in education, in society, and in workplaces.
4. From assessments that provide infrequent and isolated scores to **continuous assessments** that update how a student is progressing through learning trajectory over time.

The experts expressed their vision by three examples of goals for the field that could be achieved in future years. These are discussed below.

Within five years, AI-based assessment might produce a robust profile of a learner as a writer across contexts. This could include an ability to anticipate or predict how the learner might perform given particular writing tools, in a subject matter, in a new setting, or within a workplace. A specific scenario considered learner variability: a student who did their best writing work when they could first "talk" their ideas either into a text-to-speech tool that would capture them or in conversation with a peer. It also considered how opportunities to include diagrams, drawings, pictures, and other representations would be expected as part of the writing process. Further, it considered how the writing environment might include scaffolds, supports, or prompts—and note the kinds of supports that enabled peak writing outputs. Data could be summarized across many writing experiences in a student's life, rather than limited to writing as a component of the grade of an individual course.

Within 10 years, AI-based assessments might enable students to document competencies that are just beginning to be captured now, for example, their demonstrated skills in collaborating as part of a project team, a student's ability to use a simulation to investigate a scientific phenomenon, or their ability to design and engineer a tool to solve a challenge. Other scenarios build on examples today of game-based assessment, where roleplay in game offers students an opportunity to demonstrate what they know and can do in a more realistic environment that includes elements like real collaborators who work together on complex goals (compared to a conventional test).

With an uncertain timeframe, AI-based assessments might offer students considerable choice in what kind of experience they engage in to demonstrate their knowledge. Just as speech translation technologies can now translate from English to Spanish or recommendation engines can notice deep similarities, AI technologies might recognize "equivalencies" across contexts for showing a skill (e.g., different tools and settings in which a student might show they know how to conduct a scientific experiment). This could enable assessment designers to offer "baskets" of potential experiences that a student could choose among. Even though each student might choose different experiences for their basket, AI technologies might help in establishing the fairness of judgements of the competencies the students demonstrated.

While considering these nontraditional assessment scenarios, experts identified both benefits and risks of AI. Both students and teachers might benefit from a reduction in the activities that make conventional assessment painful: Students might appreciate less time spent taking tests, more flexibility in how they show their knowledge, and by being able to demonstrate a broader range of skills. Teachers might appreciate less time spent grading, more relevant and timely information to help them adapt their teaching to student needs, and alignment to broad and realistic educational goals that transcend a particular course—like preparing students to succeed in writing in a range of contexts. Society might benefit by documenting a broader range of abilities, learning how to support variable learners to express their skills, and more face validity in the relationships of the testing situation (e.g., a simulation) to the real-world situation.

Many of the weaknesses of AI that were introduced above were also identified as risks by the expert panel. With regard to bias and fairness, psychometrics as a field has long-standing techniques for how to reduce bias and establish fairness—but as yet these are poorly connected to emerging AI (Jones & Thissen, 2007). Likewise, the assessment discipline of evidence-centered design is highly applicable to constructing more ambitious assessment rationales, yet is rarely employed in AI products (Mislevy et al., 2017). Experts believe today's AI-based assessments are emerging without being systematically tested with broad enough populations or with enough expertise about learner variability and equity. They worry that if these flaws continued, AI might unknowingly perpetuate issues in today's large-scale assessments. Further, when assessment could happen at any time (such as during routine student collaboration on a project), concerns about privacy and surveillance emerge. And even if the above concerns are handled, the difficulty in explaining AI algorithms to the general public might undermine public confidence in the fairness of AI-based assessments.

In addition to the research and development needed on these risks, experts identified two major issues that could be a focus of research for the field. First, whereas conventional assessments are taken by the individual student in isolation, AI-based assessments may involve contexts in which students work in individual, small group, and large group modes. Students may have more choice, be able to access scaffolds or help, and take advantage of adaptive or personalized features of the assessment context. Students may be able to use complex tools during the assessment, like a simulation—and all students may not be equally familiar with how to operate within the simulation. Research is needed to reduce the combinatorial complexity of these expanded contexts for assessment, the impacts they have on the fairness of assessment results (did students have equitable opportunity to demonstrate what they know and can do?), and the reliability and validity of the results.

Second, experts shared that building meaningful teacher interfaces to assessments is a very hard problem. Experts believe that today's student information "dashboards" rarely meet a teacher's full needs and that they are cumbersome for teachers to use. Experts foresee AI doing more of the mundane aspects of assessment (e.g., grading), and freeing a teacher to focus more on planning how they will support student learning. But making the connections between the automation and actionable insight remains hard. In addition, teachers need to understand strengths and limitations of what an AI can do (e.g., coarsely score an essay's quality, but not analyze its meaning to the level that a human can) and they need to understand how the information can inform their instructional moves. Ultimately, the alignment and coupling of the roles of AI and teacher in a formative assessment system that is continuous, competency-based, and based around performances and portfolios is an unsolved problem worthy of much further research.

Discussion: Three Kinds of Risks

Along with their orientation to opportunities, as the experts discussed scenarios, they were continually attentive to risks. When AI is orchestrating learning, it could be amplifying inequitable participation. When AI is augmenting human intelligence, it could be amplifying a biased form of reasoning. When AI is expanding natural interactions, it might present obstacles to learners with certain disabilities, preferences or needs. Further, the designs for interactions may encode a historical bias that harms particular student groups. When AI is broadening our sense of measurable competencies, it may not give all people a fair opportunity to demonstrate those competences. When AI is revealing connections, these new findings could be used either to help or harm people. Within all the metaphors for AI and the future of learning, experts called our attention to the risks of amplifying the wrong signals.

We asked experts to return to the issues of risks several times during the convening. As we expected, data practices are a prominent risk. We currently lack strong enough policies, practices, and standards around the use of learning data in technology systems. Experts are very concerned with the issues around data, and because many of these have been described well elsewhere, we do not dwell on them here.

The experts informed us, however, that data is not the only important risk. Experts were also concerned with design flaws—designs may amplify the undesirable interactions among teachers, learners, and resources, and these could have negative consequences for individuals and contribute to societal challenges we'd rather overcome. Limitations to our design processes are equally important barriers to limitations in our ability to handle data responsibly. Generally speaking, experts recommended having more perspectives at the table during design and early-stage research on new designs to avoid these flaws. Waiting until the system is fully designed to obtain evaluative feedback from practitioners, students, and other stakeholders will not suffice.

A third kind of risk and barrier that experts called our attention to was the need to rapidly educate stakeholders about AI and to create participatory processes for their involvement. AI can have a mystique which is attractive, but ultimately becomes a barrier to trust and to the quality of designs. When the design process is opaque, it is hard for educational experts to participate in shaping them to fit educational needs and values. Experts often conveyed that participation of educators in decision making was needed to address the risks, but that a barrier to doing so was the need to bring participants up to speed in this complicated technical area.

Recommendations

At the close of our Expert Panel, we directly asked the experts to frame recommendations. They were prolific in contributing possible recommendations. These were captured on a whiteboard (see Figure 2) and also in an extensive discussion. We organized the experts' contributions into the seven major recommendations below.

1) Investigate AI Designs for an Expanded Range of Learning Scenarios

For many years, AI in education has deeply explored a few types of applications (e.g., intelligent tutoring systems). Yet as AI capabilities emerge, they are opening new possibilities that may prove equally or more important. Many important opportunities, such as AI agents to support learning in open-ended science inquiry environments, social studies simulation tools, or curricula to encourage design thinking, are still under-investigated. Likewise, AI learning scenarios may support better preparation for the workplace. Thus, it is important to describe a fuller typology of what is possible and to strengthen our knowledge of potential benefits, risks, consequences, and advances for each type of application.

2) Develop AI Systems that Assist Teachers and Improve Teaching

Experts were aware that today's AI systems have dashboards and other interfaces for teachers, but that these often fall short of being usable, friendly, or instrumental for teacher's work. They fall short of the idea of augmenting the teacher's intelligence and helping the teacher to grow, and often only make more work for teachers. Much prior investment in AI has been student-oriented, with not enough exploration of teachers' needs. Experts called for a vision of AI in the classroom that is more centered in assisting and supporting teachers as they orchestrate classroom experiences and for providing teachers with continuous opportunities to learn and grow. Experts also noted the need for much more research on how AI could help teachers learn and improve their teaching.

3) Intensify and Expand Research on AI for Assessment of Learning

Although AI already has been used in assessment of writing, science, and mathematics, much work is still needed to expand the bounds of the student learning activities that can be automatically assessed, the range of competencies that can be captured, and the breadth of assessment across settings and over time. Concurrently, ensuring that assessments are reliable, valid, and fair requires a new generation of analytic processes and capabilities to establish the quality of assessments, related to but not limited to existing psychometrics. Indeed, there is a need for new psychometrics and new AI techniques to co-evolve.

Figure 2: Initial Recommendations from Experts. These were later elaborated in discussions. (Names have been obscured.)

4) Accelerate Development of Human-Centered or Responsible AI

Limits in design processes and approaches can be as much of a barrier as issues with how AI collects and uses data. Included in this call is the need for AI that addresses learners with disabilities, learner variability, and the need for universal design for learning in AI applications. Although the experts did not recommend one name for the emerging discipline that is needed, it is clear that a pervasive design focus on human values is important (e.g., Shneiderman, 2020; Sambasivan & Holbrook, 2018). Learning engineering could incorporate this discipline into larger-scale products (Wagner & Lis, 2018).

5) Develop Stronger Policies for Ethics and Equity

In the expert panel discussions, there was a clear need to rapidly intensify the work to understand what core standards, guidelines, policies and other forms of guidance are for effective, equitable, and ethical practices in this emerging area (e.g., Kitto & Knight, 2019). Researchers doing the work have to participate in building the guidance that helps the field grow in a safe and credible manner. Practitioners, policymakers, and other stakeholders need to be equally involved. Policies must address the needs not only of researchers, but also of start-up companies and larger platforms and services. Policies and their implementation need to be transparent so that educators and the public can hold developers accountable.

6) Inform and Involve Educational Policy Makers and Practitioners.

Experts saw data and design risks with AI, but also the risk that would occur if educators were not well-informed. Newly emerging technologies may have quite different challenges from what educators are now familiar with (Reich, 2020). To participate in making decisions, building capacity among practitioners to understand AI is important. Capacity building is also important so that educators have the infrastructure to test and evaluate emerging AI and so they can inform design decisions. Schools and other educational institutions may need incentives to get more involved in evaluation and policies. Policy makers are learning about AI in general, but may be less aware of specific risks and barriers in education that need policy attention. Therefore, disseminating knowledge to help policy makers grapple with the issues is also important.

7) Strengthen the Overall AI and Education Ecosystem

Experts saw strong ecosystems of educational leaders, innovators, researchers, industry leaders, start-up companies, and other stakeholders as an important mechanism for shaping AI for educational good. Many of the dark scenarios, in contrast, involved poor information sharing or imbalances of power—and ultimately, one industry player acting alone. The expert panel called for stronger requirements for sharing data, for groups that are independent from industry to set standards, and to develop the field of learning engineering to be able to build safe and effective AI-based learning environments. These are all steps that could strengthen responsible use of AI technologies. They could also help start-ups to not only develop an exciting innovation, but also to address potential biases and risks. Experts also repeatedly called for more attention to building infrastructure for collaboration and techniques for partnerships among researchers, practitioners, policy makers, developers, industry, and other stakeholders.

The experts in our panel were researchers and it was therefore most clear how researchers might act on these recommendations. Researchers can write proposals to explore divergent emerging applications of AI in learning, and report in a balanced way on the risks as well as the possible benefits. They could organize knowledge of possible futures in a typology, to help others make sense of what is possible and to consider the risks of each approach. Researchers could focus more attention on how AI can support teachers and enable their growth, beyond what is possible with today's dashboards. Researchers can develop new forms of assessment to measure additional student competencies and can analyze validity and fairness. Research projects and labs can be hothouses for incubating new design practices that are responsible or human-centered. Researchers need to put more effort into developing guidance for their own community on how to develop equitable and ethical uses of AI, and could share their guidance to other communities.

Researchers can also strengthen the broader impacts of their work. They can seek to contribute more to learning engineering, where their knowledge may be put into use to build safe, effective, and equitable large-scale systems. They can get more involved in educating practitioners and policymakers about the issues of AI in education. They can refine the metaphors and other key ideas that help the public to make sense of what is possible and to understand the risks. They can participate in industry standards groups or educator associations to shape broad guidance. They also are part of an overall ecosystem and can become more involved in non-research forums where these issues are being discussed.

The experts also continually returned to their desire to see practitioners, industry, and policy makers become involved in this work. Particularly with regard to the ecosystem and building infrastructure and capacity, policy makers may have a big role. Market failures could occur because of poor availability or access to information among purchasers. Experts commented that the impacts of education within a lifetime are parallel in importance to impacts of healthcare, and thus policies that enable the growth of marketplace solutions that reflect the societal goals of education is important.

Policy makers also have a big role in supporting research that tackles issues on a horizon that is beyond what industry will invest in—as many of the issues are future-oriented ones. Researchers welcome more practitioner involvement in design issues. They often structure design-based research to invite co-design or partnership around exploring new designs. Practitioners, with their focus on centering students and their needs, are also clearly very important to defining what human-centered or responsible AI looks like in education. Finally, industry has often been a good partner with researchers in addressing challenging learning technology issues. Researchers believe that industry participation in future discussions on the themes of this report is important to a future in which AI is utilized for educational good.

References

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In *Handbook of research on learning and instruction*, 522-560.

Aung, A. M., Ramakrishnan, A., & Whitehill, J. R. (2018). Who are they looking at? Automatic eye gaze following for classroom observation video analysis. *Paper presented at the International conference on educational data mining (EDM) (11th, Raleigh, NC, Jul 16-20, 2018).*

Brusilovsky, P. & Peylo, C. (2003). Adaptive and intelligent web-based educational systems. *International Journal of Artificial Intelligence in Education, 13,* 156–169.

Carbonnel, J. (1970). *AI in CAI: An artificial intelligence approach to computer-aided instruction*. https://doi.org/10.1109/TMMS.1970.299942

Clabaugh, C., Mahajan, L., Jain, S., Pakkar, R., Becerra, D., Shi, Z., Deng, E.C., Lee, R., Ragusa, G., & Matarić, M. (November 2019). Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Frontiers in Robotics and AI.* https://doi.org/10.3389/frobt.2019.00110

Center for Innovative Research in Cyberlearning (CIRCL) (2020). Ambitious mashups: Reflections on a decade of cyberlearning research. [Report]. Digital Promise. https://circlcenter.org/resources/reflections-report/

Chin, M. (2020). These students figured out their tests were graded by AI—and the easy way to cheat. *The Verge.* https://www.theverge.com/2020/9/2/21419012/

Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher coaching in a simulated environment. *Educational Evaluation and Policy Analysis, 42*(2), 208-231. https://doi.org/10.3102/0162373720906217u

Dillenbourg, P., Nussbaum, M., Dimitriadis, Y., & Roschelle, J. (2013). Design for classroom orchestration. *Computers in Education, 69,* 485-492. https://doi.org/10.1016/j.compedu.2013.04.013

Englebart, D.C. (1962). Augmenting human intellect: A conceptual framework. Menlo Park, CA: SRI International. https://apps.dtic.mil/dtic/tr/fulltext/u2/289565.pdf https://doi.org/10.21236/AD0289565

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130–160. https://doi.org/10.3102/0091732X20903304

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, *7*(1), 1-9. https://doi.org/10.1057/s41599-020-0494-4

Greene, G. (2020). *The ethics of AI and emotional intelligence: Data sources, applications, and questions for evaluating ethics risk.* https://www.partnershiponai.org/wp-content/uploads/2020/07/PAI_The-ethics-of-AI-and-emotional-intelligence_073020.pdf

Holland, B. (2020). Artificial intelligence (AI) in K-12. Version 1.0 consortium for school networking. Consortium for School Networking and Microsoft. https://emma-assets.s3.amazonaws.com/paqab/71de919fd20c5e29fc83197664a59888/CoSN_AI_Report_2020_R8_4.24.pdf

Jones, L. V., & Thissen, D. (2007). A history and overview of psychometrics. In C.R. Rao and S. Sinharay (eds). *Handbook of Statistics, 26: Psychometrics* (pp.1-27). Amsterdam: North Holland. https://doi.org/10.1016/S0169-7161(06)26001-2

Kim, Y. et al. (2018). CIRCL Workshop: Robots, young children, & alternative input methods. https://circlcenter.org/wp-content/uploads/2018/10/WorkshopReport-RobotsYoungChildren-2018.pdf
Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology, 50*(6), 2855-2870. https://doi.org/10.1111/bjet.12868
Krumm, A., Means, B., & Bienkowski, M. (2018). Learning analytics goes to school: A collaborative approach to improving education. Routledge. https://doi.org/10.4324/9781315650722
Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). A learning management system-based early warning system for academic advising in undergraduate engineering. In *Learning Analytics* (pp. 103-119). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-3305-7_6
Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1)*,* 42-78. https://doi.org/10.3102/0034654315581420
Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901. https://doi.org/10.1037/a0037123
Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*(3), 1-47.
Minsky, M.L & Papert, S.A. (1968). Perce*ptrons: An introduction to computational geometry*. MIT Press.
Mislevy, R. J., Yan, D., Gobert, J., & Sao Pedro, M. (2020). Automated scoring in intelligent tutoring systems. In *Handbook of automated scoring* (pp. 403-422). Chapman and Hall/CRC. https://doi.org/10.1201/9781351264808-22
Mislevy, R. J., Haertel, G., Riconscente, M., Rutstein, D. W., & Ziker, C. (2017). Evidence-centered assessment design. In *Assessing model-based reasoning using evidence-centered design* (pp. 19-24). Springer. https://doi.org/10.1007/978-3-319-52246-3_3
Murphy, R. F. (2019). Artificial intelligence applications to support K–12 teachers and teaching: A review of promising applications, challenges, and risks. RAND Corporation, https://doi. org/10.7249/PE315.html
Nathan, M.J., Williams-Pierce, C., Walkington, C., Abrahamson, D., Ottmar, E., Soto, H. & Alibali, M.W. (2019). *The future of embodied design for mathematical imagination and cognition.* https://circlcenter.org/wp-content/uploads/2019/08/NathanEtAl-EMIC-Whitepaper.pdf
National Research Council. 2001. *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., and Glaser, R., editors. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press. https://www.nap.edu/catalog/10019/knowing-what-students-know-the-science-and-design-of-educational
OECD (2018). PISA 2021 mathematics framework (Draft). https://pisa2021-maths.oecd.org/files/PISA%202021%20Mathematics%20Framework%20Draft.pdf
Olsen, J. K., Rummel, N., & Aleven, V. (2020). Designing for the co-orchestration of social transitions between individual, small-group and whole-class learning in the classroom. *International Journal of Artificial Intelligence in Education*, 1-33. https://doi.org/10.1007/s40593-020-00228-w

Page, E.B. (2003). Project essay grade: PEG. In Shermis, Mark D., and Jill Burstein, eds., *Automated essay scoring: A cross-disciplinary perspective*. (p. 43). Lawrence Erlbaum Associates, Mahwah, New Jersey.

Pane, J. F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2015). *Continued progress: Promising evidence on personalized learning.* RAND Corporation. https://doi.org/10.7249/RR1365

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor Algebra I at scale. *Educational Evaluation and Policy Analysis, 36*(2), 127-144. https://doi.org/10.3102/0162373713507480

Papert, S. A. (2020). *Mindstorms: Children, computers, and powerful ideas.* Basic books.

Paquette, L., Baker, R. S., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. In *International conference on intelligent tutoring systems* (pp. 1-10). Springer. https://doi.org/10.1007/978-3-319-07221-0_1

Pardos Z.A. & Nam A.J.H. (2020) A university map of course knowledge. PLoS ONE 15(9): e0233207. https://doi.org/10.1371/journal.pone.0233207

Pardos, Z. A., Chau, H., & Zhao, H. (2019). Data-assistive course-to-course articulation using machine translation. In A. Rafferty and J.R. Whitehill (Eds.) *Proceedings of the sixth ACM conference on learning @ scale.* Chicago, IL. ACM. Pages 1-10. https://doi.org/10.1145/3330430.3333622

Penuel, W.R. & Johnson, R. (2016). *NEPC review: Continued progress: Promising evidence on personalized learning.* National Education Policy Center. https://nepc.colorado.edu/thinktank/review-personalized-learning

Peterson-Ahmad, M. B., Pemberton, J., & Hovey, K. A. (2018). Virtual learning environments for teacher preparation. *Kappa Delta Pi Record, 54*(4), 165-169. https://doi.org/10.1080/00228958.2018.1515544

Prieto, L. P., Holenko Dlab, M., Gutiérrez, I., Abdulwahed, M., & Balid, W. (2011). Orchestrating technology enhanced learning: a literature review and a conceptual framework. *International Journal of Technology Enhanced Learning*, *3*(6), 583-598. https://doi.org/10.1504/IJTEL.2011.045449

Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education*. Harvard University Press.

Richter, A., Gacic, T., Koelmel, B., Waidelich, L., & Glaser, P. (2019). *A review of fundamentals and influential factors of artificial intelligence. International Journal of Computer and Information Technology, 8*(4), 142-156. https://www.ijcit.com/archives/volume8/issue5/IJCIT080501.pdf

Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics, 42(1),* 85-106. https://doi.org/10.3102/1076998616666808

Sambasivan, N., & Holbrook, J. (2018). Toward responsible AI for the next billion users. *Interactions, 26*(1), 68-71. https://doi.org/10.1145/3298735

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495-504. https://doi.org/10.1080/10447318.2020.1741118

Simon, H.A. (1969/1996). *The sciences of the artificial* (Third Edition). MIT Press. ISBN: 9780262193740

Strain, D. (2020). New $20 million center to bring artificial intelligence into the classroom. *CU Boulder Today.* https://www.colorado.edu/today/ai-education

Suresh, A., Sumner, T., Jacobs, J., Foland, B., & Ward, W. (2019). Automating analysis and feedback to improve mathematics teachers' classroom discourse. In Proceedings of the

AAAI conference on artificial intelligence (Vol. 33, pp. 9721-9728). https://doi.org/10.1609/aaai.v33i01.33019721
Teasley, S. D., & Kelly, H. (2020). How a new focus for learning analytics could transform the relationship between learning and employment. In *Rapid Community Report* Series. https://repository.isls.org//handle/1/6848
Toyama, K. (2015). *Geek heresy: Rescuing social change from the cult of technology.* PublicAffairs. https://doi.org/10.13021/G8itlcp.9.2017.1853
U.S. Department of Education (2017). *Reimagining the role of technology in education: 2017 national education technology plan update*. Washington, D.C.: U.S. Department of Education, Office of Educational Technology. https://tech.ed.gov/files/2017/01/NETP17.pdf
Van Lehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://www.tandfonline.com/doi/abs/10.1080/00461520.2011.611369
Van Lehn, K., Cheema, S., Wetzel, J., Pead, D. (2016). Some less obvious features of classroom orchestration systems. In L. Lin and R. Atkinson (Eds.), *Educational technologies: Challenges, applications and learning outcomes* (pp. 73–94). Nova Science Publishers, Inc. http://www.public.asu.edu/~kvanlehn/Stringent/PDF/16OrchestrationChap.pdf
Wagner, E. & Lis, J. (2018). The rise of learning engineering. *eLearn, 8*. https://doi.org/10.1145/3265972.3242101

Appendix: Agenda for the Expert Panel Meeting

AI & The Future of Learning Expert Panel
Virtual Convening
Monday, June 29th, 11:00 am - 2:30 pm EDT
Tuesday, June 30th, 11:00 am - 2:30 pm EDT

Overview:
AI, machine learning, and related technologies will have powerful impacts on the future of learning. We know there are both potential benefits, but also considerable risks to be addressed. Although the greatest impacts are likely 5-10 years out, educational planning needs a long horizon to be effective, so the time to start is now. Through an invited, facilitated, day-long convening among experts in AI & Future of Learning, we seek to elicit deep contributions to two questions:

- What will educational leaders need to know about AI in support of student learning in order to have a stronger voice in the future of learning, to plan for the future and to make informed decisions?
- What do researchers need to tackle beyond the ordinary to generate the knowledge and information necessary for shaping AI in learning for the good?

Co-Chairs: Dr. James Lester, North Carolina State University; Dr. Jeremy Roschelle, Digital Promise; and Kasey Van Ostrand, Digital Promise

Hosts: Karen Cator, Digital Promise; Barbara Mean, Digital Promise; Missy Bellin, Digital Promise; Judi Fusco, Digital Promise; Bernadette Adams (Dept. of Education Project Lead), U.S. Department of Education; Jake Steel, U.S. Department of Education; Karen Marrongelle, National Science Foundation; Tatiana (Tanya) Korelsky, National Science Foundation; and Amy Baylor, National Science Foundation.

Expert Panelists: Dr. Russell Almond, Florida State University; Dr. Ryan Baker, University of Pennsylvania; Dr. Avron Barr, IEEE; Dr. Gautam Biswas, Vanderbilt University; Dr. Justine Cassell, Carnegie Mellon University; John Cherniavsky, National Science Foundation; Dr. Sherice Clarke, University of California, San Diego; Dr. Chris Dede, Harvard Graduate School of Education; Dr. Sidney D'Mello, University of Colorado, Boulder; Dr. Janice Gobert, Rutgers Graduate School of Education and Apprendis; Dr. Cindy Hmelo-Silver, University of Indiana; Dr. Susanne Lajoie, McGill University; Dr. Diane Litman, Dr. Rose Lucklin, University College London, Knowledge Lab; Dr. Maja Mataric, University of Southern California; Dr. Danielle McNamara, Arizona State University; Dr. Jaclyn Ocumpaugh, University of Pennsylvania; Dr. Amy Ogan, Carnegie Mellon University, Human-Computer Interaction Institute; Dr. Zach Pardos, University of California, Berkeley; Dr. Brian Smith, Drexel University; Dr. Kurt Van Lehn, Arizona State University; and Dr. Marcelo Worsley, Northwestern University.

Participants: Kevin Johnstun, U.S. Department of Education; Sara Trettin, U.S. Department of Education; Adam Safir, U.S. Department of Education; Christina Chhin, U.S. Department of Education; Edward Metz, U.S. Department of Education.

Agenda:

Day 1: Monday June 29th, 11 am EDT to 2:30 pm EDT Join Zoom Meeting: <removed>	
Start Time (EDT)	Topic
11:00 am	Welcome Please check-in on Mural. Opening Remarks *Jake Steel, Deputy Director, Office of Educational Technology - U.S. Department of Education* *Karen Marrongelle, Assistant Director of the National Science Foundation for Education and Human Resources*
11:30 am	Experts Share Breakthroughs and Barriers *Experts share their initial thoughts on the strengths, weaknesses, opportunities, and barriers related to the use of AI in education.* Link to Mural board.
12:30 pm	Break Link to Sign-Up Sheet.
1:00 pm	Experts Respond to Educator Questions *Experts respond to questions and concerns that educators have about the use of AI.* Links to Mural Boards: • Learning Environments • AI & Teachers • Assessment
2:00 pm	Closing & Preview of Day 2 Link to Mural board.
Day 2: Tuesday June 30th, 11 am EDT to 2:30 pm EDT Join Zoom Meeting: <removed>	
Start Time (EDT)	Topic
11:00 am	Welcome Back *Discussion question: What do you want to bring to this meeting today?*
11:30 am	Elaborate & Annotate Possible Futures

	Experts draft possible future uses of AI in education and outline the strengths, weaknesses, opportunities, and barriers of those uses. Links to Mural Boards: • Learning Environment: Multiple Contexts • Learning Environment: Social Scaffolding • Teachers & AI: Orchestrating Experiences • Teachers & AI: Teacher Feedback • Assessment: Market Basket • Assessment: Connecting Information
12:30 pm	Break
1:00 pm	Discussion of Barriers and Accelerators Across the Scenarios Discussion Questions: • Why isn't the field full-on tackling these now? What can be changed? • Do we need to rethink how industry, educators, and researchers relate and share? • How do we educate decision makers and purchasers? How do prepare school boards to think about this? • What kinds of policies (about data, IRB, etc) need to be addressed? • How do we develop more people who are highly capable of tackling this? • What are examples of a field that successfully got in front of a wave of innovation and how did they do it?
2:00 pm	Writing Exercise & Closing Remarks *Experts share recommendations for the most important barriers or accelerators to highlight in a subsequent report.* Link to Mural board.