Snow Li:

Hi everyone. Welcome. My name is Snow Li. I'm a senior program officer on the Early Learning Team at the Gates Foundation and I have the pleasure of kicking us off. Thanks so much for joining us today. At the Gates Foundation, we've been very fortunate to be supporting a team at SRI over the last few years and a couple projects really centered around classroom quality in Pre-K and in particular the ways that technology can assist in looking at classroom quality and can assist in teacher coaching. I wanted to share a few things about what really spawned this project. One of the learnings that the SRI team has shared with me over the last few years in this work is that the secret sauce to coaching is the relationship between a coach and a teacher and that there are ways that technology can allow coaches to do their work better when they're in a classroom helping a teacher and really looking at what children are doing.

And some of those opportunities are in doing video-based coaching. Some of it are in simplifying the workflow for coaches. Some are in making insights and more actionable for teachers. I was just in a conversation with a state agency leader last week actually in a state where there are a lot of rural communities, and she told me that her dream for the future is that one day her coaches could establish relationships with teachers and then be able to do virtual coaching so that every single teacher in the state had access to that. And so this is really the why behind this work.

For this particular project, the focus is to understand how AI can better enable classroom observations and coaching. When the SRI teams started this project, they started by talking to folks in the field who have been doing this work including entrepreneurs, systems leaders, practitioners and researchers, just to understand, "Well, what are the biggest leverage points for AI in helping folks understand classroom quality?" And then they came back and they hypothesized that, "Well, the AI is good enough to focus on two approaches. The first is to understand what's the instructional format that's actually happening in the classroom," and that's a quote that some of y'all who joined a little earlier may have seen on this slide.

"Are kids playing? Are they in transition? Are they in small group? Are they in whole group? Are they independently playing and working?" And that helps when I'm a coach because I want to understand what percentage of time is spent doing these things, "Where are kids throughout the day?" The second thing that they think that AI will be able to solve for is to identify within a video whether math or literacy or some other type of instructional content is being taught. And so that's going to be the focus of the presentation in the webinar today, is the SRI team led by Leigh Ann and Sarah and Gullnar will be talking about the findings from that research.

They will also be making the codebooks and the prompts that they had working with the generative AI available, publicly available, with the hopes that these resources can be used and picked up for others who are doing this work out there. And the last thing I'll say is that we have a couple of great speakers that are on the call today. So I just mentioned the project team led by Dr. Leigh Ann Delyser, sorry, I just said your name right, Delyser? And Sarah Gerard and Gullnar Syed. And then we also have presentations by Dr. Stephanie Curenton, Dr. Chavaughn Brown and Dr. Lauren Klein. And with that, I will pass it over to Leigh Ann, who will kick us off.

Dr. Leigh Ann Delyser:

Great. Thank you, Snow. And I want to just say that we were so excited to do this work and to lean into the possibilities for how AI could support the adults in Pre-K. And so we're going to kick this off with a little bit of background, share a bit about those use case interviews, give you some of the findings that we had in the fundamental algorithms we tested and then we want to make sure we have enough time for a rich panel discussion and Q&A from you. So very excited to get started. But first, I want to invite

people to say hi in the chat. Who's here? Where are you joining from and what kind of role do you play? Are you an instructional designer? Are you building tools of your own? Are you an early childhood educator or leader who's skeptical about AI?

And while we're getting to know each other a bit, we have a quick little poll for you as well, so we can get a feel for the folks who are here, whether or not you're excited by AI, concerned by AI or a little of both. And just to be clear, we're talking about AI for the adults in early childhood classroom settings, not AI for the children themselves. So we'll give another few seconds for voting. And I know I gave you two things to do at the same time, but I'm looking forward to seeing both those poll answers and some hellos. It looks like we've got some folks from Florida joining, the National Head Start Association. Welcome, Sierra, Jamie, Malika and to other folks who are jumping in and saying hi. We're really glad you're here.

All right, how are we doing on the poll? Getting to a response rate where we can look at the results and move on? So I think Keith who's running the poll, or Sarah, one of you've got to close it, so that we can see the responses. Thank you. All right, so we've got a mix here of folks interested in AI in early childhood settings. And I think it was really important for us in leading this work that we considered, "Is this a niche? Is this a place where we're on the cutting edge but nobody really wants?" And so it was important for us to start with a really market analysis here, right? So companies are investing in AI and education and these are from Holon IQ as they look at the market of EdTech and venture capital.

And so that companies with AI in their model like Leap Scholar, Magic School AI and Campus, you may have heard of them, they accounted for nearly half of all of the EdTech venture funds that were spent in quarter one of 2025. So it's clear that AI is making its way into the EdTech market. Additionally, Grand View Research shared that Pre-K is the fastest growing segment of EdTech market. And so it would stand to reason that we're going to see more and more Pre-K EdTech companies thinking about AI and thinking about uses of AI with children and adults.

And Pre-K and elementary school and middle school and high school look really different, right? New AI tools really expand on the recommendation and practice engines that we've seen in the one-to-one child model, but we're still seeing, in K through 12, AI really focused on serving the student. The chatbot is offering help during practice time and even supporting the teacher in lesson preparation. In Pre-K models, you'll see less of that device-to-child ratio. And so we focused on using video as a way to think about how we observe the classroom and support the teachers. And as you heard Snow early on, think about how we support the coaches and teacher pairing, so that those two humans can do their best job together the most efficiently and with the most accuracy.

So we need to think about AI not being a decider or a driver of student learning, but instead as a partner to those educator program leaders and curriculum providers. And so with all of that in mind, I'm going to turn it over to Sarah who's going to share a little bit about how AI might be used, specifically with classroom videos based on conversations that we've had over the past eight months. Sarah?

Sarah Gerard:

Thanks Leigh Ann. So one of the first things we did as part of this project, as Snow mentioned, was have a lot of conversations with Pre-K teachers and instructional coaches, EdTech developers and other researchers working in this space. So we asked the teachers and coaches, "What is challenging? What takes a lot of your time and what do you wish you could do faster so they can spend more time actually teaching and engaging with young children?" We asked the EdTech developers and the researchers about what they were working on, what felt promising and what were they hearing from educators about what they would like to see.

And after those conversations, our team identified four crosscutting themes that are really shaping the feasibility of AI-enabled tools in early childhood education. The biggest was trust. We heard that across the board. There are real concerns about video data security, privacy of students and of teachers, how is the data going to be used and whether or not any of these AI-enabled analyses are accurate. So our team really felt like addressing these concerns is the first step and that really needs to happen before any of these deeper investments and new tools and adoption of this in classrooms.

Next up, we heard that productive engagement was a really important theme. So there was a lot of interest in being able to assess classroom dynamics and thinking about those interactions between teachers and peers and materials and other students. So being able to get a better understanding of how those interactions are happening, which includes also the nonverbal interactions, can really help us get a better understanding of what's happening in a classroom and how that impacts children's experiences and would also allow us to have more effective teacher support.

Next step was interest in instructional grouping. So what we're meaning here is being able to look at an early childhood classroom and identify where children are in the classroom at that moment in time. Are they in a whole group book read aloud, doing some literacy instructions? Are they in small group receiving math small group time? Are they in one-on-one instruction or are they in the middle of a classroom transition between different activities? Being able to get a better sense of what those instructional groupings are can allow us to reduce that computational demand for video processing that would enable faster and more efficient analysis.

And then the last one here was seamless documentation. So we know that capturing student progress can take a lot of time and effort. We heard that time and time again from teachers and from coaches. "You're trying to jot down things in a notebook. You put it on a Post-it. Maybe you have a different system for recording this piece of data on the iPad over here. It can be a lot." So there's definitely potential here for AI to align images and videos and student work samples with developmental criteria and that right there can reduce some of that burden of documentation, preserve instructional time for actual teaching, and it also allows for potential to create some tools that can better capture what children can do.

These key themes led us to think about four main potential users of AI to support young children. So primarily, we were looking at the use cases for teachers and instructional coaches. And I'll also note that, of course, none of these users are children. These are all use cases to support the adults in the room. So for teachers, we were looking at the use cases to support collecting those classroom artifacts, like I talked about, collecting assessment data and log that in the appropriate place, thinking about in particular IEPs, DRDPs. These are both really labor and time-intensive data collection efforts that teachers and coaches express interest and be able to make that work faster.

There's also opportunities to help analyze assessment data and also share feedback with parents. And finally, also an opportunity to receive professional development resources through this mechanism. Next up, thinking about instructional coaches. So really I think the big use case we thought about a lot here was being able to speed up the process of identifying classroom video clips to target teachers' coaching needs. So by utilizing some of this instructional grouping we talked about, maybe identifying other pieces, coaches can really narrow the amount of time they have to sit either watching a teacher live or in person or clicking through a video to find the part that they're looking for and be able to grab that video clip faster and spend more time in the actual coaching relationship.

The other asset coaching support is being able to get a better sense of whether a curriculum or lesson plan is being implemented as it's designed. And then finally to also make it a little bit faster to provide some professional development resources to teachers. Next step, we also thought about curriculum developers. There are some opportunities here to think about how this can support curriculum fidelity

or teachers enacting a curriculum as it was designed. "Is that what was meant to be a small group math activity? Is that actually happening in a whole group? Do we need to think about how that's being implemented?"

And then lastly parents as another user here. So also we're hearing from some parents who are also teachers about some interest in being able to receive more detailed information from teachers about what their child is doing all day and also about their child's learning goals and their progression. And there are opportunities to think about how AI can make some of that work a little bit faster. So now, I'm going to turn it over to Gullnar who's going to share our progress on developing two fundamental algorithms to try to address some of those use. Gullnar?

Gullnar Syed:

Awesome, thank you, Sarah. So yeah, we're going to get into the nitty-gritty of our foundational algorithm testing, but first, we wanted to share why did we take this approach. So we picked this approach to developing fundamental or foundational models to serve as templates, to lay the groundwork for the application of the many use cases that Sarah just mentioned. And it's now mentioned at the top of our conversation, the models, frameworks or codebooks of the work is going to be shared publicly alongside white papers that detail really the technical pieces of the model. So the hope is that you all or any other researchers or folks interested could pick up this work for your own work and your own applications.

So the two automated supports that we developed were identifying instructional groupings, so things like small group, transitions, centers and math and literacy instructional content, so things like rhyming or counting. So for instructional groupings, our study team developed a framework that was used as a prompt to train ChatGPT to identify the grouping. We envision a potential use case for this type of work could be filtering video clips of transitions that might take longer than planned to highlight for coaches and teachers as moments to maximize learning time.

Second, we were looking at identifying math and literacy content and that was built upon existing NSF-funded models. The models were trained on math and literacy content standards that were aligned with the Head Start ELO framework and they were accurate in identifying math and literacy content in YouTube videos. So we apply this model to classroom video recordings. A similar use case for math literacy would be potentially during implementation of a new math curriculum, again filtering video segments of classroom recordings to maybe when math is occurring to understand and dig deeper into fidelity.

We also wanted to highlight at this stage that no AI was trained using the videos in this project. They were only evaluated by the models. So the videos that we used for this project, this study, will not be released to the public. We also use an enterprise or a paid version of ChatGPT so that any of the prompts and the videos are not captured and retained. So the free version, anything that you put is used to train the models, so we wanted to make sure that nothing was captured or stored to retain privacy of the folks in our videos. So first then, delving into instructional grouping. So we found that AI does a pretty good job in identifying instructional groupings when compared to human coders.

So just to talk about our process first, our human coders were trained to reliability on our framework. We annotated 14 hours of video and assigned a grouping to each video. Our AI experts then fed that framework as a prompt into GPT-4o Mini and GPT-5o Mini and is told to analyze the video and generate a category that describes what is happening in that video. So both a full framework, that's a full set of codes and subcodes and a simple framework which cuts some of the nuances around circle time specifically for whole group. They both performed reasonably well.

The best performing model is in this orange box on the screen here. It was a simple model that was run through GPT-5o mini where the accuracy between human annotations and the model was above 85%. And the bottom right here, we show an example of what the output looks like. So in this, case we see that a human categorized this segment as transition and independent time. And the GPT model also categorized it as independent time and gave a reason for that determination. To then share a snapshot of the results on our next slide, we have a confusion matrix on the right side. So this is showing the number of segments of predictive labels with the model labeled, so that's on the x-axis versus with the human labels, which is on the y-axis.

So ideally, we want the majority of numbers and a darker shading to occur in the diagonal in that orange from the top left to the bottom, right? So the most frequently observed in our dataset are centers free play, small group and whole group as shown by those three darker shadings on the diagonal. And there are less instances of independent time or transitions. So there are some outliers, especially between small group and free play where the model identified they confused some of those groupings. So the model identified small groups as free play pretty often.

So if we're thinking about small group and free play in ECE classrooms, those are pretty similar, right? So the difference really lies in the addition of instruction for small group versus free play where teachers can be floating around the classroom, monitoring in certain situations. So they do look very similar. So overall, the model, given that the model has a pretty high accuracy rate, we could say that it's pretty good, reasonable at identifying segments of video or coaching for a coaching use case, but if we're thinking of information on how much time potentially you spend in free play versus small group, the model might need refining. So for a full set of our results, be sure to check out our technical white paper which will be released in the coming weeks.

Next, I'm going to start talking through math and literacy. So we had a similar process where we had humans code. We coded 113 one-minute video segments. Our AI experts then bought that audio data from transcripts into an existing machine learning model, two models, so separately, one for math and one for literacy, and is told to analyze a transcript and generate any categories that are present. So in this table on the bottom here, we present precision and recall averages across the 10 literacy codes and five math codes. So you'll see that we set a false positive rate of 20% and this was chosen because we would have a higher amount. We would much rather ...

Yeah, we would much rather have a higher false positive rate because the presence of these instructional moments are much more varied in videos. So we use an example of cancer testing. So we'd much rather catch instances where there are no cancerous cells and not need treatment rather than having a go unnoticed. So by having that false positive rate higher, we're making the model more sensitive to detecting those codes. And this was selected based off of our review of the overall results and we didn't really find it. So as additional data becomes available or when these models get used with additional data, these numbers could change.

So that being said, the precision recall was higher for math than for literacy content. So that means, for math content, things like numerals adding and subtracting and subitizing were more readily identified by the AI models than literacy content. So next, we show two more confusion matrices of codes that we thought were interesting. So like I mentioned before, they're great for thinking about how the model got things wrong and where. So as you can see here in the darker shading in the top left, the model ... Well, we did a good job for when both the model and the human thought that there wasn't anything happening.

On the left-hand side in letter name, there were a couple of times that a human identified instruction on letters, but the model missed it. And for this code, this is one case where we think that in the original model that was trained on YouTube, there might have been more visuals in the background, maybe on a

whiteboard or a poster that help with that identification rather than something like a classroom video, which takes a broader view of the classroom and there's a lot more noise in that type of video. It might not be as focused on the materials that are present in a small group, for example.

For numerals on the right-hand side, the predicted label or what the model identified is more common than what the human identified. And this might be because students to teachers may have been saying numbers but not actual instruction with those numerals. And it also might be more challenging to know for certain if a teacher is pointing to a number of objects, for example, again, because of that YouTube video versus classroom video example. So what we can take away from these results in both of these examples and the majority of our codes across both math and literacy, the model picks up on the negative samples pretty correctly most of the time with different patterns across the individual codes. Again, for the full results, and please check out our white paper, which will be coming out in the next few weeks.

And now I'm going to hand it back to Sarah who's going to bring it back to implications for practice.

Sarah Gerard:

All right, thanks Gullnar. So just to recap those key takeaways from Gullnar, so it is possible to automatically identify elements of a preschool classroom recording using video or audio. And with some refinement, educators could use these models to identify math and literacy instructional content and also instructional groupings in those preschool classrooms. So this is a really exciting first step on the path to video filtering for teachers and coaches and school leaders who want to dive deeper into understanding these classroom interactions. However, none of this matters without some context, so we are now going to bring some context into the discussion. So I'm so excited to bring in three additional perspectives to share their experiences how AI already has or could be useful to them. We'll first hear from Dr. Stephanie Curenton, then Dr. Chavaughn Brown and then Dr. Lauren Klein. Stephanie, over to you.

Dr. Stephanie Curenton:

Okay, hello, everyone. So what I want to talk to you today about is a video-based classroom observation tool that centers equity, inclusion and individualized responsiveness as a foundational principle to classroom quality. The tool is called ACSES, Assessing Classroom Sociocultural Equity Scale and it was developed about eight years ago by myself and other colleagues who are Shana Rochester, Iheoma Iruka and Tonia Durden. And it was designed to be able to look at issues of classroom quality related to instruction and also socioemotional aspects of the classroom like culture, identity development, peer connections and harsh discipline and bias.

So the ACSES tool has six dimensions and 40 items that look for behaviors such as conversational back-and-forth exchanges, question types like open-ended questions and closed questions, peer collaboration and even use of multi-modality materials. So there's several research studies about this tool and they have found that it has concurrent and divergent validity with some of the more well-known classroom observation tools. And there's also studies that find that teacher scores on ACSES are associated with children's scores on executive functioning and math, especially when those classrooms are more diverse.

But presently, these ACSES observations are coded manually by humans, which is time-consuming, complex and can be cost prohibitive. So what I'm going to talk about today is how AI might be able to help with this observational coding in order to reduce the burden on graders and how AI might be able to summarize this data for coaches and teachers. So next slide. So what I imagine is being able to use AI

to score observation videos and then provide data-driven information about teachers' positive practices during learning interactions that can promote inclusion and equity and responsiveness.

I also imagine a system in which AI can be linked back to a comprehensive data system of evidence-based professional strategies for which the AI can suggest professional development resources to teachers based upon their access scores. And so this approach is using the data summation and organization power of AI to simplify life for humans, by saving time and by compiling and synthesizing professional resources. And in this way, AI is a helpmate to humans, but humans would still be driving the growth and improvement interaction with teachers during coaching.

So for example, AI would summarize the data about the observation into a report that then the human coach would be able to use to interact with the individual teachers. And in this way, the AI-generated data could be used as the foundation for those conversations between the teacher and the coach. But that dynamic goal setting and learning that happens during relationship-based coaching would still be reliant on human intelligence and human connection. So my idea is, by using AI for rating and for a compilation of individualized coaching resources for teachers, it makes it easier for the tool to be used more widely and it also makes the jobs of coaches more easier as well. And now, I would like to pass it over to my colleague Dr. Chavaughn Brown, who will talk about her work at AppleTree.

Dr. Chavaughn Brown:

Thanks so much and I'm so excited for this conversation. I think there's lots to learn today. I'm Chavaughn Brown. I am at AppleTree Institute for Education Innovation. Our work at AppleTree really sits at the intersection of research and practice. We operate early childhood schools in Washington, DC and we are also working to scale our Every Child Ready instructional model nationally. Across both spaces, our focus is on increasing children's access to developmentally appropriate high-quality early learning experiences and meeting the adults that serve them where they are in their professional growth.

As we move into this exciting growth phase for Every Child Ready, high quality implementation at scale while maintaining outcomes for children and educators is top of mind. We can go to the next slide please. We're really deeply interested in how classroom video and AI might help us better understand what's really happening for children during play, small groups, read alouds and other parts of the school day. And although we're still very early in our exploration of the tool, we're curious about how AI could help us spend less time in activities such as sorting through video and more time supporting teachers and that human work that Stephanie referenced that's so important and so impactful for child level outcomes.

So when we think about how AI can focus human effort to scale effective interventions, we're really looking at it through two lenses, which is depth and breadth of impact. And our directly operated schools in Washington, D.C. the deepest impact comes from strengthening teacher-child interactions and providing meaningful coaching. But coaching takes time. As many of us on the call know, reviewing video is incredibly time-consuming and not only reviewing the video, but identifying the key moments, preparing feedback, all of this is something that AI could take on and really help us spend less time sorting such as finding, as was mentioned earlier, where small group math happened or where language modeling was strongest. Less time sorting those clips and more time for coaches to spend on the human work such as building trust, planning goals and supporting instructional change.

And then at a national scale, the challenge looks a little bit different. So when we think about our breadth of impact, schools using Every Child Ready may not have the same level of touchpoints that we have in Washington, DC from a coaching perspective. So AI could really help surface patterns, implementation fidelity, lesson pacing, use of key practices across hundreds or even thousands of

classrooms that would allow AppleTree to offer actionable research-based insights without requiring a significant number of coaching hours across what we know to be very varied conditions of implementation across the country.

So this is some of the reason why we're so excited about today's conversation, and obviously, this work relates directly to some of what we're thinking about at AppleTree. So we're looking forward to continuing the conversation. And I will pass it over to Dr. Lauren Klein of Stanford Center on Early Childhood.

Dr. Lauren Klein:

Thank you so much. I'm happy to be here. So at the Center on Early Childhood, I work on this filming interactions to nurture development program or find where we use video coaching to give educators feedback on the interactions that they're having with children that are really responsive and supportive. So we'll record and securely upload videos in natural learning environments. A coach will then create a feedback video for that educator or caregiver. And as a result, we see some really exciting improvements in adult-child interactions and outcomes like child behavior, reduced challenging behavior, increased auditory comprehension, increased expressive language, some really exciting things going on there.

But as my fellow panelists have mentioned, that takes a really long time to create coaching videos. And when we start to think about how would we integrate AI into that practice, a big challenge, which Gullnar touched on too, is that there are some things AI does really well that bring this level of excitement and then other things that it does really poorly. I don't know who amongst you guys have been playing around with some of these tools and seeing transcription models don't pick up child speech that well, especially in a really noisy classroom. And video language models tend to make a lot of assumptions about what children and adults are doing when they're playing together, because there's not that much training data that these big companies are using in this specific context.

So when I joined the center, a lot of what we thought about was, "How do we leverage what AI can do so far, paired with this intervention that we've been running? FIND's been ongoing for over 15 years now. How do we pair those together to maintain the same exact product that teachers are seeing when they receive these coaching videos while enabling ourselves to deliver that more at scale, reach more teachers and children?" So if you'll go to the next slide, I'm not sure if my updated slide made it in there. Fantastic. So at Stanford FIND, we've actually developed our own video editor. We looked at a bunch of options that were out there.

Exactly, Jay, let's talk about that later. Thank you for your comment. So we found that there are video editors out there that use AI, but not really ones that combined AI and the editing component in the way that we needed it, so we built our own. Here on the left, you can see, and I know depending on the size of your screen, maybe you can see, maybe not as much, so happy to catch up on this later, first, a coach will select, "What type of responsive caregiving am I looking to give feedback on?" Our ML pipeline will then automatically look for suggestions of those moments upon the video being uploaded to our secure AWS platform.

And it will give those suggestions to the coach, but it will allow that coach to make edits because we know that AI is imperfect. You can edit the text, the timing that appears. A coach can add completely new clips and then we'll format that highlight video for you to put it into the format that we've been researching for those past 15 years. So through this approach, we've seen a time reduction of about 50%. Happy to discuss a little bit more about that, but we really are focusing on, "How can we leverage the benefits that AI brings to the table now while maintaining that fidelity with relying on that human

intelligence?" that Stephanie has mentioned. So if you are a practitioner, administrator, systems level player, please feel free to reach out and happy to field your guys' ... your all questions today.

Sarah Gerard:

Thank you so much. So now we're going to go ahead and pose a question to each of our three panelists. I think, Stephanie, if it's okay, I'm going to go ahead and start with you. So, Stephanie, you already talked really about some of this in your conversation earlier, but I'm just curious for a minute or two from each of you, how might you use some of the AI approaches that Gullnar talked about earlier in your work? Oh, Stephanie, you're on mute still.

Dr. Stephanie Curenton:

Sorry about that. Yes, so happy to start. And yes, my answer to that is very much about pointing back to some things that I talked about in my slides. So how could I use AI to focus human efforts to support scaling? So my heart lies really in terms of, like I said, using AI as a helpmate to make human jobs easier and be able to be done more quickly. And so I see AI as being able to do a lot of the grunt work that is associated with coding because it takes a long time to code the videos and then also doing some of that curating work with looking for resources that can help educators change their practices.

Sarah Gerard:

Yes, I think many of us who have spent time coding videos before know I agree with you, that can take a long time. Chavaughn, how about for you? How could you see some of these approaches being used in your work?

Dr. Chavaughn Brown:

Yeah, I think ... I mean, there's so many opportunities at a large scale. I think it really does bring ... The possibilities are endless when we think about understanding at scale what early educators are implementing in their classroom, especially with the commitment to high-quality instructional materials, what that means at the state and system level, where our theory of change holds true and where there are areas where the implementation is either too heavy of a lift or the conditions don't provide such environments for different parts of the model to be implemented.

So I think at a high level, AI has proven to have sensitivities to know what is happening in the classroom at a high level. I think at a coaching level, which is so important and what we're hearing from the workforce, all educators, regardless of their experience, are really looking for more tools in their toolkit, especially as they're receiving children with more varied ranges of developmental trajectories. AI could really say, "You had this very specific coaching goal. We're going to look across the whole coaching day to see all of the opportunities where you could have modeled language."

That can happen in every component of the day, so let's really hone in on that and let you know how it's going, let you know how you're progressing towards your goal as an educator, so that hopefully educators are being met with what they're asking for and we keep more high-quality educators in this space.

Sarah Gerard:

I love that. Lauren, how about for you?

Dr. Lauren Klein:

I agree completely with what Stephanie and Chavaughn have mentioned. Our main use of AI so far in the coaching space is looking for those moments where an educator has engaged in a specific type of interaction with a child, like naming what the child is doing or supporting and encouraging what that child is doing. And sometimes in the case of naming, that's language specific, but in supporting and encouraging what a child is doing, that could be verbal or nonverbal, especially with really young children. So we use currently primarily a text-based model and we have ongoing research for our video-based model to look at what do we think the child and adult are doing moment over moment throughout that coaching video, so that we can make those suggestions to the coach, which they can and then edit and show back to the teacher.

Sarah Gerard:

Great. Well, I think Leigh Ann, at this point in time, why don't we go ahead and flip this over to a bit more open Q&A and I'll let you be in charge of pulling the question and picking who gets to answer that one.

Dr. Leigh Ann Delyser:

Yeah, absolutely. We've got a bunch in the Q&A and I've been encouraging people to share in the chat, so I'm watching both of these things live. Please continue to share your questions. I think one of the themes I've seen both in a Q&A question and in one of the chat that came in is this notion of tech works effectively or really well in research settings, but when you deploy it into the field, it gets a little messy. It's not as accurate. What can we do to make that better? How can we account for the real-world circumstances and help improve our systems or the implementation of our systems in order to get the kind of outcomes that they were designed for?

And I think, Lauren, you shared this amazing video, that tool that you all have to help support coaches. Can you talk a little bit about how that works in real settings and ways in which you support educators and coaches in using the tools so that it works as well as your researchers who designed it?

Dr. Lauren Klein:

Yeah, absolutely. So the videos that we train on are all taken from natural learning environments, not in a lab setting. And so some of them are from home settings with parents and children. More recently, we've been using classroom videos, sometimes with multiple educators and almost always multiple children. So I would say it's a combination of using training data from the context where we're looking to evaluate and deploy. And then also I think sometimes it does need to be a process-wide strategy. So we are using a tool called IRIS Connect, moving forward to collect those teacher videos. Sometimes the teacher can also use a lapel mic to get better audio, so we have a little bit of better insight into what's happening.

And then I also think that human-in-the-loop approach rather than relying solely on AI to code the video is a really critical bridging step between now and when we see that next jump in AI performance.

Dr. Leigh Ann Delyser:

And we're working with imperfect tools, but I think it's really important to acknowledge that just as the technology is imperfect, so are our humans, right? And so you heard Gullnar talk a little bit about how we said a 20% false positive rate being tolerant to the AI making mistakes, but making mistakes in a way that gave us more information and not less, right? And so I think we have to own up to the fact that the technology's not perfect, but even a human coach, a human teacher isn't going to be perfect in their initial entry into a setting and we think about ways to support them in that learning. And so we have to

think about ways to support the humans who are using the technology to best make use of the ways that it does really well and the ways that it doesn't work so well.

I'm seeing some other questions here and Sierra put a great one into the Q&A and said, "There's strong evidence that teacher-mediated media experiences have a completely different developmental outcomes than passive solo viewing, like a teacher sharing a video with students, I think a learning video with instruction rather than students watching it alone. When thinking about AI and video insights, how should tools developers account for and design around this facilitation and co-regulation?"

Now there's the image in our minds that we all see when we go out into the world of children with iPads and they're getting content, but I think as we're thinking about this as AI that supports the adults. In what ways are we looking for the AI to support that human-human relationship between coaches and teachers as opposed to the AI just showing the teacher something alone? Stephanie, Chauvaghn, if the two of you could talk a little bit about that coaching interaction because we've also got some questions about like, "What is high-quality coaching and how supports from the technology or the AI really supports that interaction as opposed to like, 'Oh, this is just something a teacher can watch alone on their computer at the end of the day and get what they need out of it'?"

Dr. Chavaughn Brown:

Stephanie, would you like to kick off? I'm happy to learn from your insights first or I'm happy to share whatever you prefer?

Dr. Stephanie Curenton:

I'll start. So yeah, so I would say that definitely it is critically important that the coach and the educator are watching the video together and that the video, research has found, is more effective when it is that actual teacher watching herself. So she's serving as a mirror for herself in terms of those interactions. And teachers have been very responsive to that type of interaction with coaches and they have found that it's helpful. And there's some research that I think it's Jin Lee and his model that it's actually related to change that teachers feel efficacious around being able to change.

Dr. Chavaughn Brown:

I agree. So when we think about it at AppleTree, trust is so important. It's one of our core values. And I think AI gets at this perhaps indirectly in an unexpected way, which is if we bring in teachers from the design and discovery phase and ask teachers what they would like to learn about their practice and what their goals are, so it's more of a collaborative goal setting coaching experience, then the teacher is entering this video, which can also be rather nerve wracking to begin with, but entering this space differently and the coach actually modeling that reflective practice and being willing to go a different direction in partnership with the teacher in service of whatever goals they've identified as something that's centered to their practice, I think is a way where it will deepen not only the learning which is and more efficiently perhaps, but also can help foster and build trust in a way that perhaps the standard or traditional coaching relationship misses because there's less co-discovery of a tool.

Dr. Leigh Ann Delyser:

Yeah, I'm really glad you brought up trust, and I think if you were paying attention when Gullnar was presenting our results, you'll notice in the first instructional grouping results show, we asked the AI to explain why it made the choice it made, which I think would be a great cue not only to the coach and the teacher who are watching the video together and thinking about it from the instructional practice that the coach that they're orienting to, but what did somebody else notice or what did the AI notice that

wasn't familiar perhaps to the coach and the teacher? And that piece of explainability is something that's so often cited when we're talking about trust, right?

And so as we're thinking about these tools, what does it mean to foster trust with this technology? We're seeing questions here around equity, around students with different languages about inaccuracies depending on the scale and setting, which are, to me, all fundamental components underneath this like, "Do we actually trust this to do a good job?" So I'd love to hear quickly from each of you, what would it mean to foster trust? In the use and development of these tools, what features really support that?

Dr. Stephanie Curenton:

Okay, so I guess I'll start again. So for me, the foundation is that trust is really built upon systems, being transparent about people's data, how the data is being used and how the data is being stored. So if we're going to enter into these partnerships and these relationships in which we're going to be using data to help guide the coaching relationship, we want to make sure that it is clear and that it is explained to the educators and to the families how this data will be used and that it will only be used for that purpose and that it will be stored securely. And also taking that extra step as you all did with your study, making sure that it is stored privately, securely so that this data won't just be out there in the world.

Dr. Leigh Ann Delyser:

Yeah. Chavaughn, what does trust mean to you and how might the developers, researchers, practitioners here on the call help build it?

Dr. Chavaughn Brown:

I think trust is built from listening. I think it's built from understanding both an educator and then also the family and community perspective on whatever tool you're making. In the world where AI is evolving by the minute, I think there is a heightened awareness of, "What is the role of technology, especially in early education? Is my child going to be interacting with AI?" It raises lots of questions. And so making sure that it is fully communicated in a way that is framed ideally, authentically and that it's improving teacher practice and outcomes and experiences for learning. And if you see a drift from that, then you're reflective as a leader and you adjust and say, "You know what? I got that wrong. Our core purpose is this and this is how we're going to use this and we're not going to let it drift into something else that is perhaps a nice to have, but not core to who we are in advancing our mission."

Dr. Leigh Ann Delyser:

Yeah. Lauren?

Dr. Lauren Klein:

I couldn't agree with you more, Chavaughn. I think, when reflecting on my experience as a student in computer science, it can be so easy to create a prototype and throw it out there just because of all the tools that exist for coding and generating prototypes these days. And so it can be tempting to do that sometimes, but I think part of the reason I feel so lucky to be at a large research institution and specifically the Center on Early Childhood is their deep experience with engaging communities in the research and holding the same standard for program development that includes AI as we do for program development, that doesn't include AI.

Really what needs to be studied is the interaction between the output of whatever our tool is generating and the coach who's presenting it and that teacher, that always needs to be forefront. It's easy to focus sometimes on and important also to focus on the accuracy of the AI, but making sure that that community-focused and community-integrated design and research is always at the forefront, I would say is my main takeaway and suggestion from the work I've been doing.

Dr. Leigh Ann Delyser:

Yeah, And I think one of the things that's super important to us to think about is, do the datasets that we're using to train, to test, to evaluate these really represent that multicolored, multitextured classrooms that exist out in the real world, right? So as we're thinking about, are we ... The question came up about, "Well, what does multilingual classrooms look like?" right? As we're thinking about teachers who infuse their own personality into the way they instruct and the way they teach, how are we getting opportunities to observe that, include that in the training data and make transparent that we've included that?

So if people haven't seen or read the Coded Bias book by Joy Buolamwini about making sure that our datasets really represent the populations that we're trying to assess or determine, especially in things like video, we need to make sure that we are being transparent about the way in which we're diverse datasets to make sure that they match along with the end use, the far-down-the-field end use of these tools that we're producing. All right, so lightning round panelists. So I'm going to take one of the questions in the Q&A and turn it into a you-only-get-like-one-sentence-to-answer. So no taking a breath in the middle. It might be a run-on sentence, but you get one sentence to answer lightning round.

So the question was about workload and providing access, especially in rural sites. So they asked a lot of, "I wonder what," but you can't answer, "I wonder what," in one sentence, so I'm going to flip it a little bit and say, what are you most excited about for technologies like this one to help improve the access to coaching or coaches? And in what ways do you think ... Not and, or in what ways do you think that this might improve outcomes for teachers and students across? So one sentence, what are you excited about for something like these tools with AI and video in early childhood classrooms that could improve outcomes for students and teachers?

Dr. Stephanie Curenton:

Okay. So I'm most excited about the fact that being in this work, I have learned that there are safe, private, secure ways in which we can do this and explore this. So that has piqued my interest more and made me want to continue this work. And I think that that ultimately helps children and the educators because we're able to use this technology in a way that we can think is safe for them. That's the most exciting thing for me at this point.

Dr. Chavaughn Brown:

For me, this tool offers a sustainable solution for rural schools, for schools that are just getting funding for early education or for highly funded schools and allows the tool to be coaching to be more efficient and potentially impactful and protects and provides teachers with opportunity for coaching. Even in conditions where districts might be slashing budgets, coaching is often first to go. So this could sustain overcuts and maintain impact.

Dr. Lauren Klein:

I would say I'm excited for the potential of AI to support personalization and also rapid iteration, so that, as you mentioned, coaches can pivot as needed and meet the needs of many different types of sites.

Dr. Leigh Ann Delyser:

Awesome. Well, you all survived the lightning round and so I hope and I'm sure that I hear thunderous applause coming through from all of our virtual participants. And thank you so much for joining us today to help give context to these fundamental algorithms that we explored as a part of this project. So thank you so much to our panelists for joining us. This work was part of a larger initiative at SRI called RICiEE, the Reimagining Instructional Coaching in Early Education. Our final reports will be posted on the RICiEE site, and if you're interested in doing work like this with us, you can reach out to me or Sarah at the two emails that are on the slides and there's a link online.

Oh, thank you for telling me I'm not sharing the right screen. There we go. And there's a link online that we'll share as well where you can find out more about these tools. So our folks at SRI are really excited for the ways in which AI can be used to improve educational outcomes. We have a number of projects in both the Pre-K as well as the K through 12 and postsecondary space. And as always, you can get the latest news and connect with us through many of our sites. I especially want to point to our Student Behavior Blog that's really been posting some interesting interviews with educators and researchers. You can check out our podcast at the Homeroom. Find us on social media or just visit us on sri.com.

And we thank you for your time today. Extra thanks again to the members of the project team, to the Gates Foundation for the funding that supported this work and to our panelists who brought such excellent insights to the work that we're doing. Thank you everybody and have a great Thursday.