Pedagogical Responsibility in the Age of AI: Preservice Teachers' Reasoning About Automated Scoring Technologies

Jessica Watkins, Vanderbilt University, jessica.watkins@vanderbilt.edu

Abstract: The use of AI for automated scoring of student work has been positioned as both transformative for and damaging to ambitious and equitable teaching. Yet throughout these debates, teachers' voices have been noticeably absent. This paper considers how preservice teachers reason about the use of AI for assessing students' science models. Within a teacher education course on science modeling, students discussed current literature on machine learning methods for assessment. Drawing on discourse analysis, I highlight how students made sense of the ideological, axiological, epistemological, and ontological dimensions of their role as future educators alongside their meaning making about the use of AI. The students alternately position AI as a "useful tool" for, "replica" of, or "unproductive barrier" to teachers' work. With this analysis, I argue that supporting teacher learning about these technologies should involve helping them develop their pedagogical judgment about not just *how*, but *whether* these tools should be used.

Introduction and significance

There has been increased attention to developing and implementing machine learning tools for assessing student work. For instance, science education researchers have developed tools that automate scoring of complex tasks, such as students' arguments or models (Mao et al., 2018; Wilson et al., 2024; Zhai et al., 2022). Proponents of these tools advocate that they can reduce teachers' workload and support them to assign richer assessment tasks (Mintz, 2019; Zhai & Krajcik, 2024), while critical scholars point to the ways these technologies replicate or exacerbate inequitable learning environments for students from marginalized communities (Cheuk, 2021; Dixon-Román et al., 2020). Notably, throughout these debates, teachers' voices are rarely heard in the development of, concerns about, and research on these technologies.

The use of AI for assessing student work has been portrayed as a train that has "left the station" (Zhai & Nehm, 2023), contributing to a narrative of inevitability about these technologies. I argue that this narrative works in tandem with current pushes for the deprofessionalization of the teaching profession to erase teacher's agency. If tools like automated machine learning scoring are inevitable, then teachers do not have a choice in whether they use these tools. Indeed, research on teachers and AI typically focuses on whether they understand and "accept" these technologies (Nazaretsky et al., 2022; Sun et al., 2024), positioning them as passive recipients rather than agentive deciders of what tools they use to support their students' learning.

In this paper I argue the need for greater empowerment of teachers around these technologies, not just to participate in their design (Lin & Van Brummelen, 2021) but also to develop their pedagogical judgment around whether and how these tools should be used. I present an analysis of a class discussion with undergraduate and graduate science education students to show how these future educators make meaning about assessment and how they position machine learning (ML) assessment methods in relation to these meanings. This analysis unpacks how these students draw on ideological, axiological, epistemological, and ontological meanings about science teaching to consider the use of ML methods in science classrooms. I conclude with considerations of how teachers need opportunities to not just learn about these tools, but to develop their pedagogical judgment about their use, arguing that this learning needs to be embedded within the complexities of their practice.

Theoretical framing

Much of the work on teacher learning and AI has focused on how teachers understand particular concepts and make use of particular tools. A situative perspective on teacher learning emphasizes the need to examine how teachers make meaning with these concepts and tools in relation to other meanings they encounter about who students are, what it means to teach, and the nature of learning a particular discipline like science (Putnam & Borko, 2000). I draw on sociocultural theories of learning that position teachers' pedagogical reasoning as sites for negotiating these meanings and developing their pedagogical judgment about how to support students' learning in their classroom (Horn & Garner, 2022).

Teachers' pedagogical reasoning necessarily involves epistemological, axiological, ideological, and ontological dimensions about what it means to teach. Epistemologically, teachers' pedagogical reasoning can incorporate how they understand what counts as knowledge and learning (Elby & Hammer, 2010), for instance, what kind of knowledge is needed for teaching. Axiologically, teachers' reasoning can integrate what they value in learning environments, including the kinds of social relations, experiences, cultures, and environments they aspire to with their students (Philip, 2019). Ideologically, teachers' reasoning is shaped by their taken-for-granted assumptions about "how teaching works," for instance, how to understand the role of the teacher and relationship to students (Philip, 2011). These different dimensions are deeply entangled—ideologically positioning teachers in the role of "assigning points" is entangled with epistemological assumptions about the kind of knowledge assessment provides for teachers alongside values about what is worth teaching (and assessing). These dimensions also involve issues of ontology in teaching—is teaching an inherently individual endeavor or is it socially constructed by individuals and broader systems and structures (Horn, 2005). These ontologies matter for how teachers reason about and act in response to pedagogical problems, that is, whether they focus on changing their individual practices or school norms and policies (Chen & Marshall, 2018). I argue that these dimensions—ideological, axiological, epistemological, and ontological—of teachers' pedagogical reasoning will also inform how they understand the role of AI in their teaching.

Teachers' pedagogical reasoning is shaped by broader societal discourses and institutional logics about teaching and learning, particularly in the disciplines they teach (Watkins, 2023; De Lucca, et al., 2024). For instance, dominant narratives of science as meritocratic—only for the talented few—can shape how teachers understand their responsibility to support students to perform well on standardized tests and demonstrate their worthiness to pursue this discipline. Likewise, narratives about teachers and teaching, like those that deprofessionalize teachers' work, can shape teachers' judgment about their role to, say, deviate from curriculum or standardized assessments. Similarly, narratives about the inevitability of AI as a technology for schools or that schools are "behind the times" in the use of technology can shape how teachers understand their role and whether they have a choice in opting into these technologies. Given this perspective on teacher learning, to support teachers' learning about AI, there needs to be opportunities for them to make sense of these technological tools within the complexity of their practice. This paper considers what these opportunities might entail, examining how pre-service teachers reason about the use of ML methods in science assessment.

Methods

I analyze a classroom conversation from a course I taught in spring 2024, Modeling in Secondary Science. The goal of the course is to engage undergraduate and graduate science education students in expansive forms of science that broaden what counts in the discipline and that connect to power, identity, and historicity (Watkins, et al., 2024). Alongside their scientific inquiries on light and color, plant genetics and pigment, students read articles and watch videos that emphasize heterogeneity in ways of knowing and communicating in science (Bang et al., 2013; Brown, 2019; Kimmerer, 2013; Warren & Rosebery, 2011), as well as those that highlight the political dimensions of scientific pursuits (Learning in Places, 2020; Montgomery, 2020; Prescod-Weinstein, 2021). There were eight students enrolled in the course; three were undergraduate secondary education undergraduates, one was an education minor with the goal of teaching at the collegiate level, and four were Masters students.

This class discussion took place toward the end of the semester, after we had developed more expansive understandings of science modeling and reflected on how scientific models are situated in sociopolitical histories. The students read and commented on an article that highlighted issues of racial equity in using AI for science assessments (Cheuk, 2021). At the start of class I provided context for the article in relation to broader efforts to develop ML methods and we discussed advancements in the use of AI since its publication. I then introduced a review article that synthesized a current conversation in science education on the ethical dilemmas in using AI for assessment (Gouvea, 2024). This review summarized a research article and three commentaries. I presented key findings and limitations from the research article and claims from the commentaries, using them as the basis to form a series of prompts about which I asked students to form their opinion (Table 1). The prior week I had noticed students' energy was low and wanted students to move, so I placed paper at four corners for students to stand near to indicate whether they either strongly agreed, agreed, disagreed, or strongly disagreed.

My own pedagogical judgment in facilitating this discussion was informed by the idea that students needed opportunities to engage with competing arguments about the use of these technologies in education. I specifically sought to resist early ideological convergence (Philip et al., 2018) by praising students for taking different positions, framing the conversation as a place to play with ideas, and engaging myself by adopting multiple positions during the conversation, particularly when there were only a few students taking up an idea. During the conversation I remember being excited by the range of ideas that students expressed, but also cautious about the ways particular dominant narratives (e.g., the inevitability of AI, blaming individual teachers for systemic educational problems) seemed to go unquestioned by students. I debriefed the conversation with a colleague and reflected on the ways that this conversation about ML methods surfaced not only students' ideas about technology, but their meanings about teaching more broadly.

Table 1
Prompts for Four Corners Discussions

Prompt	Reference
This is compelling evidence that we can pursue ML assessments of student modeling.	(Zhai et al., 2022)
Tasks should be constrained to better assess using ML algorithms.	(Zhai et al., 2022)
The inability to engage with diversity undermines the utility of ML algorithms for teachers.	(Li et al., 2023)
ML algorithms may supplement teachers' work to allow for more complex tasks.	(Zhai & Nehm, 2023)
ML algorithms are no more biased than teachers and these are issues in all assessments	(Zhai & Nehm, 2023)
Researchers and educators should be investing resources in using ML for assessments.	

To analyze the conversation, I drew on these pedagogical reflections alongside discourse analytic tools (Fairclough, 2010; Gee, 1999). I transcribed the entire episode and watched the video several times with our research group. This team included graduate researchers, including one who was a participant-observer, as well as the director of the secondary teacher education program. These conversations surfaced analysis of the meanings teachers were making about the value of assessment and the kind of knowledge it provides. I also viewed the video with a research group consisting of other teacher education researchers and educational technology researchers. This conversation brought up ideas about how teachers understood the role of technology as entangled with their role as teachers, and what that meant for working toward educational change.

These conversations led me to consider how teachers were drawing on different dimensions of pedagogical meanings—ideological, epistemological, axiological, and ontological. I developed a detailed memo using discourse analysis to highlight these different dimensions in students' talk, while also analyzing the ways these dimensions were connected. To highlight the ideological dimensions, I looked for how students positioned teachers' roles and relations to students, and what that implied for the distribution of resources. To examine the epistemological dimensions, I analyzed students' talk for what assumptions they were making about nature of knowledge and knowing about dimensions of teaching—in this case, what kind of knowledge do teachers gain from assessment? Axiologically, I examined their meanings for what students foregrounded as values, particularly for assessments—for instance, whether the value of assessment to assign a grade or to allow for student creativity. Lastly, I looked at ontological dimensions based on whether they framed assessment as the practice of individual teachers or as constructed by individuals and broader systems and structures.

Findings

This analysis focuses on the class discussion for the first prompt, which lasted just over 10 minutes and had over half the students speaking. In this episode, I unpack how students are making meaning about what and how to assess learners' understanding and the role that ML methods might play in assessment. I argue that their meaning making about ML assessment methods involves their reasoning about the ideological, axiological, epistemological, and ontological aspects of teaching science.

The first prompt for the four corner discussions centered on the finding in Zhai et al. (2022) that showed relatively high levels of agreement between human coded scores and ML predictive scores on six items designed to assess an NGSS performance expectation that asked students to create models. I stated that the authors argue that this finding provides robust evidence for the usability of automated ML scoring methods.

Instr: Do you strongly agree, agree, disagree or strongly disagree? That this is compelling evidence that we should, can pursue ML assessments of students' models. (pause) I'm just excited that people went to different places.

Several students moved to the side of the room to disagree with this statement, with two students in between agree and disagree and two students on the agree side. I pointed out how students are scattered at different points in the room, not just in four corners, and expressed approval for having a range of perspectives. I asked students who agreed with the statement to contribute first.

Leo: Okay, so I only said agree to strongly agree because I mean, I think this is compelling evidence that we should start to pursue it because to say disagree to this, to me almost feels like a resisting the future, to where I think with these if, like teachers, educators started to use these in their classrooms. I think it would mean to be like, okay, all these people pass by machine learning. I need to go and look at all the people who did not pass and do my own assessment of the do not pass.

Instr: So it could be like a marker of like, the obvious ones you don't have to look at. But the ones that are the ones that pass, you don't have to look at it, then you could really dig in to understand the why of the people that didn't pass the machine learning.

Leo: Yes, because then I think you can like get to the root of like, what our common mistakes that are being made in the do not pass category. And also, if they should have passed, you could give them points. If the machine learning does not catch those-

Instr: Got it. So it's like gives me a nice tool for reducing the workload and allowing, like to dig into some of those mistakes.

Ling: I think he pretty much captured all I thought and I feel like it's pretty intuitive to me like that align with the teachers' grading according to the rubric. So I feel like that is pretty compelling to me. And now, but I just wonder like, it might have something to do with the rubric or the teacher themselves might have some like bias, but I don't know. But like that connection seems pretty like clear like that they I think do the same way as the teacher.

Leo's argument first rested on the idea that using ML assessment methods is inevitable and that not pursuing these technologies would be "resisting the future." He then imagined how these methods could be used to as a first check on the people who "pass." In this scenario, the role of the teacher in assessment is to determine who passes or not, with ML methods as an aid to make that more efficient. I revoiced Leo's points, shifting the categories from "pass" and "not pass" to those that are "obvious" and those that need "dig[ging] in." In doing so I positioned the role of the teacher as understanding the reason for why some students do not pass, with ML as a tool for determining who would get this attention. Leo then considered that ML could support another role for the teacher—identifying common mistakes. He elaborated that in deciding who does or does not pass, the teacher's role is to give points, not taking up my bid that their role is to dig into students' underlying reasoning. I highlighted an implication of Leo's idea that ML could be a tool to reduce teachers' workloads and re-emphasized the idea that teachers should "dig into" understand student reasoning underlying their mistakes.

Ling was the other student on the agree side of the room and she positioned herself in alignment with Leo's argument. She emphasized the evidence showing that that ML assessment results were mostly aligned with teacher's grading according to the rubric, arguing that it was intuitive that this evidence warrants pursuing these tools. In her stance on assessment, the benefit of ML methods is to use the rubric in "the same way" as teachers, even as she recognized the possibility of bias. In contrast to Leo, who positioned ML methods as a tool for teachers' assessment, Ling positioned ML methods as a replica of teachers' work.

At this point, I invited students to move positions if they were convinced by one another's argument. When no one moves, I asked students who are standing on the disagree side of the room if they want to share.

Mei: I can share my thoughts first. So like for me, like I think in a science classroom, when you do like encourage, like critical thinking and creativity, but I just like somehow think about the machine decoding sort of machine learning kind of like, disagree about the creativity, because, like, if they just follow the database, just follow the information they receive, they probably would not accept, like, any creative work, but like our human teacher can think, like when we grade our student where we can feel like learning about our students, but like, I'm just have a question here. Whether machine learning can still like keep learning when a school my students like our student work.

Mei started by naming her values for science learning: critical thinking and creativity. She positioned ML as "follow[ing] the database," which I interpret her to mean following predetermined instructions for assessing student work. Therefore, ML methods would conflict with her values of creativity because it would not accept work that does not fit within predetermined instructions. Mei explicitly positioned the role of the teacher as "learning about our students" and questioned the capabilities of ML to "keep learning." In this way she challenged the notion that ML can be a replica of teachers, because it cannot learn about students in the ways that teachers can and in ways that allow creativity since that would mean deviating from predetermined norms.

Esme: Yeah, I think like similar, like, the way I thought about it is like it can't capture the nuances that a teacher could. And if you do use it to sort of chunk away the obvious ones, then you're kind of widening the gap, where the education gap and and how, what the bubble, kids get focused on way more, and they get more resources and attention than the other students. And I'm not strongly disagree, because I mean, I think I agree with Leo in that. It's like, it's going to be used, whether we like it or not. And I think there could be some merits to it in terms of like, teachers are completely overworked and reducing workload. But I also think like, machine learning is going off of the norm. And same with like a bunch of human score. So obviously, they're going to be the same, but that doesn't mean that they're good at scoring. It doesn't mean that they're assessing student understanding well, just because they're in agreement, and I think it's easier like at teacher education programs like this and like in your own schools like for us to change how individuals think about their own students learning rather than change the whole system. Fast enough. So if, if this is like, too accessible, I think it could cause more problems in trying to move forward and more equitable ways.

Instr: So I hear- I just want to, I want to summarize what I heard really quickly, just to make sure, like, I want to pull out the points really fast, but I see other people are responding ready to respond. But I hear some of the echoing of Mei's comments. But then I also hear, like the process that was described of like, the people who pass the rubric, you don't have to look at them. We didn't when we focus on the students who didn't pass the rubric, that's already sending some messages about, what you call bubble kids are like, what messages are you think it's sending?

Esme echoed Mei's claim that ML methods cannot replicate teachers' work, adding that it cannot "capture the nuances" of students' thinking. She then challenged the earlier positioning of teachers' role in assessment as grouping students into "obvious ones" and not-obvious ones for their attention. She considered the potential implications of using this tool to label students and distribute resources, suggesting it might widen an "education gap." While she did not elaborate on what she means by this gap, it is possible she was referencing readings from earlier education courses that reframe achievement gaps as opportunity gaps for minoritized communities (Milner, 2012). She used the term "bubble kids" to talk about how resources might be distributed in ways that exacerbate historical inequities in resources and attention.

Esme agreed with Leo's claim that these methods seem inevitable in schooling and highlighted possible benefits for reducing teachers' workloads. However, she challenged the neutrality of ML methods. She described that ML methods are "going off the norm," similar to Mei's argument about how they "follow the database." Expanding on this idea, Esme not only positioned ML methods as restricted by these norms, but also humans as well. This challenged the idea that ML scoring the same as humans is good evidence for pursuing this technological tool, because there are systemic biases that can influence both teachers' rubrics and ML methods.

Esme introduced the idea that the broader educational system is implicated in assessment. She nominated teacher education programs and school cultures as sites for changing this system, starting locally by changing how individuals think about and interact with students. She contrasted this approach with changing "the whole system." While it is not clear what the "system" is, given that her argument was about disagreeing with the evidence to pursue ML methods, it seems plausible that she was positioning ML as embedded within the whole system that is challenging to change.

Given the multiple arguments in Esme's reasoning, in my revoicing I attempted to unpack different points for deeper consideration. I interpreted what she might mean by education gap, suggesting that this use of ML will send messages students receive about who is competent and who needs extra attention, setting up a hierarchy.

Esme: Um, well, I mean, I think, I think not only does it send messages, but it prevents the teacher from fully, like we've said, kind of learning our students and like, yes, some students are going to need extra support, and likely those who like, don't pass. But that doesn't mean you should take away support from kids who did pass it, because that also doesn't necessarily like we know, these tests aren't perfect. And that doesn't ensure their understanding for other topics in the class, or even this topic in itself. And you might not understand like, you might just assume that the way that they're thinking about it, just because they got it like right, is the way that NGSS wants them to think about it, but it might not actually. And that can cause problems when teaching like sort of a disconnect between teachers and students.

Instr: Got it. So it might get in the way of the relational building that you want to do to get to know your students and their ideas in a more holistic way, as opposed to in this way. And then one other point that you raised was, how does it this is engaged with the larger educational system? Like how is what is this going to provide? Oil? Like I'm thinking of like, like, how does this grease that some things and like, create friction and other ways to make things smoother to happen? Or, or was it harder to happen?

Esme: Was it last semester when like, there's a reading, where like, John, there's a boy named Jonathan, and he like was talking about who's explaining the sun and asking, like, questioning why the sun is considered nonliving and provided reasoning why he thought it would be. And so like, things like that those things would get lost and the nondominant ways of thinking, because machine learning is informed by dominant ways of thinking are going to be further ignored by machine learning. And they're more likely to get ignored by machine learning, because it's harder to harder to keep feeding that than it is to change individual teachers.

Esme acknowledged my interpretation about messages, but re-oriented back to the idea that ML should not determine distribution of resources. Instead, teachers need to use their judgment to determine who gets support, especially given the fallibility of assessments. Esme weaved together concerns about the limitations of narrow tasks and standards with relying on ML methods for assessing students' understanding. By making it easier for teachers to make assumptions about student thinking, ML can circumvent teachers' connections with students. In Esme's stance, ML methods are positioned alongside narrow tests as a barrier to teachers' work.

I amplified the end of Esme's contribution, revoicing that ML might prevent teachers' relationship building, keeping teachers from understanding their students beyond whether they obtained right or wrong answers on a test. I also connected back to Esme's point about how to make systemic changes. I explicitly positioned ML methods as an entity within the larger educational system, which can facilitate or hinder different processes. This move addressed not just how ML methods impact individual teachers' work, but how these methods might shape broader values and processes.

Esme then provided an example from an earlier reading to illustrate how ML methods might have systemic influence. Highlighting an episode from a reading in our course (Warren and Rosebery, 2010), in which a student, Jonathan, questioned the categorization of the sun as nonliving, she anticipated that reasoning like this would "get lost" with the use of ML methods because they are informed by dominant ways of knowing. Like Mei, Esme positioned ML methods as less capable of learning than teachers. But while Mei's comment referred to the ways teachers need to learn about their students, Esme talked about how teachers need to change to fix the inequity in the broader educational system.

Sarah: I was just gonna say that, like, I came over here, because like, I kind of get stuck on the whole, like, kind of what Esme was saying. But it's been a little bit of a different, like, I get stuck on like, the students that are needing as extra support in the classroom are having to do their modeling in a different way than the other students. So like, special ed child-, students, students, like English language learners, and all those that are also built into the classroom. And like, how will the use ML assessments, assess those students modeling when they're probably going to look different than the standard general education students? So like, if a student and I think someone brought it up in like, the readings as well, like if a student is using translanguaging, in their models? Is this assessment going to be able to like pick up on that? Or is it going to mark it as like incorrect or whatever? And then if a teacher doesn't take the time to go through the ones that were marked as incorrect, are they going to catch that? Or is it going to just kind of make that kid less feel less than because they used a different language even though they're doing it correctly? So that's just that's why I'm not fully compelled yet.

Sarah's comment nuanced the labeling of pass/not pass by considering that those labels might obscure what supports students need. She instead drew on other labels (special ed, ELL, general ed) to remark on how ML will accommodate these pre-existing categorizations. In contrast to Esme, she was not challenging the system that creates these labels, but wondering about how ML methods might incorporate them since these delineations are "built into the classroom." She used the example of translanguaging, which she positioned as a practice that ELL students might use as compared to "general education students." She questioned whether these practices will be recognized by the system and considers the impact on students—not just whether they pass or not, but how they might experience schooling. She introduced the idea that teachers might be fallible, i.e., not take the time to check the ML designations, which was part of Leo's proposal for how ML methods might work.

In this episode, students negotiated what it means to assess students' scientific models and the role of the teacher and ML methods in this work. *Ideologically,* they considered whether and how ML should play a role in distributing resources, in particular who gets graded points or who the teacher devotes attention to. They challenged the idea that using ML methods to reduce teachers' workload is a neutral endeavor since it will determine who gets more pedagogical opportunities for learning, which has historically been an inequitable distribution for learners from minoritized communities. *Axiologically*, the students explicitly grappled the values of science learning and what that means for assessments. For instance, Mei argued explicitly to value creativity and diversity in ways of knowing, raising challenges for how ML methods might recognize nondominant ways of knowing, communicating, and relating to the natural world. *Epistemologically,* the class negotiated what kind of knowledge assessments provide. When students are advocating that there is strong evidence for the use of ML methods to assess learners' models, they positioned assessment providing knowledge about what points to award, while students who disagreed shifted to treating assessment as providing knowledge about students' reasoning. *Ontologically,* the students considered how to conceptualize ML methods within teachers' work, whether it is just a tool for individual teachers or whether it implicates a larger educational system and set of values.

Discussion

This analysis highlights the ways that students alternately positioned ML methods as a "replica" of teacher's work, a "useful tool" for assessing student models, and a "unproductive barrier" for teachers. At first glance, these categories echo Tegmark's (2018) three different visions of AI: (1) a Utopian belief in technology as solving human problems, (2) a pragmatic perspective advocating for controlled development to balance risks and benefits, and (3) a techno-skeptic stance that doubts the capabilities of these technologies and highlights the harm they can cause. Examining these categories in students' meaning making, however, points to the nuanced ways these views about these technologies intersect with how these future educators make sense of what it means to teach. If students are reasoning that assessment is primarily about "assigning points," then it follows that AI technologies should be evaluated by how well they serve as a replica of teachers' grades and make grading more efficient. If they are considering assessment as a mechanism to provide useful data to inform their teaching, then it might make sense to position teachers "in the loop" with AI (Li et al., 2024; Mosqueira-Rey et al., 2023) to tailor their teaching with automated scoring, obtaining benefits from these technologies while also maintaining human control. Alternately, if students are foregrounding the ways assessments—and teaching more broadly—need to be responsive and relational (Kang et al., 2023; Krist, 2024) so that teachers can get to know students' strengths, creativity, and diverse ways of understanding the world, particularly if they are from nondominant communities, then it is sensible to recognize that offloading scoring to AI would be a barrier to teachers' work.

These findings point to the importance of considering teachers' learning about AI technologies in the complex contexts of their practice. While prior research has examined how to support teachers' understandings and acceptance of these technologies, there is a need for an integrated approach in which teachers' learning about these tools in relation to their pedagogical responsibility, that is, what they feel obligated to in their teaching, whether that be ethical principles, institutional norms, or situational constraints (Horn & Garner, 2022). While this conversation focused on concerns around automated scoring technologies on teachers' pedagogy, other issues might inform how teachers conceptualize their responsibility, including the histories of these technologies with respect to racialized surveillance (Benjamin, 2019) and the drain on Earth's resources (Crawford, 2021).

For teacher educators, this analysis shows that engaging teacher candidates with ongoing debates about the use of AI, in this case ML scoring methods, can surface diverse meanings about the work of teaching, which are important to engage with in teacher preparation. In this episode, the students consider how their future role as science educators might involve assigning points, learning about their students, and promoting creativity and diversity in ways of knowing. These different meanings also have implications for how resources are distributed, what ways of knowing and communicating are valued in disciplinary spaces, and how assessment is related to teaching and learning. They also negotiate issues of ontology, specifically how to think about teaching, namely assessment, as individual's isolated work versus work that is situated in and constructed by broader systems. While these aspects of students' reasoning might emerge in other conversations about teaching, making sense of a potential disruption in the standard "ways of doing things" in schools provides a fruitful opportunity for classes or learning communities to reflect on what, how, and why to teach.

Conclusion

The diminished teacher voice in AI educational research intersects with the ways that teaching is increasingly devalued in this era of deprofessionalization. Yet, just as AI research has had a "simultaneous reliance on a concept of the human that it seeks to undermine" (p. 80, Baker et al., 2023), AI research in education has had a simultaneous reliance on a concept of what it means to teach that it also seeks to undermine. Developing better automated scoring requires understanding the work of assessment by ambitious and equitable teachers, while also seeking to shift their roles away from that work (Gentile et al., 2023). However, in response to expanding efforts to incorporate technology in classrooms, Kerssens and van Dijk (2022) argue the need for pedagogical autonomy, both at the levels of institution (schools) and individual (teachers). Professional pedagogical autonomy refers to the "degree[s] of freedom teachers have to perform pedagogical practices and make pedagogical decisions… independent of digital education platforms." (p. 286)

Teacher education has an important role to play in realizing professional pedagogical autonomy. To understand their pedagogical responsibility in relation to AI technologies, teachers need opportunities to make sense of how tools like ML scoring can be used, not just to learn about their features and affordances, but also to consider the ways that they might constrain their teaching or cause harm to students, particularly those from marginalized communities (Cheuk, 2021; Shaw et al., 2024). Rather than proclaiming inevitability, it is imperative to remember that teachers, just as others, have a choice in these technologies (Benjamin, 2024), and therefore we should position them as agentive deciders of not just how, but whether these tools are used in their classrooms.

References

Baker, B., Mills, K. A., McDonald, P., & Wang, L. (2023). AI, Concepts of Intelligence, and Chatbots: The "Figure of Man," the Rise of Emotion, and Future Visions of Education. *Teachers College Record, 125*(6), 60–84.

Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2013). Desettling expectations in science education. *Human Development*, *55*(5–6), 302–318.

Benjamin, R. (2019). *Race After Technology*. Polity Books.

Benjamin, R. (2024). The New Artificial Intelligentsia. *LA Review of Books*. https://lareviewofbooks.org/article/the-new-artificial-intelligentsia/

Brown, B. (2019). *Science in the City: Culturally Relevant STEM Education*. Harvard Education Press.

Chen, G., & Marshall, S. A. (2018). Mathematics teachers' learning to enact change in oppressive systems. *International Conference on Urban Education Conference Proceedings*.

Cheuk, T. (2021). Can AI be racist? Color-evasiveness in the application of machine learning to science assessments. *Science Education*, *105*(5), 825–836.

Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

De Lucca, N., Watkins, J., Swanson, R. D., & Portsmore, M. (2024). Examining interactions between dominant discourses and engineering educational concepts in teachers' pedagogical reasoning. *Journal of Engineering Education*, *113*(1), 30-52.

Dixon-Román, E., Nichols, T. P., & Nyame-Mensah, A. (2020). The racializing forces of/in AI educational technologies. *Learning, Media and Technology*, *45*(3), 236–250.

Elby, A., & Hammer, D. (2010). Epistemological resources and framing: A cognitive framework for helping teachers interpret and respond to their students' epistemologies. In L. D. Bendixen & F. C. Feucht (Eds.), *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice.* (Issue 2, pp. 409–434). Cambridge University Press.

Fairclough, N. (2010). *Critical Discourse Analysis: The Critical Study of Language*. Routledge.

Gee, J. P. (1999). *An Introduction to Discourse Analysis*. Routledge.

Gentile, M., Città, G., Perna, S., & Allegra, M. (2023). Do we still need teachers? Navigating the paradigm shift of the teacher's role in the AI era. *Frontiers in Education*, *8*, 1161777.

Gouvea, J. S. (2024). Ethical Dilemmas in Current Uses of AI in Science Education. *CBE—Life Sciences Education*, *23*(1), fe1-3.

Horn, I., & Garner, B. (2022). *Teacher Learning of Ambitious and Equitable Mathematics Instruction: A Sociocultural Approach*. Routledge.

Horn, I. (2005). Learning on the Job: A Situated Account of Teacher Learning in High School Mathematics Departments. *Cognition and Instruction*, *23*(2), 207–236.

Kang, H., Talafian, H., & Tschida, P. (2023). Expanding opportunities to learn in secondary science classrooms using unconventional forms of classroom assessments. *Journal of Research in Science Teaching*, *60*(5), 1053–1091.

Kerssens, N., & Van Dijck, J. (2022). Governed by Edtech? Valuing Pedagogical Autonomy in a Platform Society. *Harvard Educational Review*, *92*(2), 284–303.

Kimmerer, R. W. (2013). *Braiding Sweetgrass*. Milkweed Editions.

Krist, C. (2024). Striving for Relationality: Teacher Responsiveness to Relational Cues When Eliciting Students' Science Ideas. *Cognition and Instruction*, *42*(2), 207–242.

Learning in Places Collaborative. (2020). *Framework: Ethical Deliberation and Decision-Making in Socio-Ecological Systems*. Learning in Places.

Li, T., Miller, E. A., & He, P. (2024). *Culturally and linguistically "Blind" or Biased? Challenges for AI Assessment of Models with Multiple Language Students*. 1323–1326.
Li, T., Reigh, E., He, P., & Adah Miller, E. (2023). Can we and should we use artificial intelligence for formative assessment in science? *Journal of Research in Science Teaching*, *60*(6), 1385–1389.
Lin, P., & Van Brummelen, J. (2021). Engaging teachers to co-design integrated AI curriculum for K-12 classrooms. In *Proceedings of the 2021 CHI conference* (pp. 1-12).
Mao, L., Liu, O. L., Roohr, K., Belur, V., Mulholland, M., Lee, H.-S., & Pallant, A. (2018). Validation of Automated Scoring for a Formative Assessment that Employs Scientific Argumentation. *Educational Assessment*, *23*(2), 121–138.
Milner, H. R. (2012). Beyond a Test Score: Explaining Opportunity Gaps in Educational Practice. *Journal of Black Studies*, *43*(6), 693–718.
Mintz, S. (2019). Can Technology Make Grading Fairer and More Efficient? *Inside Higher Education*. https://www.insidehighered.com/blogs/higher-ed-gamma/can-technology-make-grading-fairer-and-more-efficient
Montgomery, B. L. (2020). Lessons from Microbes: What Can We Learn about Equity from Unculturable Bacteria? *mSphere*, *5*(5), e01046-20.
Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review*, *56*(4), 3005–3054.
Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, *53*(4), 914–931.
Philip, T. M. (2011). An "Ideology in Pieces" Approach to Studying Change in Teachers' Sensemaking About Race, Racism, and Racial Justice. *Cognition and Instruction*, *29*(3), 297–329.
Philip, T. M. (2019). Principled Improvisation to Support Novice Teacher Learning. *Teachers College Record*, *121*(4), 1–32.
Philip, T. M., Gupta, A., Elby, A., & Turpen, C. (2018). Why Ideology Matters for Learning: A Case of Ideological Convergence in an Engineering Ethics Classroom Discussion on Drone Warfare. *Journal of the Learning Sciences*, *27*(2), 183–223.
Prescod-Weinstein, C. (2021). *The Disordered Cosmos: A Journey into Dark Matter, Spacetime, & Dreams Deferred*. Bold Type Books.
Putnam, R. T., & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have to Say About Research on Teacher Learning? *Educational Researcher*, *29*(1), 4–15.
Shaw, M. S., Toliver, S. R., & Tanksley, T. (2024). The Internet Doesn't Exist in the Sky: Literacy, AI, and the Digital Middle Passage. *Reading Research Quarterly*, *59*(4), 690–705.
Sun, F., Tian, P., Sun, D., Fan, Y., & Yang, Y. (2024). Pre-service teachers' inclination to integrate AI into STEM education. *British Journal of Educational Technology*, *55*(6), 2574–2596.
Tegmark, M. (2018). *Life 3.0: Being human in the age of artificial intelligence.* Vintage Books.
Warren, B., & Rosebery, A. S. (2011). Navigating Interculturality: African American Male Students and the Science Classroom. *Journal of African American Males in Education*, *2*(1).
Watkins, J. (2023). "That is Still STEM": Appropriating the Engineering Design Process to Challenge Dominant Narratives of Engineering and STEM. *Cognition and Instruction*, *41*(4), 405-435.
Watkins, J., De Lucca, N. A., & Pao, S. R. (2024). Fostering expansive and connective sensemaking with preservice secondary science teachers. *Journal of Research in Science Teaching*, *61*(5), 1093-1133.
Wilson, C. D., Haudek, K. C., Osborne, J. F., Buck Bracey, Z. E., Cheuk, T., Donovan, B. M., Stuhlsatz, M. A. M., Santiago, M. M., & Zhai, X. (2024). Using automated analysis to assess middle school students' competence with scientific argumentation. *Journal of Research in Science Teaching*, *61*(1), 38–69.
Zhai, X., He, P., & Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*, *59*(10), 1765–1794.
Zhai, X., & Krajcik, J. (2024). *Uses of Artificial Intelligence in STEM Education*. Oxford University Press.
Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, *60*(6), 1390–1398.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2145260. I thank TLES team members and the Joy and Care research group for their thoughtful comments on these data. I acknowledge that Natalie Rae suggested the Cheuk paper for the course years ago. I am also grateful for the Modeling students for sharing their ideas and perspectives with us.