When AI Breaks, Teachers Repair: Pedagogical Repair Work in Situated Classroom Practice

YAXUAN YIN, University of Wisconsin–Madison, USA
SHAMYA KARUMBAIAH, University of Wisconsin–Madison, USA
YUJIN KIM, University of Wisconsin–Madison, USA
DEVANSH SAXENA, University of Wisconsin–Madison, USA

As AI tools increasingly enter classrooms, teachers are not simply adopting new technologies. They are navigating the breakdowns these systems introduce into everyday educational practice. Yet as AI advances in education, Responsible AI values are often treated as properties to be optimized or audited within systems, and how these values are actually experienced, negotiated, and sustained in the breakdowns teachers face in everyday classroom practice remains unclear. In response, our work adopts a sociotechnical perspective that asks who is affected when breakdowns occur and how teachers respond to them in practice. Drawing on eight focus groups with 26 K–12 teachers, our findings show that pedagogical repair work emerges around three recurring sites of breakdown: misalignments between algorithmic measures and teachers' judgments of learning and engagement, shifts in classroom authority and responsibility, and disruptions to relational and growth-oriented pedagogy. By foregrounding breakdown and repair as central to Responsible AI in education, our work argues for AI systems and governance approaches that support teachers' ongoing efforts to align technology with the lived realities of classroom life, rather than relying on technical safeguards alone.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; • **Social and professional topics → K-12 education**.

Additional Key Words and Phrases: Responsible AI, AI in Education, Pedagogical Repair, Human-AI Interaction

ACM Reference Format:
Yaxuan Yin, Shamya Karumbaiah, Yujin Kim, and Devansh Saxena. 2026. When AI Breaks, Teachers Repair: Pedagogical Repair Work in Situated Classroom Practice. In *The 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT '26), June 25–28, 2026, Montreal, QC, Canada.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3805689.3812281

1 Introduction

Artificial intelligence (AI) is increasingly promoted as a solution for advancing teaching and learning [36, 52], with tools being introduced for a wide range of classroom purposes, including scaffolding science learning [42, 44], automated assessment [15, 40], and classroom management [2, 7, 12, 64, 66]. Yet alongside this enthusiasm, educators are recurringly expressing apprehension about how these systems affect their students and their practice [17, 19, 34, 65], such as plagiarism, over-reliance on AI-generated content, and the development of students' social and emotional skills. The accumulation of these risks has led to a surge in the development of

Authors' Contact Information: Yaxuan Yin, University of Wisconsin–Madison, Madison, Wisconsin, USA, yaxuan.yin@wisc.edu; Shamya Karumbaiah, University of Wisconsin–Madison, Madison, Wisconsin, USA, shamya.karumbaiah@wisc.edu; Yujin Kim, University of Wisconsin–Madison, Madison, Wisconsin, USA, yujinkim@cs.wisc.edu; Devansh Saxena, University of Wisconsin–Madison, Madison, Wisconsin, USA, devansh.saxena@wisc.edu.

This work is licensed under a Creative Commons Attribution 4.0 International License.
FAccT '26, Montreal, QC, Canada
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2596-8/2026/06
https://doi.org/10.1145/3805689.3812281

frameworks and toolkits aimed at facilitating the practical design of Responsible AI (RAI) [62], which advocates for the responsible design, development, and use of AI systems around values [38, 45, 63].

Yet the challenge of responsible AI in education is not only a matter of design and governance, nor of developers and users alone. A growing body of scholarship in science and technology studies (STS) has argued that modern societies systematically overvalue innovation while rendering invisible the ongoing work of keeping systems functional, equitable, and meaningful over time [47, 58]. This line of work calls attention to a broader and often overlooked cast of actors whose labor is equally central to how technological systems function in practice [16, 23, 55]—in our study, that cast is teachers. Central to this perspective is the recognition that breakdown is not exceptional but ordinary: it is precisely at moments of breakdown that the values embedded in sociotechnical systems become visible and contestable [23, 24]. This perspective has been applied in prior HCI work to study domains such as health care [60], social work [48], and public health [14]. We extend this line of work to K–12 education at a particularly consequential moment: AI systems are still being introduced into classrooms rather than fully stabilized in everyday practice, which means that understanding how teachers anticipate breakdowns, assess risks, and negotiate how technologies should be used offers a rare opportunity to surface the professional and ethical labor that responsible AI requires in practice—and to inform design, governance, and deployment decisions before patterns become entrenched.

Building on this, Houston et al. [21] show that repair work is not merely technical but ethical and political—that values do not simply precede breakdown but actively emerge through it, as practitioners decide what should be preserved, protected, or changed as systems enter and disrupt everyday practice. It is precisely because values emerge at moments of breakdown and repair, rather than being settled at design, that RAI principles become a productive analytic lens: they provide the vocabulary for naming what is at stake in those moments, tracing how teachers negotiate RAI values not in the abstract but in the concrete pressures of classroom practice.

Guided by this perspective, we ask: *How do K–12 teachers anticipate breakdowns around AI in classroom practice and engage in pedagogical repair?* To address this question, we conducted eight focus groups with 26 K–12 teachers, using a web-based deliberation tool to facilitate discussion about emerging classroom AI tools, anticipated breakdowns, and the repair work teachers expect to perform as these systems are introduced into everyday practice. Through our analysis of *anticipatory breakdowns* and *pedagogical repair work*, we show how teachers become everyday repairers—not of hardware or software per se, but of the values, meanings, and relationships that are disrupted or reshaped by these systems.

Our results illustrate important implications for practitioners and stakeholders: (1) the importance of grounding AI metrics for learning, engagement, and risk in teachers' contextual knowledge and classroom realities rather than treating them as fixed or universal, (2) the need to align accountability in classroom AI with teachers' professional authority rather than displacing it, and (3) the recognition that relational and care-centered pedagogy is a core ethical dimension of AI in education that current Responsible AI frameworks often overlook. Looking ahead, we develop *pedagogical repair* as a Responsible AI orientation, pointing to future research that centers the ongoing, situated work teachers do to interpret, adapt, and sometimes resist AI systems in order to sustain trust, fairness, and meaningful learning in classrooms.

2 Related Work

Our work sits at the intersection of two bodies of literature. We first review concerns and potential harms of AI in education and why Responsible AI frameworks have struggled to address them in practice, showing that RAI values are context-sensitive, entangled, and difficult to operationalize in care-oriented settings (Section 2.1). We then turn to repair, care, and relational work in sociotechnical systems (Section 2.2), which provides the theoretical framework through which we examine the ongoing, often invisible labor through which practitioners

uphold, contest, and reconstruct RAI values in everyday practice, including teachers' anticipatory negotiations with AI systems in education.

2.1 Responsible AI in Education as Situated Practice

The integration of AI into educational technologies has been driven by the promise of reducing teacher workload and enhancing student learning. AI systems are increasingly introduced into classrooms to provide feedback to learners outside of school hours, assist with lesson planning, automate grading, and handle other administrative tasks that often burden educators [8]. Yet educational technologies are seldom adopted in straightforward ways: teachers act as intermediaries, weighing institutional mandates and AI-generated recommendations against their professional expertise and knowledge of students [20], while AI systems increasingly function as infrastructure embedding assumptions about learning, risk, and success that may not align with the lived realities of classrooms [61]. These risks range from reinforcing stereotypes [4] to amplifying existing social inequalities [11], with consequences that are especially acute for students from marginalized communities. For instance, Lee et al. [34] pointed to potential sources of bias across each stage of the LLM lifecycle, leading to inequalities reinforced by the predominance of English-language research and training data [65].

In response to these concerns, governments, industry, and academia have published Responsible AI (RAI) guidelines that outline values considered essential for ethical AI development [13, 26, 38, 45, 63]. These efforts are reflected in education-specific policy frameworks as well: the U.S. Department of Education [8] emphasizes evidence-based pedagogy, "humans-in-the-loop", and commitments to privacy, explainability, and non-discrimination, and UNESCO guidelines and national AI in education strategies similarly reflect a shared recognition that AI systems in schools require principled governance [29, 56]. Yet existing RAI frameworks concentrate attention on design and governance while rendering invisible the ongoing labor of the teachers who actually sustain, adapt, and contest AI systems in everyday practice. What remains underexamined is not what RAI principles prescribe, but how teachers navigate the value tensions those principles name when AI systems enter their classrooms—and what professional and ethical work they perform when those systems fall short.

Prior work has found that RAI values are interpreted and prioritized differently across specific use cases, domains, and individuals' demographics, social and political backgrounds, and ideology [25, 57, 68, 69]. In education particularly, Yin et al. [69] showed that teachers' RAI priorities are shaped by their pedagogical orientations and classroom context: critical pedagogy is associated with greater emphasis on transparency and safety, whereas teacher-centered and behaviorist orientations are negatively associated with fairness and safety. Moreover, teachers' priorities—and the tensions between values such as performance and transparency—shift across grade levels and institutional contexts [68]. Together, this body of work underscores that RAI values are not independent dimensions to be optimized in isolation but are deeply entangled and context-dependent, a challenge that becomes especially consequential in care-oriented settings like education where relational and pedagogical realities resist standardization. What is needed is a framework capable of making visible the situated work through which these values are actually upheld, contested, and reconstructed in practice—and it is here that repair theory provides a productive lens.

2.2 Repair, Care, and Relational Work in Sociotechnical Systems

A growing body of work in HCI and science and technology studies has emphasized that technological systems are sustained not only through design and innovation, but through ongoing practices of repair, care, and relational work [47, 58]. This critique has particular resonance in institutional and care-oriented settings, where the people responsible for maintaining systems and adapting them to human needs are rarely recognized as doing technical or innovative work at all. Jackson's notion of *broken-world thinking* deepens this perspective by framing infrastructures as always in partial states of breakdown, requiring continuous human effort to keep them

functional and meaningful in everyday life [22, 24]. From this perspective, repair extends beyond fixing technical failures to include adapting systems to local conditions, preserving relationships, and negotiating which values should be upheld as technologies are taken up in practice.

Recent research on AI in workplaces has similarly shown that frontline workers play a central role in shaping how systems function in real settings. Studies of algorithmic tools in organizational contexts highlight that successful AI adoption depends on workers' values [5, 9, 39], social dynamics [6], and domain expertise [48]. Rather than acting as passive implementers, workers actively interpret, adapt, and sometimes resist AI systems in order to align them with their professional goals and ethical commitments. Much of this maintenance and adjustment work remains unseen despite being essential to system functioning [16, 47, 55]. This invisibility becomes especially consequential as AI systems enter high-stakes institutional settings, where the costs of misalignment fall on vulnerable people rather than on those who design the systems.

These dynamics are especially pronounced in forms of care work, including education [17], social work [49], and primary care [39]. In these settings, practitioners are responsible not only for efficient task completion but also for maintaining relationships, responding to individual needs, and exercising moral judgment. When AI systems introduce values such as efficiency, standardization, or optimization, they can come into tension with commitments to equity, dignity, and individualized care [30]. As a result, care workers often engage in emotional, interpretive, and organizational labor to make new technologies workable for the people they serve. In healthcare, clinicians routinely reinterpret or override electronic systems to ensure that patient care reflects lived conditions rather than automated prompts [60]. In social services, caseworkers negotiate algorithmic risk tools to preserve professional judgment and fairness [48, 49]. Houston et al. [21] conceptualize these moments as *values in repair*: sites where practitioners actively decide what should be preserved, adapted, or resisted when technologies fall short of the relational and ethical demands of care work.

3 Method

We conducted eight in-person focus groups [33, 50, 53] with 26 K–12 teachers between July and November 2025, recruited through professional networks and educator listservs. Due to the constraints of in-person study, participants were all based in the Midwest, United States, but spanned varied school contexts, grade levels, and subject areas, with differing levels of prior AI experience (see Table 2).

Each session lasted approximately 90 minutes and was structured around four phases using RAID (Section 3.1): (1) a 10-minute introduction to study goals and the three card types; (2) a 10-minute guided example to familiarize participants with the interface; (3) a 40-minute main activity in which participants documented anticipated or experienced AI breakdown scenarios and the values implicated in each; and (4) a 30-minute whole-group discussion in which participants shared their cases and collectively reflected.

3.1 RAID: A Tool for Structured AI Deliberation

Deliberations of AI in education often suffer from a lack of shared vocabulary, even when teachers refer to the same tool, they may hold very different understandings of its purpose, capabilities, and limitations. Teachers also bring varied mental models of how AI works in classroom settings. To support aligned discussion and shared understanding, we developed Responsible AI Deliberation (RAID), a web-based tool that uses teacher-friendly language and structured scaffolding to guide reflection without constraining interpretation (Figure 1).

RAID structured deliberation around three interconnected card types, guiding participants through a systematic process of identifying where and how classroom AI systems might fail and break down. (1) **Application Cards** built on prior work using input-model-output representations [27, 67] to help participants reason about how AI systems function. Each card described a category of classroom AI tool, explained underlying mechanisms such as inference and data processing, and identified the stakeholders involved (e.g., teachers, students, administrators).

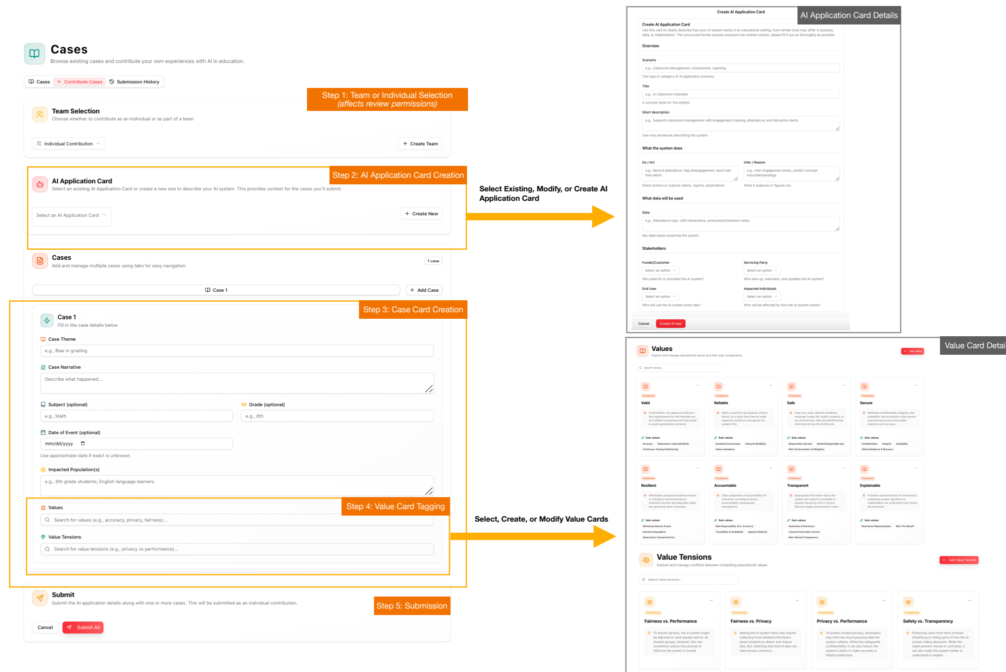

Fig. 1. Case contribution workflow in RAID (Responsible AI Deliberation). Contributors follow a structured five-step process: (1) **Team or Individual Selection** where contributors choose to submit as an individual or as part of a team, with team members able to review each other's contributions; (2) **AI Application Card Creation** where contributors select an existing, create, or modify an AI Application Card capturing the system's scenario, behaviors, data inputs, and stakeholder roles (top right); (3) **Case Card Creation** where contributors document the case theme, narrative, affected populations, and relevant metadata such as subject, grade level, and date (middle left); (4) **Value Card Tagging** where contributors link the case to relevant values and value tensions by selecting from, creating, or modifying value cards drawn from a curated taxonomy of ethical conflict pairings (e.g., Fairness vs. Privacy, Safety vs. Transparency; bottom right); and (5) **Submission** where contributors finalize and contribute the case to the repository.

(2) **Case Cards** were linked to Application Cards and anchored each AI application in concrete breakdown moments participants had experienced or anticipated. Participants described specific teaching situations, intended uses, and points at which the system might fail or generate tension—moving discussion from abstract descriptions toward situated accounts of how AI could disrupt classroom practices, relationships, and institutional constraints. (3) **Value Cards** were connected to each Case Card to support reflection on responsible AI values in context. Drawing on definitions adapted from Jakesch et al. [25], participants could revise existing definitions or introduce new ones, using the cards to surface and negotiate how RAI values became implicated in each case.

3.2 Qualitative Analysis Methods

We analyzed the focus group transcripts using an affinity diagramming approach [43] focused on identifying breakdowns, value tensions, and repair practices in teachers' discussions. All authors met regularly throughout the process to ensure consistency and reflexivity. We began with open coding to mark moments where participants described anticipated or experienced breakdowns in classroom AI use, along with how they interpreted and responded to these situations. We then clustered related excerpts based on shared forms of tension or repair,

examining how values such as fairness, autonomy, and accountability were invoked and negotiated in context rather than treated as fixed categories. Finally, we synthesized these clusters into higher-level themes capturing recurring patterns in how teachers anticipate, interpret, and seek to repair AI-related disruptions to classroom practice. In the following results section, participants are labeled by group and number (e.g., P1 of Group A is represented as A1). For clarity, we refer to eight groups in total, labeled A through H.

4 Results

We organize our findings around three recurring sites of breakdown that teachers anticipate as AI systems become embedded in classroom practice. These breakdowns concern (1) how AI systems redefine what counts as learning, engagement, and student performance through standardized metrics and automated judgments (Section 4.1), (2) how AI redistributes authority and responsibility between teachers, administrators, and algorithmic systems (Section 4.2), and (3) how AI tools reshape the relational and care-centered foundations of pedagogy by mediating how students are seen, supported, and evaluated (Section 4.3).

These themes recurred across specific AI system types as teachers deliberated over Application Cards representing different classroom tools. To make this mapping explicit, Table 1 organizes our findings by AI system type, drawing on established typologies [18, 70], and links each breakdown to the subsection where it is discussed in detail.

4.1 When AI Redefines What Counts as Learning

4.1.1 Repairing Metric-Induced Breakdowns in Fairness.

> "The system flagged a few students as 'at risk,' but I know they're doing fine. It just can't see the full picture." – D2

Teachers emphasized that what AI systems failed to capture went far beyond academic performance. They described how profiling, prediction, and early warning systems systematically erased dimensions of learning that mattered most in their classrooms, including students' language backgrounds, learning differences, emotional states, cultural norms, and nonstandard ways of participating (A3, B1–2, D2, E2, F3–4, H2–4). As one teacher (E3) reflected on how students learn, *"The system can mark some students as disengaged just because they're quiet, but they're actually listening and processing."*

Moreover, these misclassifications were not experienced as isolated technical errors but as the imposition of a narrow and culturally specific model of what learning is supposed to look like. As one teacher (H2) explained, *"A student with a disability, a kid of a different culture... who's defining what is appropriate eye contact or engagement? Because in certain cultures, eye contact is actually disrespectful."* Another teacher (H4) similarly shared, *"Depending on what kind of classroom you have... I teach in [a city in the USA]; it's often loud, and I like that... different groups are going to behave differently because of cultural values."* Yet these profiling and prediction systems reduced diverse forms of participation to a small set of quantifiable indicators, effectively redefining engagement in ways that privileged certain students while rendering others invisible. Teachers anticipated that without intervention, these metrics would not simply mislabel students but would reshape how learning itself was recognized, rewarded, and acted upon in their classrooms.

Across these cases, profiling and prediction systems produced breakdowns not simply because they were inaccurate, but because they imposed rigid definitions of learning, engagement, and risk onto diverse classroom realities that require contextual and relational judgment. In doing so, these systems risked displacing teachers' professional interpretations with simplified numerical proxies, reshaping how students were seen and how educational decisions were justified.

4.1.2 Repairing Breakdowns in Instructional Validity.

Table 1. **Anticipated breakdowns and pedagogical repair across classroom AI system types.** Each row corresponds to a subtheme in our analysis, organized by AI system type. RAI values indicate the Responsible AI principle(s) most at stake in each breakdown.

AI system type	Breakdown	Description	RAI values	Repair work
Intelligent tutoring & adaptive learning *Also touches: automated assessment*	*Metric-driven distortions of authentic learning* (§4.1.3)	Systems reward visible activity—clicks, time on task, on-screen completion—over genuine understanding, mistaking behavioral traces for intellectual engagement.	Autonomy Transparency	Teachers re-center instruction on reflection and dialogue; sideline dashboard indicators to preserve space for curiosity and struggle.
	Growth-oriented learning displaced by output focus (§4.3.3)	AI systems fail to distinguish effort, revision, and development from polished performance, reducing learning to outcomes rather than trajectories.	Autonomy Transparency	Teachers redesign assignments to surface process and revision; create space for explanation so growth remains recognizable when AI is involved.
Automated assessment & feedback *Also touches: intelligent tutoring*	*Instructional validity breakdowns* (§4.1.2)	AI feedback contradicts teachers' instructional guidance, creating student confusion about whose judgment counts and fragmenting coherent instructional meaning.	Accountability Transparency	Teachers reframe AI feedback for students, selectively override outputs, and reassert their own evaluative criteria.
	Performance vs. development conflation (§4.3.3)	Systems privilege surface features—correctness, fluency, completeness—over developmental trajectories, misrepresenting where students actually are as learners.	Accountability Fairness	Teachers create tasks that surface process and effort; use observations to supplement or override algorithmic representations of student progress.
Profiling, prediction & early warning *Also touches: classroom monitoring*	*Metric-induced fairness breakdowns* (§4.1.1)	Engagement and risk metrics impose culturally narrow definitions of learning, mislabeling students from diverse linguistic, cultural, and disability backgrounds as disengaged or at risk.	Fairness Accountability	Teachers contextualize metrics for administrators, flag cultural and linguistic biases in outputs, and resist treating algorithmic labels as objective truth.
	Predictive profiling preceding relationships (§4.3.1)	Risk scores and performance forecasts produce early judgments that precede teacher–student relationships, freezing students in deficit categories before teachers can know them through interaction.	Fairness Transparency	Teachers set aside system profiles, rely first on their own observations, and contest algorithmic labels with colleagues and administrators.
Classroom monitoring & orchestration *Also touches: profiling, prediction*	*Surveillance-induced breakdown in classroom authority* (§4.2.1)	Real-time monitoring tools undermine psychological safety and trust; systems framed as student support are repurposed as managerial oversight of teachers' competence.	Autonomy Privacy Accountability	Teachers enable or disable features selectively, reinterpret dashboards contextually, and push back when algorithmic representations conflict with their classroom knowledge.
	Mandated adoption displacing professional autonomy (§4.2.2)	Institutional mandates make AI use effectively compulsory, shifting accountability from whether students are learning to whether teachers demonstrate visible compliance with systems.	Autonomy Accountability	Teachers negotiate how much of a system to accept, translate institutional demands into workable classroom practices, and resist encroachment on pedagogical decision-making.
	Behavioral monitoring misframing emotional vulnerability (§4.3.2)	Automated behavioral alerts translate complex emotional realities into simplified risk signals, misframing student vulnerability as behavioral risk and reorienting teacher attention away from relational care.	Fairness Privacy Autonomy	Teachers treat alerts as provisional; rely on their own observations and relationships as the primary basis for responding to student needs.

Note. RAI values refer to Responsible AI principles: Fairness, Autonomy, Accountability, Transparency, Privacy. "Also touches" notes indicate findings that cut across system type categories; full discussion of each breakdown appears in the corresponding results subsection (Sections 4.1–4.3). System type categories draw on Zawacki-Richter et al. [70] and Holmes et al. [18]

> *"Some of the feedback was not what I would have written or said… it was the opposite of what I was telling them."* – C1

Drawing on their direct experience using automated assessment and feedback tools for grading and writing feedback, teachers described how these systems frequently conflicted with their instructional goals, producing breakdowns in what counted as valid evaluation. Participants (A2, C1, D2, E3, F2-3, G1-2, H2-3) described situations in which AI feedback contradicted their own guidance, creating uncertainty for students about whose judgment should be trusted and what kind of work was actually valued. As one teacher (E3) explained, *"Students would come to me confused, asking why the computer told them one thing when I had been telling them something else the whole time."*

Even when teachers provided detailed rubrics, participants observed that these automated assessment tools applied rigid and brittle rules that failed to capture the substance of students' thinking. Teachers reported cases from their own classroom use in which creative responses were penalized, synonyms were marked as wrong, or nuanced arguments were misunderstood. One teacher (D2) described how this forced them into ongoing repair work: *"I would read what the system gave them and think, that's not what I was trying to teach at all. Then I have to go back and explain to the student what to ignore and what actually matters."* Another participant (G2) added, *"It ends up being more work because now I'm not just giving feedback, I' m correcting the AI's feedback too."*

For teachers, instructional validity in AI-supported assessment involves more than technical accuracy. It depends on alignment with pedagogical intent, lesson goals, and the developmental trajectory of student learning. Based on these experiences, participants described repairing these breakdowns by reframing AI feedback for students, selectively overriding system outputs, and explicitly reasserting their own criteria for good work. Through these practices, teachers sought to preserve coherent instructional meaning in classrooms where automated assessment and feedback tools threatened to fragment or distort it.

4.1.3 Repairing Metric-Driven Distortions of Authentic Learning.

> *"It was just a checklist. It wasn't helping those kids learn."* – F1

Teachers drew on their direct experience with intelligent tutoring and adaptive learning systems to describe how these tools reduced complex learning processes into narrow, easily countable signals. Systems that tracked clicks, time on task, or on-screen activity were seen as mistaking visible activity for intellectual engagement, producing representations of learning that were shallow, misleading, and misaligned with pedagogical goals. Participants (A2, B1–3, C1, D2–3, F1, G4) emphasized that these metrics did not simply measure learning but began to redefine it. As one teacher (B2) explained, *"They learn how to game the system. They look busy on the screen, but they're not actually learning anything."* Another teacher (G4) described how students quickly adapted to what the system rewarded: *"They figure out what makes the dashboard happy, not what makes them understand."*

Teachers also worried that metricized engagement, across both intelligent tutoring and adaptive learning systems and automated assessment and feedback tools, would shift classroom priorities away from relationships and sensemaking toward data compliance. As one participant (B3) put it, *"If somebody's doing that, they're not doing anything with others."* In this way, dashboards were not neutral reporting tools but active participants in shaping what teachers and students treated as important.

Across these cases, intelligent tutoring and adaptive learning systems and automated assessment and feedback tools produced metric-driven breakdowns not because they were simply inaccurate, but because they imposed a narrow definition of learning that privileged what could be easily measured over what mattered educationally. When time, attention, and effort were redirected toward satisfying the system, opportunities for curiosity, struggle, and growth risked being displaced. Teachers anticipated repairing these distortions by deliberately re-centering learning around reflection, dialogue, and student development rather than dashboard indicators. By reframing or

sidelining metrics, they sought to preserve authentic learning as something that unfolds through interaction, effort, and meaning, not just data traces.

4.2 When AI Reassigns Authority and Responsibility in Classrooms

4.2.1 Repairing Surveillance-Induced Breakdowns in Classroom Authority.

> "My students would freak out, and the parents probably would too. That's a very Big Brother thing." – C4

Across all focus groups, teachers described AI enabled monitoring tools such as real time engagement dashboards, automated attendance systems, and behavioral analytics as sites where they anticipated breakdowns in how classroom authority, responsibility, and trust normally function (A2-3, B2, C1-3, D3, E1–E3, F1, G3, H2-4). Instead of classrooms being organized around teachers' situated judgment, participants anticipated a shift toward classrooms governed by algorithmic traces.

From students' perspective, teachers anticipated that constant monitoring would undermine trust and psychological safety by making learning feel like surveillance rather than participation. As one teacher explained, students would experience these tools not as support but as scrutiny. The opacity of data collection intensified this concern. As E3 put it, *"It tracks their clicks, their logins, even how long they pause, but no one really told us what happens with that data."* When the meaning and use of data were opaque, teachers anticipated a loss of their ability to explain, justify, or stand behind decisions made in the classroom.

Teachers also anticipated a breakdown in how responsibility is understood by students. Engagement risked becoming something that was measured and flagged by systems rather than something students took ownership of. As G3 put it, *"It makes me imagine there has to be a camera in the room recording everything... and I would not be comfortable with my students being watched like that."* Even the possibility of being monitored was expected to shape behavior [32], creating a classroom atmosphere that required active management to preserve trust.

These breakdowns extended beyond students to teachers themselves. Several participants expected that monitoring systems would be repurposed as tools of managerial oversight. As A3 described, *"My principal may notice it was flagging more students in my classes than anyone else. He was concerned about my classroom management."* What began as a student facing support system became a mechanism through which teachers' competence was evaluated, shifting authority away from professional judgment toward algorithmic indicators.

In response to these anticipated breakdowns, teachers described enabling or disabling features, to reinterpret and contextualize dashboards for students and administrators, and to push back when algorithmic representations conflicted with what they knew about their classrooms. These practices highlight that surveillance oriented AI does not simply add new data to classrooms. It destabilizes how authority works, and teachers' repair labor becomes essential to keeping classrooms governable, trustworthy, and pedagogically meaningful.

4.2.2 Repairing Autonomy Under Mandated AI Adoption.

> "It feels less like a choice and more like a checkbox. We're supposed to teach, but now it's about showing we used the tool." – D1

Teachers described district and school level mandates to adopt AI systems as producing a different but equally consequential kind of breakdown (A1-3, B2-3, C2, E1-2, G4, H1). Rather than disrupting trust through surveillance, mandated adoption was expected to disrupt professional autonomy by embedding AI systems into the institutional routines that define what counts as legitimate teaching. Although many tools were introduced as optional supports, participants anticipated that funding decisions, reporting requirements, and administrative oversight would make their use effectively compulsory.

This breakdown was experienced as a shift in what teachers were held responsible for. Several participants described feeling pressure to demonstrate that they were using district purchased AI systems, even when those

tools did not align with their instructional goals (B2, C2, H1). As one teacher (C2) explained, *"The school district and administrators are the ones funding it, so if this gets implemented, we might feel pressured by administration saying, 'We're paying for this, you need to use it.'"* In these situations, accountability no longer rested on whether students were learning, but on whether teachers could show visible compliance with the system.

Moreover, teachers in under resourced schools and special education settings expected mandated AI systems to intensify workload and administrative burden. As one special education teacher (E2) noted, *"AI based engagement tracking and IEPs[1] reporting could really increase my workload, especially when I'm already managing a large class and a ton of documentation."* Others described how alerts, flags, and required follow ups created a constant stream of additional work. As C2 explained, *"Every time the dashboard sends an alert, I have to double check what actually happened. It's more time on paperwork than time with my students."*

Teachers also anticipated that mandated systems would quietly reassign labor that had previously belonged to counselors, administrators, or support staff. Tools for attendance analytics, behavioral monitoring, and emotional indicators effectively shifted responsibility for monitoring and intervention onto classroom teachers [10]. As B3 summarized, *"[In this case,] teachers are told they're not co parents, but they're expected to do all the work."*

In response, teachers anticipated engaging in ongoing repair by negotiating how much of a system to accept, how to translate its demands into workable classroom practices, and when to resist its encroachment on pedagogical decision making. These dynamics show that the ethical stakes of classroom AI lie not only in what systems do, but in how institutional power structures compel teachers to live with them and how teachers must continually repair those arrangements to keep teaching possible.

4.3 When AI Disrupts Relational and Care-Centered Pedagogy

4.3.1 Repairing Relational Authority Under Predictive Student Profiling.

> *"When I'm meeting students for the first time, I don't want to be influenced by what a system, or even another teacher, has already decided about them."* – E1

Teachers described profiling, prediction, and early warning systems, including features such as engagement scores, risk flags, and performance forecasts, as introducing preemptive narratives about who students are and what their learning is likely to become. Rather than serving as neutral supports, these systems were seen as producing early judgments that could shape expectations before teachers had the chance to build relationships with students.

Participants (A1-2, C2-3, E1, F2-3, G1-3, H2-4) worried that predictive profiles would become taken-for-granted truths that followed students across classrooms. As one teacher (F3) explained, *"Once something is in the system, people start treating it like that's who the kid is, even if it's wrong."* Another participant (C3) described how risk labels could quietly structure attention and care: *"You end up watching the flagged kids differently, even when you don't mean to."*

For teachers, this was not simply a technical issue of prediction accuracy but a relational one. Relational pedagogy depends on allowing students to be known through interaction, growth, and change. When algorithmic labels precede relationships, they risk freezing students into deficit-oriented categories before students have had the chance to define themselves. As one teacher (F2) put it, *"You meet the kid already labeled, before you even meet the kid."* Several participants emphasized that this danger was especially acute for students from historically marginalized communities, where biased data and institutional inequities could be amplified by predictive systems.

Teachers anticipated engaging in repair to protect relational trust. They described deliberately setting aside system-generated profiles, relying first on their own observations, and reframing or contesting algorithmic labels

[1]Individualized Education Programs (IEPs) are legally mandated, customized learning plans designed to support students with disabilities and special educational needs in U.S. public schools.

in conversations with colleagues and administrators. As one teacher (A1) explained, *"I want to see who they are before I see what the AI says."* Through this work, teachers sought to preserve the classroom as a space where students could be known as developing people rather than as risk scores or forecasts. In this sense, repair was not about correcting data but about reclaiming the moral authority to decide who students are allowed to become.

4.3.2 Repairing Emotional Care Under Behavioral Monitoring.

> *"Forget to notice your kids and the emotions that they come in with… because you're just looking at this AI management tool."* – A3

Teachers described classroom monitoring and orchestration tools as posing risks to how attention and care circulate in the classroom (A1–3, D2, E1–2, G3–4). Participants (A3, E1, E2) shared that much of their daily work involves noticing subtle emotional cues, responding to disclosures, and supporting students through moments of anxiety, trauma, or instability. These forms of care are deeply relational and situational, emerging through presence, conversation, and trust rather than through standardized indicators.

By contrast, these classroom monitoring and orchestration tools translate these emotional realities into simplified, legible traces such as attendance patterns, risk scores, or engagement flags. Teachers worried that this translation would not only miss important context but also redefine what counts as a problem worth addressing. As one teacher (E1) reflected, *"A kid can be falling apart inside and still look fine on the screen. The system doesn't see that, but I do."* Another teacher (E2) described how an alert might be triggered by a student's irregular attendance, even when the underlying issue could be family stress that required care rather than discipline. In these cases, algorithmic signals risked misframing vulnerability as behavioral risk.

Teachers worried that emotional labor would be reorganized around what systems could detect rather than what students actually needed. As one teacher (G3) put it, *"If the tool tells you who to worry about, you stop looking for yourself."* In response, teachers described engaging in repair to protect care-centered pedagogy. They would treat AI alerts as provisional and secondary, using their own observations, conversations, and relationships as the primary basis for deciding how to respond to students. This repair work was not simply about correcting data but about resisting a deeper shift in moral authority from teachers' relational judgment to automated classifications. Through this ongoing effort, teachers sought to keep emotional care grounded in human connection rather than allowing it to be governed by classroom monitoring tools.

4.3.3 Repairing Growth-Oriented Learning Under AI Use.

> *"We've assigned virtue, value, and morality to doing your own work... but students may see their goal as just getting a good grade and choose ChatGPT accordingly."* – B1

Teachers identified a deeper breakdown in what automated assessment and feedback systems recognize, reward, and make visible as learning, particularly in tools designed to scaffold science learning, generate feedback, or evaluate student work (B1-2, D1-2, F2, D2-3, G3-4, H1-2). Growth-oriented pedagogy depends on effort, struggle, revision, and the gradual formation of understanding, yet these dimensions are difficult for AI systems to detect and are often displaced by outputs, completion, and surface features. When such systems enter classroom practice, they shift what counts as evidence of learning, creating tensions between pedagogical values and algorithmic representations.

As one teacher (D2) described, this reoriented how students approached their work: *"They're not asking, 'What do I understand?' They're asking, 'What will get me the points?' "* Another teacher (B1) contrasted two students, one who showed steady growth through revision and effort and another who maintained high grades by producing polished answers, noting that *"To the AI [evaluation] system, they look the same, but to me they're in completely different places as learners."* For teachers, both intelligent tutoring systems and automated assessment tools failed to distinguish between performance and development, even though that distinction is central to how teachers evaluate progress.

Because AI systems privilege surface features such as correctness, fluency, and completeness, they risk reducing learning to outcomes rather than trajectories. Teachers anticipated repairing this breakdown by redesigning assignments to make learning processes more visible and by creating space for explanation, revision, and reflection so that growth could still be recognized when AI was involved. In this sense, repair meant reasserting pedagogical values that intelligent tutoring, adaptive learning systems and automated assessment tools could not easily represent, keeping learning centered on development rather than display.

5 Discussion

As AI systems enter classroom life, teachers emerge as everyday repairers—not only of technologies but of the values, relationships, and learning conditions that those technologies disrupt or expose. Our findings make visible the negotiation work teachers perform when AI enters their classrooms, showing how RAI principles such as fairness and accountability, and the relational and care-centered dimensions of pedagogy that those frameworks do not yet capture, become sites of ongoing repair. Together, these findings motivate a future research agenda around *pedagogical repair* as a Responsible AI orientation that centers teachers' situated work in sustaining trust, equity, and meaningful learning.

5.1 Toward Contextualized and Teacher-Governed Metrics

Teachers' accounts of profiling, prediction, and early warning systems, as well as automated assessment and feedback tools, point to a deeper challenge in AI-supported education: many of the ethical tensions surrounding these systems arise not from model error, but from how learning, engagement, and risk are defined and measured in the first place. Participants described how standardized metrics frequently fail to align with classroom realities, misclassifying students whose participation styles, linguistic backgrounds, or neurodiverse needs do not fit algorithmic assumptions.

This challenge is not unique to education, prior work has shown that the choice of optimization target is itself a normative decision with distributional consequences, and that optimizing predictive accuracy can come at the cost of meaningful outcomes for the people most affected [59]. In child welfare, Stapleton et al. [54] similarly showed that predictive systems used to make high-stakes decisions about families often fail to capture the relational and contextual dimensions of care that practitioners consider most important, a finding that resonates with teachers' experiences in our study. In our study, teachers did not reject measurement altogether. Instead, they anticipated needing to reinterpret scores, qualify labels, and sometimes resist the use of AI outputs in high-stakes decisions [51]. These repair practices demonstrate how fairness and validity emerge through situated interpretation rather than being guaranteed by technical design alone. Ethical risk is situated not only in model performance but in how measurement frameworks are selected, interpreted, and operationalized by schools, districts, and vendors, and whose knowledge counts when those frameworks fail to fit.

For designers and policymakers, this may shift attention from improving metric accuracy to two more fundamental questions: who has the authority to define what counts as learning and risk in a given classroom, and what mechanisms exist for teachers to contest, override, or recontextualize metrics that do not fit their students? Supporting contestability, the ability to meaningfully challenge and revise algorithmic outputs [35], is not simply a technical feature to be added to existing systems but a governance commitment that requires giving teachers legitimate standing to dispute and reframe what AI systems report.

5.2 Accountability Should Be Paired with Teacher Authority

Our findings reveal a consistent pattern of accountability displacement in which teachers are expected to act on AI outputs and answer for their consequences, yet they have no meaningful standing in the decisions that shape how those systems are designed, deployed, or evaluated. This is not a measurement problem but a governance problem.

Our results show that accountability is detached from professional judgment and reattached to algorithmic compliance. Similarly, Kuo et al. [31] documented an identical displacement among frontline workers in homeless services, where AI tools introduced by institutions consistently overrode workers' relational knowledge and transferred accountability to system outputs.

This displacement is compounded by structural conditions. In under-resourced schools, classroom monitoring and orchestration tools and profiling systems are sometimes adopted to compensate for funding shortfalls, substituting for counseling staff, behavioral support, or intervention capacity that districts can no longer afford. When this happens, teachers inherit not only the accountability burden of these systems but also the institutional failures those systems were meant to conceal. Alkhatib and Bernstein [1] frame this as a defining feature of "street-level algorithms": the governance gap between where authority formally resides and where responsibility actually lands is filled by frontline workers, invisibly and without recourse.

While teachers intend to negotiate what tools to use, contextualize algorithmic reports for administrators, and explicitly refuse certain forms of compliance, they enact a form of professional accountability that the systems themselves cannot provide. These practices suggest that responsible AI governance in education requires more than transparency mechanisms or audit trails. It requires institutional structures that give teachers formal standing in AI adoption decisions, including representation in procurement processes, the right to flag misalignment between system outputs and classroom realities, and protection against administrative evaluation based on algorithmic indicators they did not endorse [11, 28, 37]. Without such structures, pedagogical repair remains a form of invisible labor that patches governance failures without ever being recognized as doing so.

5.3 Relational Pedagogy as a Missing Dimension of Responsible AI

Teachers' reflections also point to a dimension of ethical impact that is often overlooked in discussions of classroom AI: how these systems shape relationships and learning orientations. Beyond repairing AI outputs directly, teachers described needing to repair the pedagogical conditions that AI systems disrupt or expose, including the relationships, trust, and professional authority that these tools threatened to displace when introduced without adequate preparation or support. Participants worried that, whether orchestration or assessment tools, these systems could pre-shape expectations of students, narrow how learning is recognized, and encourage transactional approaches to schooling that prioritize outputs over development.

These concerns resonate with prior work showing that teaching is fundamentally relational, grounded in trust, care, and recognition of students as developing persons rather than data points [3, 41]. Critical AI scholarship has similarly shown how data-driven systems can reshape how people are perceived and treated, often privileging prediction and optimization over situated human judgment [11]. In child welfare, Saxena and Guha [48] found that caseworkers actively resisted algorithmic risk scores that threatened to freeze clients in deficit categories, a dynamic our teachers anticipated with respect to predictive profiling and early warning systems in their own classrooms. These practices reflect a relational approach to accountability: one that centers not only on whether algorithmic outputs are accurate or fair in a technical sense, but also on whether they preserve the conditions under which meaningful human relationships and professional judgment can be sustained [37].

Our results illustrate how relational and growth-oriented pedagogy is maintained through ongoing repair in AI-mediated classrooms. Teachers described how they emphasized reflection, revision, and dialogue, even when AI systems made it easy for students to optimize for grades or speed. In this way, their ethical concerns extended beyond whether AI systems were accurate to whether they supported the moral and developmental aims of education. Moreover, this foregrounds a dimension of educational AI that is often invisible to technical evaluation, but central to how teachers, students, and families experience these systems. For system developers, school leaders, and Responsible AI practitioners, it highlights that ethical impact is shaped not only by what AI predicts or generates, but by how it restructures relationships, expectations, and opportunities for growth.

Addressing this requires expanding what counts as evidence of ethical AI in education to include relational and developmental outcomes alongside technical accuracy, a shift that only becomes possible when teachers' repair work, and the values it enacts, are treated as legitimate data about how well AI systems are serving their educational purpose.

5.4 Future Work: Toward Pedagogical Repair as a Responsible AI Orientation

Our results show that pedagogical repair is not merely a descriptive label for what teachers do. It reframes how Responsible AI should be studied and governed in educational contexts. Rather than treating RAI as a checklist of properties to be satisfied at design, it foregrounds the ongoing, situated work through which values are negotiated and sustained in practice [21, 22]. Our findings show that the most consequential ethical tensions in educational AI do not emerge at the moment of model development but in the everyday work of teachers navigating AI within classrooms, schools, and institutional constraints. Teachers' repair practices reveal how harm, trust, and fairness are produced through interaction with infrastructures, policies, and relationships, not only through algorithmic outputs.

This reframing also shifts whose labor and expertise matter. Teachers are often positioned as downstream users of educational technology, yet our study shows that they are central actors in making AI systems workable and ethically tolerable [47, 55]. Crucially, this labor is not only a response to AI but also to the structural conditions that make AI adoption feel necessary. In many under-resourced schools, AI is introduced to compensate for funding deficits, replacing counselors, absorbing unsustainable grading loads, or flagging students that specialist staff can no longer support. Responsible AI in education cannot therefore be evaluated at the level of individual systems alone. Teachers' repair work is a form of situated evidence about what responsible education requires, evidence that no technical audit can generate on its own, and future research should treat it as a generative lens rather than a gap to be closed by better design [46].

Taken together, our findings point to several directions for future work, including studying how teachers' repair practices evolve as AI systems mature, how institutional policies and resource constraints shape what kinds of repair are possible, and how students experience and respond to these dynamics. For the broader Responsible AI community, pedagogical repair invites a reorientation toward practice-based accountability, where responsibility is understood not only as something built into systems but as something continuously negotiated by the people who live with them [46].

6 Limitations

While our study offers initial insights into how teachers navigate Responsible AI in classrooms, several limitations point to important directions for future work. First, our findings draw on a relatively small and regionally specific sample of teachers from the Midwest United States, which constrains the range of perspectives captured. To address this, we plan to deploy RAID, the web-based deliberation tool used in this study, across a broader teacher network to collect repair cases from more diverse educational, cultural, and institutional contexts, and to support mixed-methods analysis at scale. Second, our findings are based on self-reported anticipations of breakdown rather than direct observation of classroom practice. As AI tools become more fully integrated into classroom practice, future research should examine whether the anticipatory repair work documented here translates into observable practice, and how teachers' responses evolve as they gain direct experience with deployed systems.

7 Conclusion

This study examines how K–12 teachers anticipate and navigate breakdowns as AI systems become embedded in classroom practice. Drawing on Jackson's concept of repair [23], we analyze eight focus groups with 26 K-12 teachers to surface the anticipatory and interpretive work teachers perform to keep AI compatible with

their pedagogical commitments. We show how teachers engage in ongoing repair to preserve professional authority, sustain meaningful learning, and protect care-centered classroom relationships as algorithmic systems increasingly shape how teaching and learning are represented and evaluated. Moreover, our findings contribute to Responsible AI research by showing that values such as fairness, transparency, accountability, and trust are not simply built into systems but are actively reestablished through teachers' everyday practices of contextualizing, resisting, and reworking AI tools. This suggests that Responsible AI in education cannot be achieved through technical safeguards alone, but should also support teacher judgment, contextual interpretation of metrics, and the relational dimensions of teaching. Looking forward, we argue that pedagogical repair should be understood as an ongoing orientation for Responsible AI in education, guiding the design and governance of AI systems in ways that better sustain trust, equity, and meaningful learning over time.

Acknowledgments

Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation.

References

[1] Ali Alkhatib and Michael Bernstein. 2019. Street-Level Algorithms: A Theory at the Gaps Between Policy and Decisions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300760

[2] Pengcheng An, Kenneth Holstein, Bernice d'Anjou, Berry Eggen, and Saskia Bakker. 2020. The TA framework: Designing real-time teaching augmentation for K-12 classrooms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–17.

[3] Gert Biesta. 2015. *Good Education in an Age of Measurement: Ethics, Politics, Democracy*. Routledge, London.

[4] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 5454–5476. doi:10.18653/v1/2020.acl-main.485

[5] Ashley Boone, Aparna Arul, James Kemerait, Ian Wood, Jaeri Suh, Carl DiSalvo, and Christopher A Le Dantec. 2025. Designing Tools for Data Advocacy. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 2130–2142.

[6] Ashley Boone, Carl Disalvo, and Christopher A Le Dantec. 2023. Data Practice for a Politics of Care: Food Assistance as a Site of Careful Data Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

[7] Emma Brunskill, Kole A. Norberg, Stephen Fancsali, and Steve Ritter. 2024. Examining the Use of an AI-Powered Teacher Orchestration Tool at Scale. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (Atlanta, GA, USA) *(L@S '24)*. Association for Computing Machinery, New York, NY, USA, 356–360. doi:10.1145/3657604.3664704

[8] Miguel A. Cardona, Roberto J. Rodríguez, and Kristina Ishmael. 2023. *Artificial Intelligence (AI) and the Future of Teaching and Learning: Insights and Recommendations*. Technical Report. U.S. Department of Education, Office of Educational Technology (OET). https://www2.ed.gov/documents/ai-report/ai-report.pdf

[9] Shiva Darian, Aarjav Chauhan, Ricky Marton, Janet Ruppert, Kathleen Anderson, Ryan Clune, Madeline Cupchak, Max Gannett, Joel Holton, Elizabeth Kamas, et al. 2023. Enacting data feminism in advocacy data work. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–28.

[10] Melissa Diliberti, Heather L. Schwartz, David Grant, and Gerald P. Hunter. 2021. *Stress Topped the Reasons Why Public School Teachers Quit, Even Before COVID-19*. Technical Report. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1121-2.html

[11] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, New York, NY, USA.

[12] Shuo Feng, Lishan Zhang, Shuwen Wang, and Zhihui Cai. 2023. Effectiveness of the functions of classroom orchestration systems: A systematic review and meta-analysis. *Comput. Educ.* 203, C (Oct. 2023), 17 pages. doi:10.1016/j.compedu.2023.104864

[13] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, and Burkhard Schafer. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* 28, 4 (2018), 689–707.

[14] Sarah E Fox, Samantha Shorey, Esther Y Kang, Dominique Montiel Valle, and Estefania Rodriguez. 2023. Patchwork: the hidden, human labor of AI integration within essential work. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–20.

[15] Ishaya Gambo, Faith-Jane Abegunde, Omobola Gambo, Roseline Oluwaseun Ogundokun, Akinbowale Natheniel Babatunde, and Cheng-Chi Lee. 2024. GRAD-AI: An automated grading tool for code assessment and feedback in programming course. *Education and Information Technologies* (2024), 1–41.

[16] Mary L. Gray and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* Houghton Mifflin Harcourt.

[17] Emma Harvey, Allison Koenecke, and Rene F. Kizilcec. 2025. "Don't Forget the Teachers": Towards an Educator-Centered Understanding of Harms from Large Language Models in Education. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1064, 19 pages. doi:10.1145/3706598.3713210

[18] Wayne Holmes, Maya Bialik, and Charles Fadel. 2019. *Artificial Intelligence in Education. Promise and Implications for Teaching and Learning.*

[19] Wayne Holmes, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C. Santos, Mercedes T. Rodrigo, Mutlu Cukurova, Ig Ibert Bittencourt, and Kenneth R. Koedinger. 2022. Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education* 32, 3 (Sept. 2022), 504–526. doi:10.1007/s40593-021-00239-1

[20] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. 2019. Co-designing a teacher-AI workflow for classroom orchestration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–15. doi:10.1145/3290605.3300831

[21] Lara Houston, Steven J Jackson, Daniela K Rosner, Syed Ishtiaque Ahmed, Meg Young, and Laewoo Kang. 2016. Values in repair. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1403–1414.

[22] Steven J. Jackson. 2014. Rethinking Repair. In *Media Technologies: Essays on Communication, Materiality, and Society*, Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (Eds.). MIT Press, Cambridge, MA, 221–240.

[23] Steven J Jackson and Laewoo Kang. 2014. Breakdown, obsolescence and reuse: HCI and the art of repair. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 449–458.

[24] Steven J Jackson, Alex Pompe, and Gabriel Krieshok. 2012. Repair worlds: maintenance, repair, and ICT for development in rural Namibia. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. 107–116.

[25] Maurice Jakesch, Zana Buçinca, Saleema Amershi, and Alexandra Olteanu. 2022. How Different Groups Prioritize Ethical Values for Responsible AI. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. ACM, 310–323. doi:10.1145/3531146.3533097

[26] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399. doi:10.1038/s42256-019-0088-2

[27] Ji-Youn Jung, Devansh Saxena, Minjung Park, Jini Kim, Jodi Forlizzi, Ken Holstein, and John Zimmerman. 2025. Making the Right Thing: Bridging HCI and Responsible AI in Early-Stage AI Concept Selection. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 2992–3012. doi:10.1145/3715336.3735745

[28] Naveena Karusala, Sohini Upadhyay, Rajesh Veeraraghavan, and Krzysztof Z. Gajos. 2024. Understanding Contestability on the Margins: Implications for the Design of Algorithmic Decision-making in Public Services. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 478, 16 pages. doi:10.1145/3613904.3641898

[29] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfefer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences* 103 (2023), 102274. doi:10.1016/j.lindif.2023.102274

[30] Seyun Kim, Jonathan Ho, Yinan Li, Bonnie Fan, Willa Yunqi Yang, Jessie Ramey, Sarah E Fox, Haiyi Zhu, John Zimmerman, and Motahhare Eslami. 2024. Integrating Equity in Public Sector Data-Driven Decision Making: Exploring the Desired Futures of Underserved Stakeholders. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–39.

[31] Tzu-Sheng Kuo, Hong Shen, Jisoo Geum, Nev Jones, Jason I. Hong, Haiyi Zhu, and Kenneth Holstein. 2023. Understanding Frontline Workers' and Unhoused Individuals' Perspectives on AI Used in Homeless Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 860, 17 pages. doi:10.1145/3544548.3580882

[32] Monika Blue Kwapisz, Avanya Kohli, and Prashanth Rajivan. 2024. Privacy Concerns of Student Data Shared with Instructors in an Online Learning Management System. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 661, 16 pages. doi:10.1145/3613904.3642914

[33] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. Interviews and Focus Groups. In *Research Methods in Human-Computer Interaction*, Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser (Eds.). Morgan Kaufmann, Cambridge, MA, 187–228. doi:10.1016/B978-0-12-805390-4.00008-X

[34] Jinsook Lee, Yann Hicke, Renzhe Yu, Christopher Brooks, and René F. Kizilcec. 2024. The Life Cycle of Large Language Models in Education: A Framework for Understanding Sources of Bias. *British Journal of Educational Technology* (2024). doi:10.1111/bjet.13505

[35] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 106 (April 2021), 25 pages. doi:10.1145/3449180

[36] Jasmine Mariappan and Chitra Krishnan. 2023. Artificial intelligence: future of advance learning. *Digital Transformation in Education: Emerging Markets and Opportunities* 118 (2023).

[37] Jacob Metcalf, Ranjit Singh, Emanuel Moss, Emnet Tafesse, and Elizabeth Anne Watkins. 2023. Taking Algorithms to Courts: A Relational Approach to Algorithmic Accountability. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1450–1462. doi:10.1145/3593013.3594092

[38] Sábëlo Mhlambi and Simona Tiribelli. 2023. Decolonizing AI ethics: Relational autonomy as a means to counter AI harms. *Topoi* 42, 3 (2023), 867–880.

[39] Joy Ming, Hawi H Tolera, Jiamin Tu, Ella Yitzhaki, Chit Sum Eunice Ngai, Madeline Sterling, Ariel C Avgar, Aditya Vashistha, and Nicola Dell. 2025. Exploring Data-Driven Advocacy in Home Health Care Work. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.

[40] Fawad Naseer, Muhammad Usama Khalid, Nafees Ayub, Akhtar Rasool, Tehseen Abbas, and Muhammad Waleed Afzal. 2024. Automated assessment and feedback in higher education using generative AI. In *Transforming education with generative AI: prompt engineering and synthetic content creation*. IGI Global Scientific Publishing, 433–461.

[41] Nel Noddings. 2013. *Caring: A Relational Approach to Ethics and Moral Education* (2 ed.). University of California Press, Berkeley, CA.

[42] Ronald Perez-Alvarez, Ioana Jivet, Mar Pérez-Sanagustín, Maren Scheffel, and Katrien Verbert. 2022. Tools designed to support self-regulated learning in online learning environments: A systematic review. *IEEE Transactions on Learning Technologies* (2022).

[43] Kara Pernice. 2018. *Affinity Diagramming: Collaboratively Sort UX Findings & Design Ideas*. https://www.nngroup.com/articles/affinity-diagram/ Accessed: 2025-09-09.

[44] Donatella Persico and Francesca Pozzi. 2015. Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology* 46, 2 (2015), 230–248.

[45] Petar Radanliev, Omar Santos, Alistair Brandon-Jones, and Adam Joinson. 2024. Ethics and responsible AI deployment. *Frontiers in Artificial Intelligence* 7 (2024), 1377011.

[46] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 7 (April 2021), 23 pages. doi:10.1145/3449081

[47] Andrew L. Russell and Lee Vinsel. 2018. After Innovation, Turn to Maintenance. *Technology and Culture* 59, 1 (2018), 1–25. doi:10.1353/tech.2018.0004

[48] Devansh Saxena and Shion Guha. 2024. Algorithmic harms in child welfare: Uncertainties in practice, organization, and street-level decision-making. *ACM Journal on Responsible Computing* 1, 1 (2024), 1–32.

[49] Devansh Saxena, Seh Young Moon, Dahlia Shehata, and Shion Guha. 2022. Unpacking invisible work practices, constraints, and latent power relationships in child welfare through casenote analysis. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–22.

[50] Andrew Sears and Julie A. Jacko (Eds.). 2007. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (2nd ed.). CRC Press, Boca Raton, FL. doi:10.1201/9781410615862

[51] Neil Selwyn. 2019. *Should Robots Replace Teachers? AI and the Future of Education*. Polity Press, Cambridge, UK.

[52] Priten Shah. 2023. *AI and the Future of Education: Teaching in the Age of Artificial Intelligence*. John Wiley & Sons.

[53] Jesper Simonsen and Toni Robertson (Eds.). 2012. *Routledge International Handbook of Participatory Design* (1 ed.). Routledge, London. doi:10.4324/9780203108543

[54] Logan Stapleton, Min Hun Lee, Diana Qing, Marya Wright, Alexandra Chouldechova, Ken Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1162–1177. doi:10.1145/3531146.3533177

[55] Susan Leigh Star. 1999. *The Ethnography of Infrastructure*. American Behavioral Scientist.

[56] The Institute for Ethical AI in Education. 2021. *The Ethical Framework for AI in Education*. Technical Report. The Institute for Ethical AI in Education. https://www.buckingham.ac.uk/wp-content/uploads/2021/03/The-Institute-for-Ethical-AI-in-Education-The-Ethical-Framework-for-AI-in-Education.pdf

[57] Rama Adithya Varanasi and Nitesh Goyal. 2023. "It is currently hodgepodge": Examining AI/ML practitioners' challenges during co-production of responsible AI values. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–17.

[58] Lee Vinsel and Andrew L. Russell. 2020. *The Innovation Delusion: How Our Obsession with the New Has Disrupted the Work That Matters Most*. Currency, New York, NY.

[59] Angelina Wang, Sayash Kapoor, Solon Barocas, and Arvind Narayanan. 2024. Against Predictive Optimization: On the Legitimacy of Decision-making Algorithms That Optimize Predictive Accuracy. *ACM J. Responsib. Comput.* 1, 1, Article 9 (March 2024), 45 pages. doi:10.1145/3636509

[60] Elizabeth A. Watkins, David Ribes, and Madeleine Clare Elish. 2020. Repairing Innovation: A Study of Integrating AI in Clinical Care.

[61] Ben Williamson, Felicitas Macgilchrist, and John Potter. 2023. Re-examining AI, automation and datafication in education. *Learning Media and Technology* 48 (01 2023), 1–5. doi:10.1080/17439884.2023.2167830

[62] Richmond Y. Wong, Michael A. Madaio, and Nick Merrill. 2023. Seeing Like a Toolkit: How Toolkits Envision the Work of AI Ethics. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 145 (April 2023), 27 pages. doi:10.1145/3579621

[63] Khensani Xivuri and Hosanna Twinomurinzi. 2023. How AI developers can assure algorithmic fairness. *Discover Artificial Intelligence* 3, 1 (2023), 27.

[64] W. Xu and F. Ouyang. 2022. A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies* 27 (2022), 4195–4223. doi:10.1007/s10639-021-10774-y

[65] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. 2024. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. *British Journal of Educational Technology* 55, 1 (2024), 90–112. doi:10.1111/bjet.13370

[66] Kexin Bella Yang, Vanessa Echeverria, Zijing Lu, Hongyu Mao, Kenneth Holstein, Nikol Rummel, and Vincent Aleven. 2023. Pair-Up: Prototyping Human-AI Co-orchestration of Dynamic Transitions between Individual and Collaborative Learning in the Classroom. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 453, 17 pages. doi:10.1145/3544548.3581398

[67] Nur Yildirim, Susanna Zlotnikov, Deniz Sayar, Jeremy M. Kahn, Leigh A. Bukowski, Sher Shah Amin, Kathryn A. Riman, Billie S. Davis, John S. Minturn, Andrew J. King, Dan Ricketts, Lu Tang, Venkatesh Sivaraman, Adam Perer, Sarah M. Preum, James McCann, and John Zimmerman. 2024. Sketching AI Concepts with Capabilities and Examples: AI Innovation in the Intensive Care Unit. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3613904.3641896

[68] Yaxuan Yin, Shamya Karumbaiah, and Shona Acquaye. 2025. Responsible AI in Education: Understanding Teachers' Priorities and Contextual Challenges. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 2705–2727. doi:10.1145/3715275.3732176

[69] Yaxuan Yin, Shamya Karumbaiah, Ju Lim, Shona Acquaye, and Sara Schwartz. 2025. Toward Teacher-Centered AI Design: Exploring the Role of Pedagogical Values and Contextual Factors in K-12 Teachers' Perceptions of Responsible AI. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning (CSCL 2025)*, Jun Oshima, Bodong Chen, Frieda Vogel, and Sanna Järvelä (Eds.). International Society of the Learning Sciences, 196–204. doi:10.22318/cscl2025.817482 Long Paper.

[70] Olaf Zawacki-Richter, Victoria I. Marín, Melissa Bond, and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education: where are the educators? *International Journal of Educational Technology in Higher Education* 16, 1 (2019), 39. doi:10.1186/s41239-019-0171-0

Generative AI Usage Statement

The authors used GPT-5.2 (ChatGPT) to assist with grammar, wording, and clarity of expression. No generative AI system was used to generate research content, analysis, results, or interpretations. All claims, arguments, and conclusions in this paper were developed, verified, and are solely the responsibility of the authors.

Ethical Considerations

Our study was approved by our university's Institutional Review Board (IRB). Focus group participants were informed that while we would maintain confidentiality in reporting, anonymity could not be guaranteed within the group setting. To protect participants, we anonymized all quotes in publications, reported characteristics only at a low level of granularity, stored data securely in accordance with IRB protocols, and deleted original recordings after transcription. We also recognized that participation required additional time from teachers with already demanding schedules. In appreciation of their time and insights, each educator received a stipend of $30. Beyond compensation, we sought to maintain reciprocity by contributing to community-facing activities, including AI literacy workshops, seminars, and other teaching engagements with local educational communities. These activities sought to ensure that the benefits of the research extended beyond data collection, fostering mutual exchange between researchers and practitioners.

A Supplementary Study Materials

A.1 Participant Demographics

Table 2. Participant demographics, professional background, and AI tool usage ($N = 26$).

Category	Subcategory	Count
Gender	Female	16
	Male	10
Demographic Background	White	15
	Black	2
	Hispanic	4
	Asian	3
	Middle Eastern	2
School Level	Elementary School	4
	Middle School	10
	High School	12
School Sector	Public School	19
	Private School	7
Subject Area	English / Language Arts	8
	Mathematics	3
	Science	5
	Social Studies / History	4
	Technology / Computer Science	2
	Special Education	4
Teaching Experience	>20 years	13
	10–20 years	5
	5–10 years	4
	1–5 years	4
AI Tool Usage	Lesson planning	17
	Creating teaching materials	12
	Grading or providing feedback	11
	Supporting student learning	9
	Classroom management	6
	Professional development	9
	Administrative tasks	7