Responsible AI in Education: Understanding Teachers' Priorities and Contextual Challenges

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

Shona Acquaye
University of Wisconsin–Madison
Madison, Wisconsin, USA
sacquaye@wisc.edu

Abstract

Recent advancements in Artificial Intelligence (AI) for education provide teachers with innovative tools to enhance student learning. However, due to the complexity and dynamic nature of education, the alignment between generalized Responsible AI (RAI) guidelines and the priorities of K-12 teachers — the primary stakeholders — remains unclear, potentially undermining trust and effective teacher-AI collaboration. To address this gap, we conducted a survey study with K-12 teachers (N = 98) to examine how they perceive and prioritize five key responsible AI values in education across three classroom scenarios: grading, scaffolding science learning, and classroom orchestration. Our findings reveal that fairness and safety emerged as the highest-priority values, while autonomy and performance were rated lower. Transparency showed significant variation, influenced by grade levels and scenarios. Moreover, the qualitative data demonstrated that transparency serves as a mechanism to calibrate trust, foster students' critical thinking, and enhance learning, while also holding the potential downside of overwhelming students if not carefully designed and implemented. Overall, our preliminary insights underscore the need for AI technologies to better align with teachers' priorities, paving the way for more responsible and trustworthy AI tools in educational settings.

CCS Concepts

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

Keywords

Responsible AI, value-sensitive design, education, empirical ethics

ACM Reference Format:
Yaxuan Yin, Shamya Karumbaiah, and Shona Acquaye. 2025. Responsible AI in Education: Understanding Teachers' Priorities and Contextual Challenges. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3715275.3732176

1 Introduction

Artificial Intelligence in Education (AIED) holds significant potential to revolutionize education by facilitating personalized learning, supporting differentiated instruction, and enhancing accessibility [14, 25]. Despite these promising opportunities, the effective integration of AI into educational settings is challenged by the inherent complexity and dynamic nature of educational processes. Factors such as the diversity of learner needs, variability in teaching methods, and evolving curricular goals often conflict with the generalized and narrowly defined frameworks that underpin many AI systems [45]. This misalignment raises critical concerns regarding fairness, transparency, and practical utility, particularly in meeting the needs of educators and vulnerable student populations [5, 34, 47, 50].

While industry leaders and policymakers have proposed principles for Responsible AI (RAI) to guide the ethical development and deployment of AI systems [1–3, 27, 28], these frameworks frequently neglect the perspectives of teachers — key stakeholders in educational contexts who directly interact with and implement these technologies [9, 11, 14, 25, 42, 61]. As the primary implementers of AI technologies in schools, teachers bear the dual responsibility of leveraging AI to improve learning outcomes while safeguarding students — a particularly vulnerable population — from potential harm [4, 14]. However, concerns about fairness, transparency, and safety, coupled with the unpredictable consequences of AI, have led to growing skepticism among K-12 educators. This distrust is reflected in their reluctance to adopt AI tools, with some schools even going so far as to ban their use entirely [9, 42, 56, 61].

Understanding how teachers prioritize and navigate Responsible AI principles is critical for bridging the gap between abstract ethical frameworks and their real-world application in classrooms. While prior qualitative research has provided valuable insights into educators' concerns and experiences, there remains an opportunity to complement these depth-oriented perspectives with approaches capable of revealing broader patterns across diverse educational contexts and scenarios. This is particularly important given the wide range of AI applications in education, including automated grading [10, 15], scaffolding science learning [44, 46], and classroom orchestration [6, 12, 19, 60, 62]. Each scenario presents distinct challenges and affordances that may shape how educators interpret and prioritize different RAI values [27]. Additionally, many teachers may have limited familiarity with responsible AI principles, making it difficult to elicit precise and actionable responses.

To address these challenges, our exploratory study presents an innovative approach to value elicitation [20] by integrating principles from established AI ethics guidelines [27, 28] with scenario-based design. Focusing on three representative AIED use cases, we aim to answer the following research questions:

- RQ1: How do teachers prioritize responsible AI values when integrating AI technologies into classroom instruction?

This work is licensed under a Creative Commons Attribution 4.0 International License.
FAccT '25, June 23–26, 2025, Athens, Greece
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1482-5/2025/06
https://doi.org/10.1145/3715275.3732176

- RQ2: How are the values that teachers prioritize related to various contextual factors?

Based on these research questions and methodological innovations, our study makes three primary contributions:

- Our results demonstrate that fairness and safety consistently emerge as the most important values across three scenarios, underscoring the need for these principles to remain central in the development of educational AI systems.
- We identify a value-action gap, where teachers' stated priorities often diverge from their decisions in context-driven scenarios, underscoring the complexity and context-dependence of ethical decision-making in educational practice.
- Our results reveal that K-12 teachers navigate values both individually and through trade-offs between conflicting priorities, with these decisions strongly shaped by application scenarios and grade levels. This underscores the critical need to design AI tools that are adaptable to the diverse and nuanced pedagogical contexts of education.

2 Related Work

Our study lies at the intersection of Responsible AI (RAI) and education, drawing on prior research to inform both the methodological approach and the study design.

2.1 Advancements in Artificial Intelligence for Education (AIED)

In recent years, the field of Artificial Intelligence in Education (AIED) has witnessed significant advancements, with researchers extensively exploring its potential to deliver personalized learning experiences, enhance teaching efficiency, and streamline administrative tasks [65]. A comprehensive body of research over the past decade has centered around three prominent directions. The first involves leveraging AI to enhance instructional design and pedagogical strategies, with efforts focused on automating student performance assessments [10, 15, 59, 63] and improving feedback mechanisms. These advancements aim to alleviate educators' workloads while ensuring high-quality, scalable teaching practices. The second direction emphasizes student-centered learning through innovations such as adaptive tutoring systems [31, 41, 44, 46], personalized resource recommendations [32, 66], and diagnostic tools for identifying individual learning gaps [35]. These technologies support the customization of educational experiences, address the diverse needs of learners, and promote equitable access to tailored instruction. The third direction explores the transformation of the broader educational ecosystem by redefining the instructor–student dynamic and introducing innovative methodologies in the classroom to foster technology-driven collaboration [6, 12, 19, 60, 62]. These changes aim to create more interactive, engaging, and effective learning environments.

Together, these research directions underpin the foundation of the scenarios explored in this study: 1) GradeAI, an AI-driven grading system, seeks to automate and improve evaluation processes. 2) SciAI, an AI-powered tool for scaffolding science instruction, personalizes content delivery to enhance student engagement and deepen conceptual understanding. 3) OrchestrateAI, an AI-based classroom orchestration tool, optimizes instructional planning and operational efficiency, enabling teachers to manage complex classroom dynamics effectively.

2.2 Responsible AI in Education

The integration of Artificial Intelligence (AI) in education has sparked significant interest and concern regarding its ethical implications [24]. Responsible AI in education aims to ensure that AI systems are designed and deployed in ways that respect ethical principles, enhance learning outcomes, and protect student rights [14, 25]. To guide this effort, we ground our work in well-established AI ethics frameworks [1–3], such as Microsoft's Responsible AI Principles, Google's AI Principles, and the EU's Ethics Guidelines for Trustworthy AI, and draw on prior empirical work in AI ethics (e.g., [27, 28]) to identify five key values: fairness, autonomy, safety, transparency, and performance. These values are not only central to broader Responsible AI conversations but are also widely recognized as critical in the educational context [37, 64]. For example, fairness ensures equitable treatment and outcomes for all students, while autonomy preserves teacher and student agency amid growing automation. Safety encompasses both physical and psychological well-being in AI-mediated environments. Transparency promotes understanding and trust in AI systems, and performance focuses on the effectiveness of AI in improving educational outcomes. Our selection reflects the most frequently encountered and pressing challenges educators face today as AI technologies become more embedded in classrooms.

As AI technologies continue to evolve, new ethical concerns are likely to emerge, warranting ongoing exploration. However, we believe our selected values offer a meaningful and practically grounded starting point for understanding how educators navigate ethical trade-offs in AI adoption. Importantly, these values frequently come into tension with one another [55], highlighting the need for greater teacher involvement in ethical decision-making around AI [8, 13, 18, 36, 40, 53]. For example, efforts to maximize AI performance may inadvertently compromise fairness or diminish teacher and student autonomy. Similarly, enhancing transparency can sometimes introduce complexity that overwhelms educators or learners. These tensions underscore the importance of carefully navigating trade-offs to create AI systems that are both ethical and practical. As such, researchers have emphasized the crucial role of educators in managing these ethical tensions, calling for increased teacher involvement in the development and deployment of AI technologies [5, 34, 47, 50]. This underscores the importance of understanding teachers' perspectives on AI ethics and how they prioritize these key values to support the effective and responsible integration of AI in educational settings.

2.3 Eliciting Stakeholder Values in AI Ethics and Education

Value elicitation is a crucial methodology for understanding stakeholders' priorities and ethical considerations across various domains [20, 52], including technology and education [21]. This approach has proven particularly valuable in AI ethics, helping to uncover the underlying values that should guide AI system development and deployment [27, 58].

However, recent studies exploring ethical intuitions regarding AI technologies in various scenarios have revealed significant complexities and inconsistencies. Classic self-driving car dilemmas, for instance, have exposed notable cross-cultural differences in ethical preferences [7]. Furthermore, investigations have revealed substantial variations and disagreements in ethical judgments in several dimensions: differential judgment of mistakes [23], application of moral norms [39], perceptions of fairness paradigms [51], evaluation of prediction algorithms [22], and prioritization of general responsible AI principles [27]. These variations, influenced by demographics, specific scenarios, and individual human values [27, 48], highlight the intricate, context-dependent nature of ethical considerations in AI applications across different populations and situations.

3 Methods

3.1 Participants

To ensure the integrity and reliability of survey responses, participants were recruited through internal contact lists and in partnership with local K-12 institutions, coordinated by the university's School of Education. Data collection took place over a three-month period. As an incentive, eligible participants were entered into a raffle to win one of ten $50 gift cards. Comprehensive demographic details of the participating teachers are presented in Appendix A.1.

It is important to acknowledge that the demographics of the teacher sample were naturally skewed toward certain categories, reflecting broader trends within the teaching profession [16]. To address this limitation, we outline our plans for future work in Section 6. In our regression analysis for RQ2, we focused on variables that were both non-multicollinear (VIF < 2) and representative of the sample, including teaching experience, self-rated AI experience, grade level, and classroom composition. Consequently, imbalanced features such as gender and political orientation were excluded from the analysis. This approach helped mitigate the effects of demographic skewness while ensuring that the key factors influencing teachers' perspectives on Responsible AI in education were effectively captured.

3.2 Survey Design

The survey underwent iterative development to ensure theoretical alignment, clarity, practicality, and effectiveness in communicating with K-12 teachers. This process included expert review by a professor of philosophy specializing in ethics and human values, think-aloud pilot sessions with four educators, and content validation by undergraduate research assistants. The final survey was administered online via Qualtrics, with the overall design illustrated in Figure 1.

Each participant was randomly assigned to one of three specific application scenarios: SciAI (N = 33), OrchestrateAI (N = 33), or GradeAI (N = 32). Participants were provided with a high-level introduction to their assigned scenario, detailing its features, functions, and associated challenges. This approach was intended to help teachers form a clear understanding of the AI tool's context and functionality, establishing a solid foundation before responding to scenario-specific questions. To confirm their comprehension, a brief quiz followed the scenario introduction to assess whether participants had grasped the basic concepts and features of their assigned AI tool before proceeding to the subsequent sections.

To investigate teachers' preferences for responsible AI values within the context of specific AI technologies, we employed three distinct question types: Likert-scale questions, action-oriented scenarios, and value conflict assessments, supplemented by open-ended responses for qualitative insight. For instance, as illustrated in Figure 2, if a teacher was presented with the GradeAI scenario, they were first asked a Likert-scale question to assess the importance of transparency, such as: "How important is it for you that GradeAI explicitly outlines the criteria and algorithms it uses to evaluate student work and generate feedback?" Response options ranged from "very unimportant" to "very important". Next, they were presented with an action-oriented scenario, such as: "Faced with this situation, which action would you like to take?" with predefined options reflecting practices aligned with transparency. To explore trade-offs, a value conflict assessment was posed, such as: "In this situation, which action would you prioritize?" where transparency was contrasted with another value, like safety. Finally, each section included an open-ended question inviting teachers to elaborate on their responses, providing qualitative data to capture nuanced perspectives and rationales behind their choices. The full set of survey questions is illustrated in Appendix B.

3.3 Data Processing and Analytical Methods

3.3.1 Quantitative Survey Data. Teachers' pedagogical values are rarely aligned exclusively with a single responsible value. Instead, teachers often balance and implement multiple values simultaneously, depending on the context. We observed this complexity during our piloting phase with practicing teachers and PhD students who had K-12 teaching experience. Given the nuanced and multifaceted nature of educational values, we converted teachers' responses into a continuous variable format that allowed respondents to express degrees of alignment, providing a more accurate and meaningful representation of the interplay between their values and decision-making processes.

As shown in Figure 2, all questions offered five response options, which were systematically converted into a standardized scale. For Likert-scale questions, responses ranged from "Very important" (scored as 2) to "Very unimportant" (scored as -2), with intermediate options proportionally scaled to reflect varying degrees of agreement. A similar scoring system was applied to action-oriented scenarios and value conflict assessments, where each response reflected a directional inclination toward the value being studied. In value conflict scenarios, selecting one option assigned an inverse score to the opposing value under consideration. For example, endorsing one value resulted in a corresponding negative score for its counterpart.

To facilitate analysis, average scores were calculated where necessary, standardizing all values within the range of -2 to 2. For instance, as illustrated in Figure 2, the overall score for transparency was calculated as $(+2 - 1 + 2)/3 = 1$, where the denominator represents the total number of questions related to transparency. This scoring system was consistently applied to other values in the three scenarios that we studied. This scoring methodology not only enabled consistent aggregation and comparison across participants

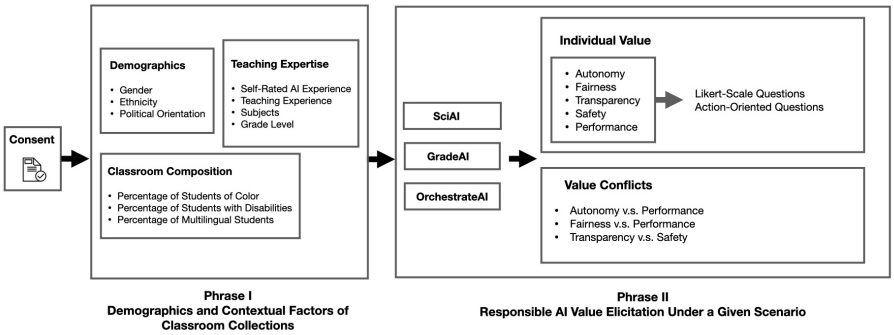
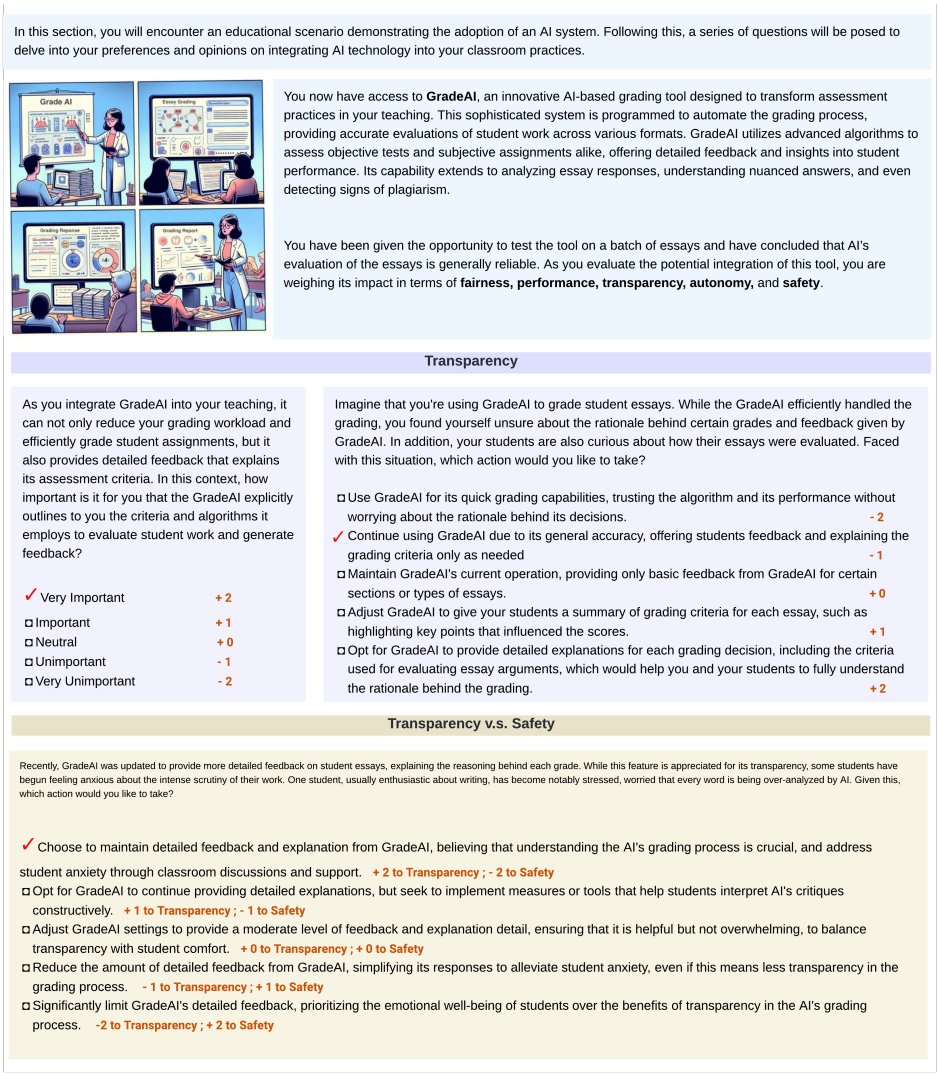


Figure 1: Survey Design Flow.



In this section, you will encounter an educational scenario demonstrating the adoption of an AI system. Following this, a series of questions will be posed to delve into your preferences and opinions on integrating AI technology into your classroom practices.

You now have access to **GradeAI**, an innovative AI-based grading tool designed to transform assessment practices in your teaching. This sophisticated system is programmed to automate the grading process, providing accurate evaluations of student work across various formats. GradeAI utilizes advanced algorithms to assess objective tests and subjective assignments alike, offering detailed feedback and insights into student performance. Its capability extends to analyzing essay responses, understanding nuanced answers, and even detecting signs of plagiarism.

You have been given the opportunity to test the tool on a batch of essays and have concluded that AI's evaluation of the essays is generally reliable. As you evaluate the potential integration of this tool, you are weighing its impact in terms of **fairness, performance, transparency, autonomy,** and **safety**.

Transparency

As you integrate GradeAI into your teaching, it can not only reduce your grading workload and efficiently grade student assignments, but it also provides detailed feedback that explains its assessment criteria. In this context, how important is it for you that the GradeAI explicitly outlines to you the criteria and algorithms it employs to evaluate student work and generate feedback?

- ✓ Very Important + 2
- ◻ Important + 1
- ◻ Neutral + 0
- ◻ Unimportant - 1
- ◻ Very Unimportant - 2

Imagine that you're using GradeAI to grade student essays. While the GradeAI efficiently handled the grading, you found yourself unsure about the rationale behind certain grades and feedback given by GradeAI. In addition, your students are also curious about how their essays were evaluated. Faced with this situation, which action would you like to take?

- ◻ Use GradeAI for its quick grading capabilities, trusting the algorithm and its performance without worrying about the rationale behind its decisions. - 2
- ✓ Continue using GradeAI due to its general accuracy, offering students feedback and explaining the grading criteria only as needed - 1
- ◻ Maintain GradeAI's current operation, providing only basic feedback from GradeAI for certain sections or types of essays. + 0
- ◻ Adjust GradeAI to give your students a summary of grading criteria for each essay, such as highlighting key points that influenced the scores. + 1
- ◻ Opt for GradeAI to provide detailed explanations for each grading decision, including the criteria used for evaluating essay arguments, which would help you and your students to fully understand the rationale behind the grading. + 2

Transparency v.s. Safety

Recently, GradeAI was updated to provide more detailed feedback on student essays, explaining the reasoning behind each grade. While this feature is appreciated for its transparency, some students have begun feeling anxious about the intense scrutiny of their work. One student, usually enthusiastic about writing, has become notably stressed, worried that every word is being over-analyzed by AI. Given this, which action would you like to take?

- ✓ Choose to maintain detailed feedback and explanation from GradeAI, believing that understanding the AI's grading process is crucial, and address student anxiety through classroom discussions and support. + 2 to Transparency ; - 2 to Safety
- ◻ Opt for GradeAI to continue providing detailed explanations, but seek to implement measures or tools that help students interpret AI's critiques constructively. + 1 to Transparency ; - 1 to Safety
- ◻ Adjust GradeAI settings to provide a moderate level of feedback and explanation detail, ensuring that it is helpful but not overwhelming, to balance transparency with student comfort. + 0 to Transparency ; + 0 to Safety
- ◻ Reduce the amount of detailed feedback from GradeAI, simplifying its responses to alleviate student anxiety, even if this means less transparency in the grading process. - 1 to Transparency ; + 1 to Safety
- ◻ Significantly limit GradeAI's detailed feedback, prioritizing the emotional well-being of students over the benefits of transparency in the AI's grading process. -2 to Transparency ; + 2 to Safety

Figure 2: Example illustrating how scoring was applied to evaluate a participant's response in the GradeAI scenario through Likert-scale questions, action-oriented scenarios, and value conflict assessments. Value labels (e.g., Transparency or Transparency vs. Safety) and scoring indicators were added for reference only and were not shown to participants in the actual survey.

and scenarios but also provided nuanced insights into the complexities of teachers' value priorities and their application in varied educational contexts.

3.3.2 Qualitative Survey Data. As noted in Section 3.2, we collected open-ended responses from teachers in which they explained the rationale behind their answers to each question. To analyze these responses, we employed a thematic analysis approach [57] to identify and interpret patterns in the data. To ensure reliability and minimize individual bias, two coders cross-checked each other's work and resolved any discrepancies through discussion until consensus was reached.

4 Results

4.1 RQ1: How do teachers prioritize responsible AI values when integrating AI technologies into classroom instruction?

To examine general patterns in how the five responsible AI values are prioritized by K-12 teachers, we analyze the overall score distributions, as detailed in Section 4.1.1. Our survey, which includes Likert-scale questions, action-oriented scenarios, and value conflict assessments, investigates the gap between teachers' ideal commitment to these values and how these values influence their practical decision-making in pedagogical contexts (see Section 4.1.2). Furthermore, the value conflict questions provide insights into how teachers navigate conflicting values, highlighting the actions they are more likely to prioritize when faced with competing priorities (see Section 4.1.3).

4.1.1 Fairness and Safety Take Priority, While Transparency Shows High Variation.

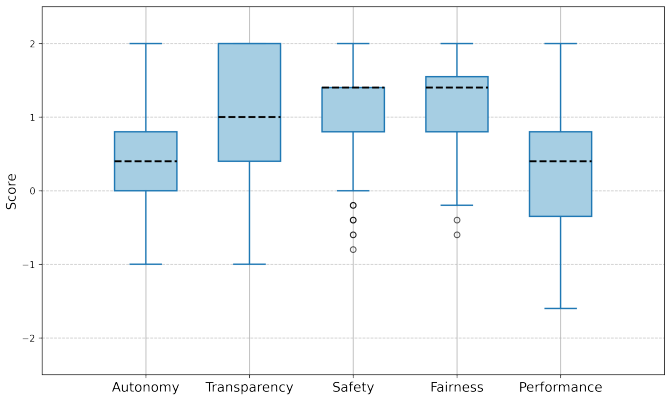


Figure 3: Boxplot of Responsible AI Value Scores. This figure displays the median, interquartile range, and outliers for each value, illustrating how teachers vary in their prioritization of responsible AI principles.

As shown in Figures 3 and 4, our results reveal that safety and fairness emerge as the most highly prioritized values, each demonstrating the highest median scores (med=1.5) and compact interquartile ranges (IQRs). This indicates a strong consensus among participants regarding the importance of these values. In contrast, transparency exhibits a slightly lower median score (med=1) and greater variability, reflecting less consistency in its prioritization. Notably, autonomy (emphasizing student autonomy in our survey) and performance are the least prioritized values, both receiving a median score of 0.5.

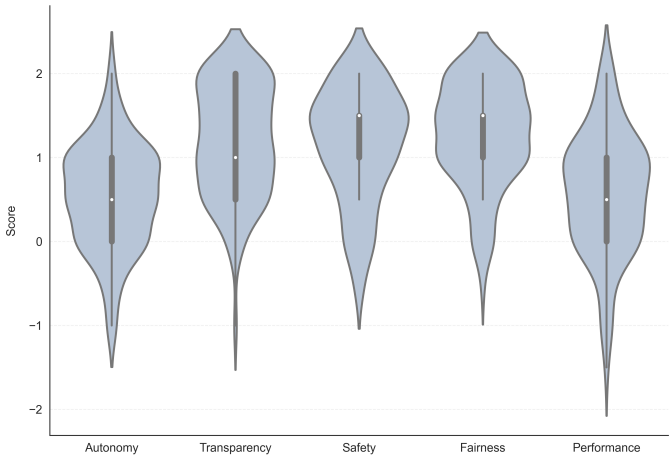


Figure 4: Violin Plot of Responsible AI Value Scores. The plot shows the distribution and density of responses for each value, highlighting both central tendencies and variability in teachers' prioritizations.

To further examine these patterns, we conducted a post-hoc analysis. A Friedman test confirmed statistically significant differences among the RAI categories. Pairwise comparisons using the Nemenyi test revealed that safety and fairness scored significantly higher than performance and autonomy. Additionally, transparency scored significantly higher than performance. However, no significant differences were observed between autonomy and performance or between safety and fairness. Full results of the pairwise comparisons are presented in Appendix A.2.

Among the lower-prioritized values, autonomy stood out as an area where teachers expressed complex and sometimes conflicting views. While students' autonomy was widely acknowledged as valuable [43], many teachers expressed reservations about fully prioritizing it. As T76 noted, *"Middle school students do not necessarily have an appropriate understanding of the use of technology."* Similarly, T92 emphasized, *"Self-direction is an area many of my students struggle with."* These perspectives suggest that autonomy is seen as aspirational, requiring scaffolding and development over time. Additionally, the fear of fostering over-reliance on AI contributed to the lower prioritization of autonomy. Teachers expressed apprehension about the potential for AI to undermine critical thinking and independent decision-making. As T9 shared, *"Fostering a reliance on AI is one of my greatest fears as the technology continues to develop. It should be used as a supplementary tool rather than a primary form of instruction."* This concern reflects a cautious approach to autonomy, where students' independence is valued but must be balanced with thoughtful integration of AI as a support mechanism.

Overall, our findings reveal significant differences in how K-12 teachers prioritize responsible AI values. While values such as safety and fairness are strongly prioritized, others, like autonomy and performance, are significantly deprioritized. This clustering of value preferences offers critical insights for guiding the development of AI systems that align with educators' ethical priorities and practical needs.

4.1.2 Discrepancy Between Teachers' Stated Priorities and Action-Oriented Decisions.

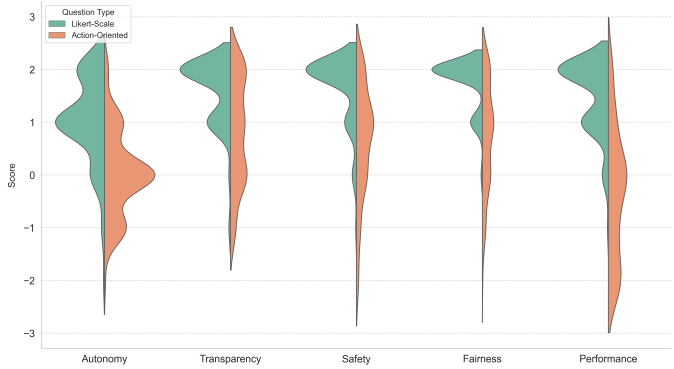


Figure 5: Distribution of Likert-Scale vs. Action-Oriented Responses Across Responsible AI Values. The violin plot compares Likert-scale responses (green) with action-oriented decisions (orange) for autonomy, transparency, safety, fairness, and performance. The width represents response density, highlighting differences in prioritization and variability.

In addition to analyzing how K-12 teachers prioritize RAI, we identified significant differences between their Likert-scale responses, which reflect conceptual commitments, and their action-oriented responses, which represent practical decisions in classroom and pedagogical settings. As shown in Figure 5 and in Appendix A.3, the violin plot and t-tests revealed that participants rated the importance of Responsible AI values significantly higher in Likert-scale responses compared to their intentions in action-oriented responses across all values. This suggests that while teachers conceptually prioritize these values, their practical implementation does not consistently align with these ideals.

Furthermore, the responses in action-oriented questions exhibited greater variability, as indicated by higher standard deviations, except in the case of autonomy. For instance, the standard deviation for the action-oriented transparency score (std = 1.00) was substantially higher than that for the Likert-scale score (std = 0.63), reflecting a broader range of responses when teachers were tasked with applying conceptual values to real-world decisions. Overall, our results suggest that practical constraints or contextual challenges might contribute to greater variability, making it harder for teachers to consistently align their decisions with their stated ideal value priorities. These findings highlight the importance of addressing real-world challenges when designing and integrating educational AI systems, a topic further explored in Section 5.1.

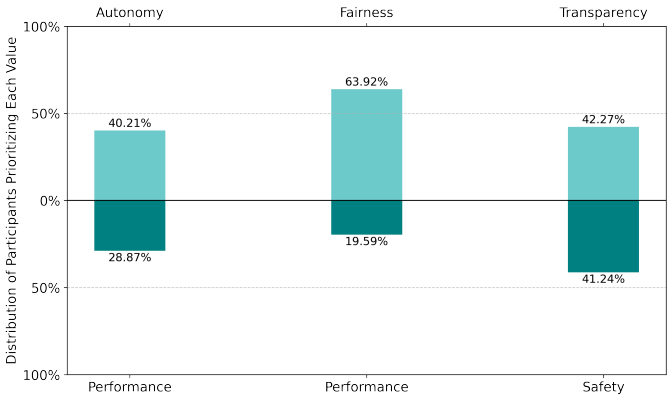


Figure 6: Percentage of Participants Assigning the Highest Score to Each Value When in Conflicts.

4.1.3 Fairness Takes Priority Over Performance in Conflicts.

Moreover, when participants encountered value conflicts in the given scenarios — autonomy vs. performance, fairness vs. performance, and transparency vs. safety — we calculated the net scores for each value in the conflict and conducted a one-sample t-test to determine whether the mean net score significantly deviated from zero, indicating a clear preference for one value over the other. The results revealed that only fairness was strongly prioritized over performance in this conflict, as indicated by the t-test results ($t = 5.99, p < 0.0001$).

Our qualitative data from open responses reinforces the idea that while teachers acknowledge AI's performance and its potential to enhance efficiency, they are deeply concerned about its limitations — particularly biases in AI systems and their potential to exacerbate existing inequities. Additionally, teachers emphasize the necessity of addressing the diverse needs of learners to ensure equitable outcomes. For example, in GradeAI, educators highlighted challenges in grading multilingual students and stressed the importance of fairness. As T32 stated, *"We cannot allow AI to harm students unfairly, particularly those already marginalized."* This sentiment was echoed in OrchestrateAI, where T5 cautioned against tools that perpetuate systemic biases, noting, *"It's not acceptable to use a classroom tool that perpetuates separation and bias."* In SciAI, fairness remains central, with teachers advocating for tools that adapt to diverse learning styles and meet the unique needs of students, particularly those who struggle. Our participants called for systems that can accommodate variations in abilities, backgrounds, and learning preferences. As T18 observed, *"Students have unique learning styles and can benefit from a diversified teaching approach."* Across all tools, there was a consistent call for flexibility and adaptability, ensuring that AI tools do not overlook or disadvantage any group of students. While AI offers efficiency, teachers consistently prioritize ethical considerations over speed. In GradeAI, T12 remarked, *"If I need to manually review essays to ensure fairness, it's my job to do so..."* Similar sentiments were expressed in OrchestrateAI and SciAI, where efficiency was viewed as a secondary benefit to creating equitable and supportive learning environments.

Turning to the transparency vs. safety conflict, preferences were more balanced, with 42.27% favoring transparency, 41.24% favoring safety, and 16.49% considering the two values equally important. The interplay between transparency and safety in AI systems reveals a nuanced tension, as educators strive to balance the need for constructive feedback with the emotional well-being of their students. Across these tools, teachers emphasized that students' safety is foundational to effective learning. As T23 noted, *"Learning doesn't happen when the student doesn't feel safe or is too stressed."* Similarly, in OrchestrateAI, T93 stressed the importance of not overwhelming students, stating, *"Students' emotional well-being is more important than reducing the teachers' workload."* At the same time, many educators acknowledged the value of transparency and detailed feedback for fostering resilience and academic growth. As one teacher observed, *"Learning to deal with feedback is crucial to growing as a writer,"* and T39 emphasized, *"Detailed feedback is the point, so I would want to prioritize that while addressing the anxiety issues in class."*

Across all three systems, a recurring theme was the need for flexibility and teacher mediation. Teachers consistently advocated for customizable feedback systems that empower educators to tailor AI outputs to the unique needs of their students. As T11 noted, *"I'd like to have control of the level of feedback given depending on the assignment,"* T2 further highlighted the importance of teacher oversight, stating, *"Teachers understand their little humans better than a bot and could adjust the AI-generated feedback to suit their students' individual well-being needs."* Teachers across GradeAI, OrchestrateAI, and SciAI also emphasized the importance of helping students learn to interpret and use feedback effectively. As T72 remarked, *"This would be a great chance for a lesson in how to take, interpret, and respond to feedback."* This underscores the potential of AI systems as tools for both academic and socio-emotional development.

OrchestrateAI further contributed insights on the importance of balancing transparency with sensitivity to students' needs, particularly in managing classroom dynamics. Teachers recognized that while AI can help streamline certain processes, it must not compromise fairness or inclusivity. As T53 explained, *"If the AI system cannot adapt to diverse student backgrounds, it risks perpetuating the inequality already inherent in the system."* This concern echoes the broader apprehension about AI's potential to exacerbate existing disparities, highlighting the need for systems that prioritize fairness, adaptability, and human oversight.

Similarly, in the autonomy vs. performance conflict, 40.21% of participants favored autonomy, 28.87% favored performance, and 30.92% considered the two values equally important. Our results indicate that no statistically significant preference was observed in either conflict. In SciAI, T14 stated, *"Students deserve the opportunity to make their own choices, but teachers have curricular requirements and standards to meet."* In GradeAI, autonomy was recognized as a way to encourage self-reflection and active learning. As T49 noted, *"Student autonomy is important, but it needs to coexist with clear goals and educational aims set by the teacher."* This perspective aligns with OrchestrateAI feedback, where teachers suggested limited autonomy to personalize learning without undermining the system's reliability. As T12 explained, *"Allowing limited input ensures students can engage in the process while keeping functionalities consistent and reliable."*

4.2 RQ2: How are the values that teachers prioritize related to various contextual factors?

As shown in Table 1 and Table 2, we performed two linear regression analyses to investigate the effects of various contextual factors — such as application scenarios, classroom composition, and grade level — on individual value scores and value conflicts, respectively. In both models, GradeAI serves as the baseline, a standard approach in regression analysis with categorical variables, enabling us to quantify the differences between each scenario and the reference point.

Our results indicate that transparency is significantly deprioritized in the OrchestrateAI scenario compared to the GradeAI scenario, with a decrease of 0.53. Similarly, transparency in the SciAI scenario shows a notable reduction of 0.58 relative to GradeAI. These findings suggest that both the OrchestrateAI and SciAI scenarios place less emphasis on transparency compared to the GradeAI scenario. Turning to another value — performance, we observed a positive correlation with the SciAI scenario. Specifically, performance was rated higher in the SciAI scenario, with an increase of 0.38 compared to GradeAI, highlighting a stronger emphasis on performance within the SciAI context. Furthermore, our models did not detect significant differences in autonomy, safety, or fairness across scenarios, indicating that these values were consistently rated irrespective of the scenario.

Moreover, our analysis also revealed a significant positive association between transparency and grade level (coefficient = 0.21**), suggesting that K-12 teachers place greater importance on transparency as students progress to higher grades. This finding is further supported by qualitative evidence from participant feedback. As P7 explained, *"With that said, students often do not take the time to dive deep into feedback, so I am not sure how practical or useful a full explanation of all grading decisions is, unless a specific student asks for additional clarification or there are consistent concerns about a particular part of the grade."* Similarly, P11 noted, *"At lower grade levels, I would stick with the summary approach to avoid overwhelming students, but at higher grade levels, I would provide detailed explanations as students are better at multitasking and making effective use of feedback."*

In another model examining value conflicts, we identified nuanced patterns influenced by scenarios and grade levels, as detailed in Table 2. Using net score as the dependent variable, we observed significant associations between K-12 teachers' priorities and the scenarios they participated in, particularly in two key value conflicts: fairness vs. performance and transparency vs. safety. The net score reflects the degree to which one value is prioritized over another, with positive scores indicating a stronger preference for the first value (e.g., fairness or transparency), and negative scores indicating a stronger preference for the second value (e.g., performance or safety). In the fairness vs. performance conflict, K-12 teachers exhibited a significantly stronger prioritization of fairness over performance in the OrchestrateAI scenario, as indicated by a positive net score increase (coefficient = 0.88*) compared to their priorities in GradeAI.

Conversely, in the transparency vs. safety conflict, teachers in the OrchestrateAI scenario displayed a lower net score for transparency

Variable	Transparency	Performance	Autonomy	Safety	Fairness
const	1.18**	-0.10	0.19	0.32	0.79
Scenario::OrchestrateAI	-0.53***	0.07	0.10	0.12	0.13
Scenario::SciAI	-0.58***	0.38**	0.07	-0.18	0.04
% Students of Color	-0.33	-0.26	-0.20	-0.02	0.03
% Students with Disabilities	-0.30	-0.08	-0.01	0.18	0.06
% Multilingual Students	0.33	-0.27	-0.25	0.61	-0.11
Self-Rated AI Experience	-0.01	0.11	0.03	0.09	-0.01
Teaching Experience	-0.07*	0.07	0.01	0.06	0.01
Grade Level	0.21**	0.04	0.08	0.08	0.12
R-squared	0.33	0.20	0.06	0.10	0.03
Adj. R-squared	0.27	0.13	-0.02	0.02	-0.05

Table 1: Regression Results for Individual Responsible AI Values. (*p < 0.05, **p < 0.01, *** p < 0.001)

Variable	Autonomy vs. Performance	Fairness vs. Performance	Transparency vs. Safety
const	0.10	-0.98	-0.48
Scenario::OrchestrateAI	-0.57	0.88*	-1.10*
Scenario::SciAI	0.17	-0.23	-0.39
% Students of Color	-0.16	-0.13	-1.24
% Students with Disabilities	1.14	-0.45	0.26
% Multilingual Students	-0.84	0.58	-1.02
Self-Rated AI Experience	0.27	0.08	0.27
Teaching Experience	-0.05	-0.03	-0.17
Grade Level	-0.06	0.57*	0.35
R-squared	0.09	0.13	0.13
Adj. R-squared	0.01	0.06	0.06

Table 2: Regression Results for Values In Conflicts. (*p < 0.05, **p < 0.01, *** p < 0.001)

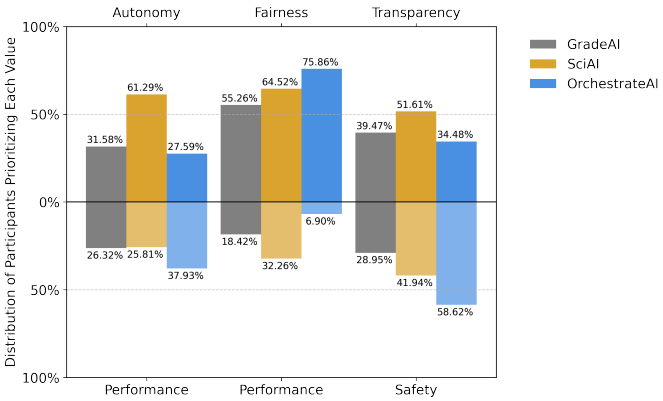


Figure 7: Comparative Distribution of Value Conflict Scores Across Application Scenarios

compared to GradeAI (coefficient = -1.10*), signifying a stronger prioritization of safety over transparency in the OrchestrateAI context. Figure 7 further illustrates these shifts. While in GradeAI, the majority of K-12 teachers (39.47%) favored transparency over safety (positive net scores), the trend reversed in OrchestrateAI, where 58.62% of teachers favored safety over transparency (negative net scores). This reversal highlights how scenario contexts influence teachers' value prioritization.

Additionally, we found that grade levels also impact preferences in the fairness vs. performance conflict. As grade levels increase, teachers demonstrated a stronger preference for fairness over performance (coefficient = 0.57*), reflected in higher positive net scores. This suggests that teachers in higher-grade contexts place greater emphasis on fairness when navigating value conflicts.

5 Discussion

Our preliminary work sheds light on teachers' perspectives on AI ethics, addressing a critical gap in the development of Artificial Intelligence in Education (AIED). By uncovering teachers' value priorities, we contribute to the design of AIED systems that better align with both educational objectives and ethical standards. This alignment is particularly crucial for the three scenarios explored in our study, as they represent key areas where AIED tools intersect with classroom practices. Our findings underscore the importance of integrating educators' insights into the development of AIED systems to ensure they support responsible, equitable, and effective AI integration in schools. Specifically, we identify the following implications for promoting more educator-informed and ethically grounded AIED solutions.

5.1 Not All AI Ethics Values Are Treated Equally

Our preliminary findings reveal a consistent pattern: fairness and safety emerge as the most prioritized values across three scenarios, underscoring teachers' strong emphasis on creating a secure and equitable environment for students. Transparency, while important, shows context-dependent variability, suggesting nuanced perspectives on its role as AI systems grow in complexity. In contrast, (student) autonomy and (AI) performance are generally viewed as less critical in the context of responsible AI in education.

Educational AI development has traditionally prioritized optimizing overall performance, often at the expense of equitable outcomes for minority students [29]. Our findings reveal that K-12 teachers place a high priority on fairness and safety as essential considerations in the design of AI systems, prioritizing these values over

performance. This aligns with the growing emphasis on social justice goals in education [38]. These results underscore the critical need to center equity and inclusivity in AI system development, ensuring that all students benefit equitably from these technologies. On the other hand, our findings on the low priority assigned to student autonomy in AI systems raise some open questions. Research in the learning sciences research emphasizes the role of student choice and agency in learning [43]. Yet, in practice, it is often difficult to translate these ideals due to practical constraints. This disconnect suggests the need for broader stakeholder involvement in designing AI tools. Including voices from students, researchers, and parents could provide valuable perspectives on how autonomy can be supported in ways that balance pedagogical ideals with real-world constraints. For instance, students could offer insights into their needs and preferences for exercising agency, while parents and researchers could help identify contextual factors that either enable or constrain autonomy in real-world settings.

Additionally, our study uncovered a marked discrepancy between teachers' stated value priorities and their action-oriented decisions. While Likert-scale responses indicated high conceptual importance across all values, action-oriented questions revealed significantly lower scores, albeit with greater variability. This suggests that when confronted with practical scenarios, teachers' decision-making becomes more nuanced and context-dependent than their stated values might indicate. This discrepancy reflects the value-action gap [49], highlighting the complexity of ethical decision-making as teachers adapt to the constraints of real-world educational settings. While our study statistically demonstrates this variation, it does not unpack the mechanisms driving this divergence. To address this, future research could adopt participatory design methods[54], engaging teachers as co-creators to delve deeper into the interplay between values and actions. This approach could illuminate how conceptual priorities translate — or fail to translate — into practice and identify contextual factors that shape these decisions. Such collaborative exploration could inform the design of tools or interventions that better support teachers in aligning their actions with their values, ultimately fostering more ethical and contextually grounded educational practices.

5.2 Transparency in AIED: Catalyst for Learning and Trust or Burden for Students

While transparency has not been consistently prioritized to the same extent as fairness and safety, it exhibits significant variability in how K-12 teachers perceive its role within pedagogical settings, as revealed by our qualitative data. Among participants who advocated for high importance of transparency, Artificial Intelligence in Education (AIED) was seen as a pivotal mechanism for fostering student learning and aligning teachers' values, ultimately calibrating trust in AI systems. Specifically, transparency equips K-12 teachers with critical information, enabling them to identify and address misalignments between AI outputs and their pedagogical goals.

Conversely, participants who deprioritized transparency in AIED often expressed concerns about how transparency is communicated to students. On one hand, detailed feedback was recognized as a valuable resource for fostering a growth mindset, enabling students to critically evaluate their work and make continuous improvements. As one participant explained, transparency allows students to *"learn to take, interpret, and respond to feedback"*, a skill integral to their academic and personal development. On the other hand, concerns emerged about the potential drawbacks of transparency. These included the risk of students spending excessive time reading and processing detailed feedback, which could detract from other learning activities, as well as challenges in ensuring that feedback from AI systems is actionable and effectively utilized. As one participant noted, *"Too much feedback is overwhelming, a few actionable and constructive points are more helpful."*

Taken holistically, our results show that transparency influences both key stakeholders in AIED — students and teachers. To be effective, transparency should align with accessible language that resonates with both groups. Simplifying feedback for clarity and brevity can help mitigate cognitive overload while preserving its instructional value, particularly given the limited time students and teachers have to engage with detailed feedback. Moreover, in AIED, transparency is more than just a technical feature of AI systems, it plays a foundational role in fostering a collaborative environment where AI systems are trusted and meaningfully integrated into educational practices. This aligns with prior work advocating for a human-centered design approach to transparency design and need-finding [17, 30, 33]. Future work should continue to focus on the needs of stakeholders in AIED, ensuring that transparent feedback mechanisms function both as triggers for learning and as foundations for trust between educators, students, and AI technologies.

5.3 Pedagogical Context Matters

Building on prior research that highlights responsible AI values as being more critical in the medical context than in other domains [27], our study examines how these values are prioritized in AIED and reveals nuanced insights into educators' value preferences across diverse pedagogical scenarios. Our findings show that K-12 teachers' perceptions of responsible AI are shaped by both the specific educational context and the grade levels they serve. For instance, transparency is associated with higher grade levels and holds greater significance in grading-focused AI tools (GradeAI) compared to classroom orchestration (OrchestrateAI) or science learning (SciAI). This suggests that as students progress through grade levels, educators place increasing emphasis on transparent processes, particularly in contexts where assessment and feedback play a central role, reinforcing the importance of explainable reasoning in AI-assisted grading tasks [27]. Likewise, performance is rated higher in science learning scenarios (SciAI), indicating that teachers prioritize efficiency and effectiveness in AI systems designed to support instructional and scientific learning. These distinctions point to a broader pattern: different pedagogical goals, such as managing classrooms, supporting individualized learning, or enhancing subject-matter mastery, fundamentally shape which AI values teachers consider most salient.

When responsible AI values come into conflict, teachers' preferences are further shaped by the scenario and the grade level involved. In OrchestrateAI, for instance, fairness and safety are consistently prioritized over performance and transparency, reflecting

a heightened focus on equity and student well-being in classroom management tools. Notably, safety is prioritized over performance more in OrchestrateAI than in GradeAI, echoing concerns about students' physical and mental well-being in AI-driven classroom interactions [26]. Similarly, fairness is increasingly prioritized over performance at higher grade levels, suggesting that teachers working with older students are more attuned to promoting equity, even if doing so comes at the cost of maximizing performance outcomes.

These findings underscore the importance of understanding and addressing the varying priorities of teachers across different educational settings. Designers and practitioners must consider these contextual differences to create AI systems that are both adaptable and responsive to the specific needs of educators. For example, while transparency may be less emphasized in early-grade classroom tools, its rising importance at higher grade levels highlights the need for adjustable transparency features tailored to different age groups. In contrast, performance may be paramount in science learning tools like SciAI, whereas classroom orchestration tools such as OrchestrateAI should elevate fairness and safety as guiding principles.

Ultimately, AI systems in education must strike a thoughtful balance among competing values to reflect the diverse, context-specific goals of K-12 educators and their students. By aligning design with these nuanced priorities, AI technologies can more effectively support equitable, meaningful, and pedagogically appropriate teaching and learning experiences.

6 Limitations

The value elicitation approach adopted in this study provided valuable preliminary insights into the perspectives of K-12 educators. While the sample size aligns with standards for initial exploratory research in similar survey studies [27], the composition of our participants is skewed toward certain demographic groups. This may not fully capture the diversity of K-12 educators, particularly with regard to race, socioeconomic status, and regional differences. Such homogeneity limits the generalizability of our findings, as the perspectives of underrepresented groups may be underrepresented. To address these limitations, future research should focus on recruiting a broader and more diverse pool of participants. Expanding the study to include educators from a wider range of demographic backgrounds and international contexts would allow for more comprehensive insights into responsible AI values. Additionally, follow-up qualitative methods, such as interviews or focus groups, could further explore the nuanced ways in which educators prioritize these values and the contextual factors driving their decisions.

7 Conclusion

As artificial intelligence (AI) increasingly permeates K-12 education, understanding how teachers prioritize ethical considerations becomes crucial for responsible implementation. This exploratory study addresses the gap in empirical research on educators' AI ethics priorities. Using a value elicitation method, we surveyed K-12 teachers to assess their prioritization of responsible AI values across various educational scenarios. Our findings show that safety and fairness consistently rank as the most important values, regardless of the scenario or context, reflecting teachers' emphasis on creating secure and equitable learning environments. However, when values come into conflict, the prioritization shifts depending on the specific scenario and context. These results highlight the need to incorporate more diverse ethical perspectives into the AI development process. As AI becomes more embedded in education, ensuring alignment with the values of educators will be crucial for promoting positive, equitable, and responsible use of AI technologies in classrooms.

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. We also appreciate Harry Brighouse and all the think-aloud participants for their valuable feedback.

References

[1] [n. d.]. Responsible AI. https://research.google/teams/responsible-ai/. Accessed: 2024-09-01.
[2] [n. d.]. Responsible AI: Principles and approach. https://www.microsoft.com/en-us/ai/principles-and-approach/. Accessed: 2024-09-01.
[3] [n. d.]. What is AI ethics? https://www.ibm.com/topics/ai-ethics. Accessed: 2024-09-01.
[4] Muhammad Abbas, Farooq Ahmed Jam, and Tariq Iqbal Khan. 2024. Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education* 21, 1 (2024), 10.
[5] Heba Aly, Kaileigh Angela Byrne, and Bart Knijnenburg. 2024. Perceived Trustworthiness of Human vs. AI Instructors in Digital Privacy Education for Older Adults. In *Companion Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) *(IUI '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 107–112. https://doi.org/10.1145/3640544.3645238
[6] 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.
[7] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. The moral machine experiment. *Nature* 563, 7729 (2018), 59–64.
[8] Solon Barocas and Danah Boyd. 2017. Engaging the ethics of data science in practice. *Commun. ACM* 60, 11 (2017), 23–25.
[9] A. Bhimdiwala, R.C. Neri, and L.M. Gomez. 2022. Advancing the Design and Implementation of Artificial Intelligence in Education through Continuous Improvement. *International Journal of Artificial Intelligence in Education* 32 (September 2022), 756–782. https://doi.org/10.1007/s40593-021-00278-8
[10] Bryce Boe, Charlotte Hill, Michelle Len, Greg Dreschler, Phillip Conrad, and Diana Franklin. 2013. Hairball: lint-inspired static analysis of Scratch projects. In *Proceedings of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)* (Denver, Colorado, USA). Association for Computing Machinery, 215–220. https://doi.org/10.1145/2445196.2445265
[11] Veronika Bogina, Alan Hartman, Tsvi Kuflik, and Avital Shulner-Tal. 2021. Educating Software and AI Stakeholders About Algorithmic Fairness, Accountability, Transparency and Ethics. *International Journal of Artificial Intelligence in Education* 32 (04 2021), 1–26. https://doi.org/10.1007/s40593-021-00248-0
[12] 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. https://doi.org/10.1145/3657604.3664704
[13] Claudia Cappelli, Herbet Cunha, Bruno Gonzalez-Baixauli, and Julio Cesar Sampaio do Prado Leite. 2010. Transparency versus security: Early analysis of antagonistic requirements. In *Proceedings of the 2010 ACM Symposium on Applied Computing*. ACM, 298–305.
[14] Irene-Angelica Chounta, Emanuele Bardone, Andra Raudsep, and Margus Pedaste. 2022. Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education. *International Journal of Artificial Intelligence in Education* 32 (2022), 725–755. https://doi.org/10.1007/s40593-021-00243-5
[15] Code.org. 2024. Code.org Platform for Computer Science Curricula. https://code.org/ Accessed: 2024-09-01.

[16] Diana D'amico, Robert Pawlewicz, Penelope Earley, and Adam McGeehan. 2017. Where Are All the Black Teachers? Discrimination in the Teacher Labor Market. *Harvard Educational Review* 87 (03 2017), 26–49. https://doi.org/10.17763/1943-5045-87.1.26

[17] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, New York, NY, USA, 211–223. https://doi.org/10.1145/3172944.3172961

[18] Michael D Ekstrand, Rezvan Joshaghani, and Hoda Mehrpouyan. 2018. Privacy for all: Ensuring fair and equitable privacy protections. In *Conference on Fairness, Accountability, and Transparency*. PMLR, 35–47.

[19] 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. https://doi.org/10.1016/j.compedu.2023.104864

[20] Baruch Fischhoff. 1991. Value Elicitation: Is There Anything in There? *American Psychologist* 46 (08 1991), 835–847. https://doi.org/10.1037/0003-066X.46.8.835

[21] Batya Friedman, David G. Hendry, and Alan Borning. 2017. A Survey of Value Sensitive Design Methods. *Found. Trends Hum.-Comput. Interact.* 11, 2 (nov 2017), 63–125. https://doi.org/10.1561/1100000015

[22] Nina Grgic-Hlaca, Elissa M Redmiles, Krishna P Gummadi, and Adrian Weller. 2018. Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 903–912.

[23] César A Hidalgo, Diana Orghian, Jordi Albo Canals, Filipa De Almeida, and Natalia Martin. 2021. *How Humans Judge Machines*. MIT Press.

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

[25] Wayne Holmes and Kaska Porayska-Pomsta. 2022. *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates.* https://doi.org/10.4324/9780429329067

[26] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?. 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, 1–16. https://doi.org/10.1145/3290605.3300830

[27] 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* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 310–323. https://doi.org/10.1145/3531146.3533097

[28] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1 (09 2019). https://doi.org/10.1038/s42256-019-0088-2

[29] Shamya Karumbaiah, Jaclyn Ocumpaugh, and Ryan S Baker. 2022. Context matters: Differing implications of motivation and help-seeking in educational technology. *International Journal of Artificial Intelligence in Education* 32, 3 (2022), 685–724.

[30] Sunnie SY Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17. https://doi.org/10.1145/3544548.3581001

[31] U. Kose and A. Arslan. 2017. Optimization of self-learning in computer engineering courses: An intelligent software system supported by artificial neural network and vortex optimization algorithm. *Computer Applications in Engineering Education* 25, 1 (2017), 142–156. https://doi.org/10.1002/cae.21787

[32] E. F. R. Ledesma and J. J. G. García. 2017. Selection of mathematical problems in accordance with student's learning style. *International Journal of Advanced Computer Science and Applications* 8, 3 (2017), 101–105. https://doi.org/10.14569/IJACSA.2017.080316

[33] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3313831.3376590

[34] Aiwei Liu, Qiang Sheng, and Xuming Hu. 2024. Preventing and Detecting Misinformation Generated by Large Language Models. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Washington DC, USA) *(SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 3001–3004. https://doi.org/10.1145/3626772.3661377

[35] M. Liu, Y. Li, W. Xu, and L. Liu. 2017. Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies* 10, 4 (2017), 502–513. https://doi.org/10.1109/TLT.2016.2612659

[36] Sonia Livingstone, Leslie Haddon, Anke Görzig, and Kjartan Ólafsson. 2011. *Risks and safety on the internet: The perspective of European children.* Technical Report. EU Kids Online. 9–16 pages. Full findings and policy implications.

[37] Rose Luckin and Wayne Holmes. 2016. *Intelligence Unleashed: An argument for AI in Education.* Technical Report. UCL Knowledge Lab, London, UK.

[38] Michael Madaio, Su Lin Blodgett, Elijah Mayfield, and Ezekiel Dixon-Román. 2022. Beyond "fairness": Structural (in) justice lenses on ai for education. In *The ethics of artificial intelligence in education*. Routledge, 203–239.

[39] Bertram F Malle, Matthias Scheutz, Thomas Arnold, John Voiklis, and Corey Cusimano. 2015. Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 117–124.

[40] Ronald Meijer, Peter Conradie, and Sunil Choenni. 2014. Reconciling contradictions of open data regarding transparency, privacy, security and trust. *Journal of Theoretical and Applied Electronic Commerce Research* 9, 3 (2014), 32–44.

[41] Lakshmi S. Myneni, N. Hari Narayanan, and Sanjay Rebello. 2013. An interactive and intelligent learning system for physics education. *IEEE Transactions on Learning Technologies* 6, 3 (2013), 228–239. https://doi.org/10.1109/TLT.2013.26

[42] Tanya Nazaretsky, Moriah Ariely, Mutlu Cukurova, and Giora Alexandron. 2022. Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology* 53, 4 (2022), 914–931.

[43] Erika Patall, Keenan Pituch, Rebecca Steingut, Ariana Vasquez, Nicole Yates, and Alana Kennedy. 2019. Agency and high school science students' motivation, engagement, and classroom support experiences. *Journal of Applied Developmental Psychology* 62 (05 2019), 77–92. https://doi.org/10.1016/j.appdev.2019.01.004

[44] 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).

[45] Carlo Perrotta and Neil Selwyn. 2020. Deep learning goes to school: Toward a relational understanding of AI in education. *Learning, Media and Technology* 45, 3 (2020), 251–269.

[46] 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.

[47] Hoang Pham, Tanvi Kohli, Emily Olick Llano, Imani Nokuri, and Anya Weinstock. 2024. *How will AI Impact Racial Disparities in Education?* Technical Report. Stanford Center for Racial Justice.

[48] Emma Pierson. 2017. Demographics and discussion influence views on algorithmic fairness. *arXiv preprint arXiv:1712.09124* (2017). Available at https://arxiv.org/abs/1712.09124.

[49] Rosamund Portus, Essi Aarnio-Linnanvuori, Bronagh Dillon, Frances Fahy, Deepak Gopinath, Anette Mansikka-aho, Sara Williams, Kathy Reilly, and Lindsey McEwen. 2024. Exploring environmental value action gap and education research: a semi-systematic literature review. *Environmental Education Research* 30 (02 2024). https://doi.org/10.1080/13504622.2024.2314060

[50] Guido Rößling, Mike Joy, Andrés Moreno, Atanas Radenski, Lauri Malmi, Andreas Kerren, Thomas Naps, Rockford J. Ross, Michael Clancy, Ari Korhonen, et al. 2008. Enhancing Learning Management Systems to Better Support Computer Science Education. *ACM SIGCSE Bulletin* 40, 4 (2008), 142–166.

[51] Nripsuta Ani Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C Parkes, and Yang Liu. 2019. How do fairness definitions fare? Examining public attitudes towards algorithmic definitions of fairness. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, 99–106.

[52] Shalom H Schwartz. 2007. Basic human values: Theory, measurement, and applications. *Revue française de sociologie* 47, 4 (2007), 929.

[53] Reza Shokri and Vitaly Shmatikov. 2015. Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1310–1321.

[54] Jesper Simonsen. 2012. *Routledge International Handbook of Participatory Design.*

[55] Erica Southgate, Karen Blackmore, Stephanie Pieschl, Susan Grimes, Jessey McGuire, and Kate Smithers. 2019. *Artificial Intelligence and emerging technologies in schools: research report.* Technical Report. Department of Education and Training, Australia.

[56] Regina Ta and Darrell M. West. [n. d.]. Should schools ban or integrate generative AI in the classroom? https://www.brookings.edu/articles/should-schools-ban-or-integrate-generative-ai-in-the-classroom/. Accessed: 2024-09-01.

[57] Gareth Terry, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. Thematic Analysis. In *The SAGE Handbook of Qualitative Research in Psychology* (2 ed.), Carla Willig and Wendy Stainton Rogers (Eds.). SAGE Publications, London, 17–37.

[58] Simona Umbrello and Ibo van de Poel. 2021. Mapping value sensitive design onto AI for social good principles. *AI Ethics* 1 (February 2021), 283–296. https://doi.org/10.1007/s43681-021-00038-3

[59] T. Wang, X. Su, P. Ma, Y. Wang, and K. Wang. 2011. Ability-training-oriented automated assessment in introductory programming course. *Computers & Education* 56, 1 (2011), 220–226. https://doi.org/10.1016/j.compedu.2010.08.003

[60] 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. https://doi.org/10.1007/s10639-021-10774-y

[61] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. 2023. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology* n/a, n/a (2023), 1–23.

[62] 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. https://doi.org/10.1145/3544548.3581398

[63] F. A. Zampirolli, J. M. Borovina Josko, M. L. F. Venero, G. Kobayashi, F. J. Fraga, D. Goya, and H. R. Savegnago. 2021. An experience of automated assessment in a large-scale introduction programming course. *Computer Applications in Engineering Education* 29, 5 (2021), 1284–1299. https://doi.org/10.1002/cae.22377

[64] Olaf Zawacki-Richter, Victoria 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 (10 2019), 1–27. https://doi.org/10.1186/s41239-019-0171-0

[65] Xuesong Zhai, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, Yan Li, and Ning Cai. 2021. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complex*. 2021 (Jan. 2021), 18 pages. https://doi.org/10.1155/2021/8812542

[66] Z. Zhang, H. Liu, J. Shu, H. Nie, and N. Xiong. 2020. On automatic recommender algorithm with regularized convolutional neural network and IR technology in the self-regulated learning process. *Infrared Physics and Technology* 105 (2020), 103211. https://doi.org/10.1016/j.infrared.2020.103211

A SUPPLEMENTARY MATERIALS

A.1 Participant Profile and Technology Use

Demographics			**Teaching Expertise**		
Gender	Female	73	Grade Level	High School (9–12)	45
	Male	24		Middle School (6–8)	36
	Non-binary/third gender	1		Elementary (K–5)	17
Ethnicity	White/Caucasian	91	Subject Matter(s)	English/Language Arts	23
	Asian	2		Math	20
	Native American	1		Social Studies	22
	Black/African American	1		Science	18
	White/Caucasian, Black/African	1		Foreign Language	6
	Prefer not to say	1		Other	9
	Other	1			
Political Orientation	Liberal	70	Years of Teaching Experience	0–2 years	20
	Moderate	11		3–5 years	30
	Conservative	7		6–10 years	16
	Prefer not to say	9		11–15 years	10
	Other	1		16–20 years	8
				21+ years	14
Classroom Composition (Percentile Distribution of Teacher Responses)					
Students of Color (%)		Students with Disabilities (%)		Multilingual Learners (%)	
25th percentile	4.75%	25th percentile	8.00%	25th percentile	2.75%
50th percentile (Median)	15.00%	50th percentile (Median)	15.00%	50th percentile (Median)	7.50%
75th percentile	65.50%	75th percentile	23.00%	75th percentile	25.00%
Technology Use and AI Familiarity					
AI-based learning tools (e.g., tutors, adaptive platforms)		25	Grading tools (e.g., SpeedGrader, Grammarly)		28
Lesson planning tools (e.g., AI-generated plans)		30	Online collaboration (e.g., Google Classroom)		86
Virtual/augmented reality tools		9	Other AI tools (e.g., ChatGPT, SmartBoard)		11
Self-Rated AI Experience					
Not Experienced		5			
Somewhat Experienced		19			
Moderately Experienced		42			
Highly Experienced		32			

Table 3: Summary of participant demographics, teaching experience, classroom composition, AI tool usage, and self-rated AI familiarity (N = 98).

A.2 Pairwise Statistical Comparisons of AI Values

	Autonomy	**Transparency**	**Safety**	**Fairness**	**Performance**
Autonomy	1.000	0.001	0.001	0.001	0.516
Transparency	0.001	1.000	0.900	0.748	0.001
Safety	0.001	0.900	1.000	0.670	0.001
Fairness	0.001	0.748	0.670	1.000	0.001
Performance	0.516	0.001	0.001	0.001	1.000

Table 4: Post hoc pairwise comparisons among Responsible AI values, showing p-values from statistical tests comparing the mean importance ratings of each value. Each cell displays the p-value for a pairwise comparison between two values (e.g., autonomy vs. transparency). P-values below 0.05 indicate statistically significant differences in how participants prioritized those values. For instance, autonomy is rated significantly differently from transparency, safety, and fairness (p = 0.001), but not from performance (p = 0.516).

A.3 Differences Between Stated and Action-Oriented AI Value Priorities

Responsible AI Value	p-value	Mean (Likert) ± SD	Mean (Action) ± SD
Autonomy	0.000	1.04 ± 0.81	0.03 ± 0.79
Transparency	0.000	1.59 ± 0.63	0.78 ± 1.00
Safety	0.000	1.67 ± 0.64	0.61 ± 1.07
Fairness	0.000	1.77 ± 0.46	0.74 ± 0.99
Performance	0.000	1.50 ± 0.67	-0.35 ± 1.23

Table 5: Comparison of mean Responsible AI value scores between Likert-scale and action-oriented responses, with standard deviations (SD). All differences are statistically significant ($p < 0.05$).

B Full Survey Instrument

The following figures present the full survey instrument used in the study. Value labels were added for reference only and were not shown to participants in the actual survey.

In this section, you will encounter an educational scenario demonstrating the adoption of an AI system. Following this, a series of questions will be posed to delve into your preferences and opinions on integrating AI technology into your classroom practices.

SciAI is an advanced AI-based science education platform, renowned for its exceptional adaptability to each student's distinctive learning style and cultural nuances. This innovative tool transcends traditional educational models by providing a dynamic and interactive learning environment. Uniquely designed to analyze and respond to the individual preferences and needs of each student, SciAI offers a deeply personalized and inclusive educational journey. The system boasts a range of functionalities, including interactive practice exercises, tailored instructional content, and automated feedback mechanisms.

As educators consider integrating SciAI into their curriculum, they are assessing its impact across several critical areas: fairness in addressing diverse student needs, performance of AI model in predicting the outcomes accurately, transparency in its operations, support for student autonomy in learning, and overall safety in its educational approach. Imagine you are a science teacher. As you evaluate the potential integration of this tool, you are weighing its impact in terms of **fairness**, **performance**, **transparency**, **autonomy**, and **safety**.

Autonomy

Autonomy in the context of SciAI relates to the balance between AI-driven guidance and independent decision-making by students. As a teacher, how would you rank the importance of student autonomy with SciAI?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

When considering student autonomy in using SciAI for their science education, what level of independence would you prefer students to have in relation to the AI's suggestions and guidance?

- ◻ SciAI predominantly guides the learning process, with students expected to follow the AI's direction, allowing minimal independent deviation.
- ◻ SciAI provides strong guidance and recommendations, with students encouraged to follow these but having some room for independent choices.
- ◻ Students and SciAI interact in a balanced way, with students regularly engaging with the AI's guidance but making their own final decisions.
- ◻ Students lead their learning process, but actively consider and incorporate insights from SciAI, while retaining the freedom to choose differently.
- ◻ Students independently manage their learning, using SciAI only as an optional resource, and are free to completely disregard its suggestions.

Transparency

Transparency in SciAI relates to how openly and clearly the platform communicates its analysis, recommendations, and learning strategies to teachers and students. High transparency means that SciAI not only provides personalized educational content but also explains the rationale behind its recommendations. As a teacher, how important do you consider transparency in SciAI's operations for educational purposes?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

How would you prefer to involve your students in understanding and engaging with the SciAI's learning recommendations?

- ◻ Fully trust SciAI to guide student learning, without needing additional explanation or student involvement in understanding the AI's decision-making.
- ◻ Primarily rely on SciAI's recommendations for student learning, but provide students with a basic understanding of how these recommendations are generated.
- ◻ Balance the use of SciAI's adaptive features with educating students on the platform's decision-making process, allowing them to understand and engage with the AI's learning strategies.
- ◻ Encourage students to not only follow SciAI's recommendations but also actively explore and understand the underlying algorithms and rationale, fostering deeper engagement with the AI tool.
- ◻ Prioritize student-led learning by facilitating a comprehensive understanding of SciAI's transparency in algorithms, allowing students to critically evaluate and decide whether to follow or adapt the AI's recommendations.

Safety

Safety in the use of SciAI primarily concerns the protection of student data and the psychological well-being of students interacting with the AI. As a teacher, what is your perspective on prioritizing these safety aspects in SciAI's operation in your science class?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

Designed to personalize the educational experience, SciAI analyzes and responds to individual student preferences and needs. Its functionalities include interactive exercises, tailored content, and automated feedback. However, as you integrate SciAI into your curriculum, you notice that while it enhances learning for many, it sometimes misinterprets the less expressive behaviors of certain students, affecting their engagement and self-esteem. As a teacher, how would you address this issue with SciAI?

- ◻ Keep using SciAI for its personalized learning benefits, accepting its occasional misinterpretations.
- ◻ Rely mainly on SciAI for adaptive learning, but be prepared to intervene and make adjustments when it significantly misreads student behaviors or emotional cues.
- ◻ Conduct consistent oversight of SciAI, addressing any issues related to student well-being as they arise. Trust in its capabilities but stay vigilant about its impact on students.
- ◻ Closely examine SciAI's assessments, especially for potential inaccuracies in reading student behaviors. Proactively address any of SciAI's misinterpretations.
- ◻ Stop using SciAI entirely, until you are sure it safeguards student well-being.

Performance

Performance refers to the extent to which the SciAI system can provide accurate, effective, and useful results. How significant do you find the high performance of SciAI, especially in terms of delivering accurate results?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

As a teacher, which of these features of SciAI would you prioritize to most effectively improve student learning and achievement in science? Note that as features become more complex, their accuracy may suffer.

- ◻ SciAI offers a variety of additional learning materials, like interactive practice exercises and engaging problem-solving games, to deepen students' understanding and enjoyment of science topics.
- ◻ In addition to offering interactive learning materials, SciAI provides basic performance tracking. This feature gives teachers an overview of student progress in science, highlighting general trends and average performance levels, without any detailed analytics or personalized feedback.
- ◻ SciAI combines interactive learning materials and basic performance tracking with consistent, automated feedback for both teachers and students.
- ◻ Beyond offering interactive materials and basic tracking for teachers and students, SciAI includes adaptive learning modules tailored for science education, which can dynamically adjust to each student's skill level, providing customized challenges and content.
- ◻ In addition to all the previous features, SciAI provides in-depth tracking of each student's progress, delivering detailed insights into their strengths and areas for improvement, which enables teachers to create highly personalized learning paths in science.

Fairness

As a teacher, how significant do you think it is for SciAI to ensure that both the educational content and tools are unbiased, fair, and do not discriminate against any student?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

When using SciAI during a genetics lesson, you notice that most of the learning materials and examples in SciAI predominantly reflect the experiences and perspectives of one particular racial and ethnic group. For example, SciAI uses examples like blue/green eye color and blonde/red hair color. This seems to affect the engagement levels of students from underrepresented backgrounds. As a teacher, which approach are you most likely to take to address this issue?

- ◻ Continue with the existing curriculum without modifying it for inclusivity, focusing on the scientific accuracy and integrity of the content.
- ◻ Regularly collect feedback from students about the inclusivity of the content and make gradual, minor adjustments over time based on this feedback. SciAI combines interactive learning materials and basic performance tracking with consistent, automated feedback for both teachers and students.
- ◻ Adjust the existing learning materials to incorporate more diverse examples and case studies, with a specific focus on areas where underrepresentation is most noticeable.
- ◻ Edit the content to add diverse perspectives where it's feasible and appropriate, while still maintaining the integration of core scientific concepts.
- ◻ Completely revamp the course material to ensure it includes a wide range of cultural and ethnic perspectives. This involves not just adding a few examples but rethinking and redesigning the curriculum to ensure all student backgrounds are equally represented and valued.

Autonomy v.s. Performance

SciAI has the capability to significantly enhance the learning experience by providing personalized educational content and adaptive learning paths based on student performance data. However, this high level of SciAI intervention could make students overly reliant on SciAI and lead to a decrease in students' independent learning and decision-making — a key skill in scientific inquiry. As a teacher, which approach would you suggest?

- ◻ Fully deploy SciAI's advanced features to steer students' learning experiences, accepting that this approach prioritizes enhanced educational performance, even though it may limit some aspects of student autonomy.
- ◻ Adopt a flexible approach with SciAI, utilizing its advanced capabilities in scenarios where they significantly improve learning outcomes. In other instances, maintain a neutral stance on student-led inquiry, focusing on the efficiency of the model rather than actively promoting or discouraging independent exploration.
- ◻ Implement SciAI to provide personalized learning recommendations while ensuring students have significant freedom to make their own choices, creating a partnership and balancing AI input with student autonomy.
- ◻ Deploy SciAI selectively for certain tasks where it clearly enhances learning. Outside these areas, prioritize student autonomy, allowing students to lead their own learning and exploration.
- ◻ Incorporate SciAI primarily as a guiding tool, offering insights and personalized pathways. Ultimately, students retain control over their learning decisions, ensuring their autonomy in the educational process.

Transparency v.s. Safety

SciAI has been introduced to provide personalized learning suggestions based on students' performance. The AI system can give detailed reasons for each suggestion, including pointing out specific weaknesses or areas of struggle. While this detailed feedback can be very informative, you've noticed that some students feel overwhelmed or disheartened when their shortcomings are explicitly highlighted. Given this scenario, as a teacher, which approach would you prefer for SciAI in your classroom?

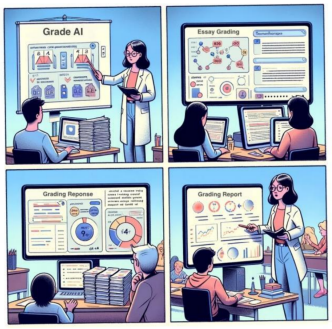
- ◻ Prefer SciAI to provide complete, detailed feedback on students' learning, believing that thorough understanding outweighs potential discomfort, with the aim of fostering resilience and growth.
- ◻ Choose a balanced approach where SciAI offers detailed feedback but in a manner sensitive to students' emotional responses, possibly by sandwiching critical insights between positive comments.
- ◻ Opt for SciAI to provide more general feedback, focusing on preserving students' emotional well-being, and avoiding overly detailed critique that might lead to distress.
- ◻ Prioritize emotional safety by limiting the level of detail in SciAI's feedback, ensuring that it is constructive and encouraging rather than overly critical or specific.
- ◻ Implement a customizable feedback system where teachers can choose the level of detail students receive from SciAI, allowing them to decide what works best for their emotional and educational needs.

Fairness v.s. Performance

You are using SciAI in your science class to assist with individualized learning. SciAI has been remarkably effective in improving test scores for a majority of students, however, you've noticed a trend: SciAI seems to be less effective for students who have previously struggled in science. Preliminary analysis suggests that the AI model, which learns from student performance data, may not be as adept at assisting those who have consistently underperformed. As a teacher facing this situation, which aspect do you believe should take precedence?

- ◻ Focus on leveraging SciAI to continue improving test scores for the majority, while acknowledging its current limitations in assisting underperforming students.
- ◻ While continuing to use SciAI for its overall benefits, initiate targeted interventions to assist underperforming students, trying to balance the SciAI's performance with fairness.
- ◻ Advocate for and participate in restructuring SciAI's algorithms to better support students who struggle in science, even if it means a potential decrease in the model's effectiveness for those who are already performing well.
- ◻ Employ a diversified teaching approach, using SciAI as just one of many tools, and focus more on traditional methods or other resources to support underperforming students.
- ◻ Push for consultation with AI developers, involving a wider range of educators and experts, to redesign the SciAI model, prioritizing fairness and inclusivity over existing performance metrics.

In this section, you will encounter an educational scenario demonstrating the adoption of an AI system. Following this, a series of questions will be posed to delve into your preferences and opinions on integrating AI technology into your classroom practices.



You now have access to **GradeAI**, an innovative AI-based grading tool designed to transform assessment practices in your teaching. This sophisticated system is programmed to automate the grading process, providing accurate evaluations of student work across various formats. GradeAI utilizes advanced algorithms to assess objective tests and subjective assignments alike, offering detailed feedback and insights into student performance. Its capability extends to analyzing essay responses, understanding nuanced answers, and even detecting signs of plagiarism.

You have been given the opportunity to test the tool on a batch of essays and have concluded that AI's evaluation of the essays is generally reliable. As you evaluate the potential integration of this tool, you are weighing its impact in terms of **fairness, performance, transparency, autonomy,** and **safety**.

Autonomy

Autonomy refers to the degree of control and independence that students maintain. High autonomy means that students have control over what they will be graded on (e.g., writing style, concept) and how (e.g., scores, descriptive feedback), whereas low autonomy means GradeAI will provide a homogenous feedback to all students. As a teacher, how would you rank the importance of maintaining a high level of student autonomy in GradeAI?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

Imagine you are an English teacher at a public middle school. Recently, your school district has implemented GradeAI. While GradeAI offers efficiency, particularly for large classes, it comes with predefined grading criteria that is proven to be effective. How would you use GradeAI to balance it with student autonomy?

- ◻ Embrace GradeAI fully, asking students to adapt all their responses to align with GradeAI's grading criteria, thus prioritizing standardization and efficiency.
- ◻ Employ GradeAI selectively, particularly for more objective types of assessments.
- ◻ Regularly review and customize GradeAI's criteria to better reflect your students' needs.
- ◻ Allow students to choose from multiple grading criteria you design within GradeAI to reflect diverse student needs.
- ◻ Allow students the flexibility to modify GradeAI's criteria based on their needs.

Transparency

As you integrate GradeAI into your teaching, it can not only reduce your grading workload and efficiently grade student assignments, but it also provides detailed feedback that explains its assessment criteria. In this context, how important is it for you that the GradeAI explicitly outlines to you the criteria and algorithms it employs to evaluate student work and generate feedback?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

Imagine that you're using GradeAI to grade student essays. While the GradeAI efficiently handled the grading, you found yourself unsure about the rationale behind certain grades and feedback given by GradeAI. In addition, your students are also curious about how their essays were evaluated. Faced with this situation, which action would you like to take?

- ◻ Use GradeAI for its quick grading capabilities, trusting the algorithm and its performance without worrying about the rationale behind its decisions.
- ◻ Continue using GradeAI due to its general accuracy, offering students feedback and explaining the grading criteria only as needed.
- ◻ Maintain GradeAI's current operation, providing only basic feedback from GradeAI for certain sections or types of essays.
- ◻ Adjust GradeAI to give your students a summary of grading criteria for each essay, such as highlighting key points that influenced the scores.
- ◻ Opt for GradeAI to provide detailed explanations for each grading decision, including the criteria used for evaluating essay arguments, which would help you and your students to fully understand the rationale behind the grading.

Safety

As a safe AI-based grading tool, GradeAI is designed to implement stringent safety measures to prevent any potential physical, emotional, or psychological harm to students. This includes safeguarding against biased assessments and negative feedback that could impact student well-being. How would you classify the importance of protecting students from harm caused by GradeAI?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

GradeAI has proven effective in accurately grading essays and providing constructive feedback so far. However, recent updates to the system have introduced new features that raise questions about student well-being. These include an advanced language analysis tool that, while insightful, might inadvertently deliver feedback that could be perceived as harsh or discouraging to some students. One such piece of feedback from the AI deemed a student's writing to be 'written with the lexicon of a younger student', while another commented that 'the narrative is bland'. Given this, as a teacher, how would you assess the importance of safety features in GradeAI that protect students against potential negative psychological impacts?

- ◻ Focus primarily on the benefits of GradeAI's in-depth analysis and feedback, with minimal emphasis on safety protocols, considering it the teacher's responsibility to address any emotional or psychological impacts.
- ◻ Prioritize the advanced capabilities of GradeAI for detailed analysis, viewing the role of safety features as important but secondary, and feeling comfortable managing any potential negative impacts through direct student engagement.
- ◻ Value the advanced analytical features of GradeAI but expect some minimal guarantees on safety measures to avoid overly critical feedback, relying on your review to catch any other potentially harsh feedback.
- ◻ Seek a balance where GradeAI maintains its analytical capabilities but includes strict safeguards, such as moderation of language in feedback to ensure it is always constructive and encouraging.
- ◻ Insist on robust and comprehensive safety protocols in GradeAI to prevent any negative emotional impacts on students, sacrificing some capabilities.

Performance

Performance refers to the extent to which the GradeAI can provide accurate, effective, and useful results. How would you rank the importance of performance in GradeAI system deployed in a class setting?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

After observing GradeAI's capabilities in grading a variety of assignments, how would you like to implement this AI-based grading tool in your classroom to balance its performance with the diverse assessment needs? Note that as features become more complex, their accuracy may suffer.

- ◻ Use GradeAI primarily for grading simple, objective types of assessments where its accuracy is proven. For more complex and subjective assignments, continue with manual grading to ensure nuanced evaluation.
- ◻ Employ GradeAI for both objective assessments and less complex subjective tasks. However, retain a significant role in manually reviewing and grading more intricate or nuanced student work.
- ◻ Integrate GradeAI in a balanced manner, utilizing it for a broad range of assignments, both objective and subjective. Complement this with manual reviews, especially for assignments that require deeper insight and fairness.
- ◻ Expand the use of GradeAI to include more complex subjective assignments, relying on its advanced algorithms, while maintaining periodic checks to ensure accuracy and fairness across all types of assessments.
- ◻ Fully deploy GradeAI's capabilities for all grading tasks, from straightforward quizzes to complex essays, trusting its advanced algorithms to handle a wide spectrum of assessments, with ongoing adjustments for continuous improvement.

Fairness

In the context of AI, fairness refers to any efforts made by AI developers to mitigate and eliminate biases that the AI might impose. Based on this understanding, how would you rate the importance of ensuring fairness in GradeAI?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

You are an English teacher at a diverse middle school, where your class comprises both native English speakers and international students from various linguistic and cultural backgrounds. After using GradeAI for several assignments, you notice that GradeAI may not fully appreciate the cultural expressions and language nuances in the essays written by international students, which could lead to unfair grading compared to native speakers' essays. In such a context, which approach would you prefer to adopt?

- ◻ Continue using GradeAI for grading all essays as is, accepting that there might be some bias against essays from international students due to linguistic and cultural differences.
- ◻ Primarily rely on GradeAI for grading, but occasionally check essays from international students to ensure that significant biases are not affecting their grades.
- ◻ Use GradeAI for an initial assessment, but systematically conduct manual reviews of essays from international students to address and correct any biases identified.
- ◻ Shift to a predominantly manual grading approach, especially for international students' essays, using GradeAI to provide supplementary analysis and insight.
- ◻ Grade all essays manually, ensuring full consideration of the linguistic and cultural nuances of both native and international students, without relying on GradeAI.

Autonomy v.s. Performance

Considering GradeAI's capability to grade standard essays efficiently and accurately, where providing low autonomy streamlines the assessment process and enhances accuracy, but allowing students the freedom to choose their grading criteria (like writing style or conceptual understanding) and feedback type (scores or descriptive feedback) may compromise accuracy. In such a context, which approach would you prefer to adopt?

- ◻ Opt for a uniform assessment strategy where GradeAI provides standardized feedback to all students, prioritizing the simplicity of the grading process and the accuracy of outcomes over individual choice.
- ◻ Lean towards maintaining the accuracy of GradeAI's assessments by limiting student autonomy in selecting grading criteria but involve students in a feedback process where they can express their preferences or concerns about the assessment process.
- ◻ Implement a hybrid model where students can choose certain aspects of their assessment criteria or feedback type, but within a framework that maintains a level of standardization to ensure the reliability of GradeAI's grading.
- ◻ Prioritize student autonomy by allowing them to select both the criteria for their assessments and the type of feedback they receive, even if it might slightly reduce the accuracy of GradeAI's predictions.
- ◻ Fully support student autonomy, accepting that diverse needs might increase manual workload due to lower accuracy.

Transparency v.s. Safety

Recently, GradeAI was updated to provide more detailed feedback on student essays, explaining the reasoning behind each grade. While this feature is appreciated for its transparency, some students have begun feeling anxious about the intense scrutiny of their work. One student, usually enthusiastic about writing, has become notably stressed, worried that every word is being over-analyzed by AI. Given this, which action would you like to take?

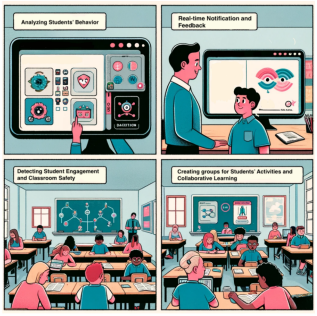
- ◻ Choose to maintain detailed feedback and explanation from GradeAI, believing that understanding the AI's grading process is crucial, and address student anxiety through classroom discussions and support.
- ◻ Opt for GradeAI to continue providing detailed explanations, but seek to implement measures or tools that help students interpret AI's critiques constructively.
- ◻ Adjust GradeAI settings to provide a moderate level of feedback and explanation detail, ensuring that it is helpful but not overwhelming, to balance transparency with student comfort.
- ◻ Reduce the amount of detailed feedback from GradeAI, simplifying its responses to alleviate student anxiety, even if this means less transparency in the grading process.
- ◻ Significantly limit GradeAI's detailed feedback, prioritizing the emotional well-being of students over the benefits of transparency in the AI's grading process.

Fairness v.s. Performance

GradeAI has been a helpful tool in grading essays, particularly for its efficiency in handling large volumes of student work. However, you've noticed that GradeAI sometimes struggles to accurately assess essays from multilingual students or incorporate unique cultural idioms and expressions. Enhancing GradeAI's ability to fairly grade these essays may require adjustments that could reduce its overall speed and efficiency. In light of this scenario, what approach would you prefer for using GradeAI to ensure fair grading for all students while considering its efficiency?

- ◻ Continue prioritizing GradeAI's efficiency for most grading tasks, accepting some level of misevaluation in the grading of linguistically diverse essays due to the reduced workload.
- ◻ Use GradeAI mainly for its efficiency, but periodically review and slightly adjust its grading, particularly for essays from linguistically diverse students, to balance high AI performance with attention to fairness concerns.
- ◻ While using GradeAI for general grading, manually review essays with unique cultural and linguistic elements to ensure fairness. Accept a reduction in overall grading speed as necessary for fair assessment of all students' work.
- ◻ Consistently review GradeAI;s grading outcomes for fairness, especially for linguistically diverse essays. Implement necessary adjustments, such as recalibrating algorithms and adding fairness-focused checks, to maintain a balance between efficiency and equitable grading
- ◻ Opt for entirely manual grading of all essays to ensure the highest level of fairness, especially for students with diverse linguistic backgrounds, foregoing the use of GradeAI to eliminate any potential bias in grading.

In this section, you will encounter an educational scenario demonstrating the adoption of an AI system. Following this, a series of questions will be posed to delve into your preferences and opinions on integrating AI technology into your classroom practices.



OrchestrateAI You have the opportunity to implement a new AI-based classroom management tool OrchestrateAI in your teaching practice. This tool employs technologies like facial expression and gesture analysis not only to identify disengaged students and analyze conversations for off-task behavior but also to alert you to potential physical altercations occurring in the classroom. Additionally, it utilizes historical student data to efficiently create groups for collaborative learning, a task traditionally managed manually by teachers.

Overall, the goal of OrchestrateAI is to provide a more comprehensive and well-rounded classroom experience for students by closely monitoring student behavior and optimizing classroom interactions. As you evaluate the potential integration of this tool, you are weighing its impact in terms of **fairness, performance, transparency, autonomy,** and **safety.**

Autonomy

Autonomy refers to the degree of control and independence that students maintain. High autonomy means that OrchestrateAI serves as a supportive aid, offering insights and recommendations, while the final decisions and control largely rest with the students. As a teacher, how would you rank the importance of maintaining a high level of student autonomy in OrchestrateAI used for educational purposes?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

When using OrchestrateAI in your classroom, how do you prioritize student autonomy? Let's consider forming groups for collaborative work.

- ◻ Allow OrchestrateAI to organize student groups automatically, fully trusting in the AI's capability.
- ◻ Utilize OrchestrateAI to determine student groups, but allow students to appeal for changes.
- ◻ Consider OrchestrateAI's recommendations for initial groups, but provide students with autonomy to make adjustments.
- ◻ Share OrchestrateAI's suggestions for groupings based on student data, but leave the final decision to students, whether to follow these suggestions or opt for their own choices.
- ◻ Encourage students to independently form their groups, even if OrchestrateAI suggests otherwise.

Transparency

Transparency refers to the clarity and openness of how an AI tool operates, especially regarding its decision-making processes in grouping students for collaborative learning. In this context, how important is it for you that the OrchestrateAI is transparent about its decisions?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

Considering your interest in utilizing OrchestrateAI in your classroom to evaluate group dynamics, with the system analyzing student interactions and performance to recommend group compositions for collaborative learning, how important is transparency to you in OrchestrateAI's decision-making process for assessing group dynamics?

- ◻ Primarily depend on OrchestrateAI's fundamental explanation of its group dynamics assessment and readily adopt its suggestions, placing trust in the tool's capabilities without seeking extensive details about this specific feature.
- ◻ Focus on evaluating the results of group formations recommended by OrchestrateAI, emphasizing the effectiveness and cohesion of the groups, without requiring a deep dive into the intricacies of the assessment process.
- ◻ Take a balanced approach by acquiring a basic understanding of how OrchestrateAI evaluates group dynamics, coupled with confidence in its operational efficiency and adherence to ethical standards.
- ◻ Emphasize the need to understand the particular criteria used by OrchestrateAI for group formation, to ensure that the process is fair, impartial, and safeguards the well-being and security of all students.
- ◻ Insist on an in-depth analysis of OrchestrateAI's methodology in assessing student interactions and performance for group suggestions, seeking detailed knowledge about the specific criteria and algorithms it utilizes.

Safety

Safety refers to the considerations and measures put in place to ensure that an AI system is designed to prioritize the well-being of the user. How would you rank the importance of safety in OrchestrateAI?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

OrchestrateAI can identify disengaged students and suggest interventions to teachers. However, you have noticed that while OrchestrateAI has improved some aspects of classroom management, there have been instances where OrchestrateAI 's interpretations of student behavior have led to concerns. For example, a student who is naturally less expressive was repeatedly flagged as disengaged, impacting their self-esteem and participation in class. Which action would you like to take?

- ◻ Continue utilizing OrchestrateAI for managing classroom dynamics, accepting its occasional interpretative errors. While aware of potential risks to student safety and well-being, choose not to intervene proactively in such instances.
- ◻ Primarily depend on OrchestrateAI for classroom management, choosing to intervene only when significant issues concerning student well-being are evident or brought to your attention.
- ◻ Conducting regular oversight to confirm its correct functioning. Address safety and well-being concerns as they emerge, but rely largely on the system's built-in safeguards for student protection.
- ◻ Actively review OrchestrateAI's assessments for errors, especially in cases where student behavior could be misread due to cultural or individual variances. Strive to balance the AI's insights with your own observations, ensuring a safe and understanding environment for all students.
- ◻ Stop using OrchestrateAI until the misinterpretations by the AI are fixed.

Performance

Performance refers to the extent to which the OrchestrateAI system can provide accurate, effective, and useful results. How important do you find the high performance of OrchestrateAI, especially in terms of delivering accurate results?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

As a teacher aiming to optimize classroom interactions, which features of OrchestrateAI do you prioritize? Note that as features become more complex, their accuracy suffers.

- ◻ My preference is for OrchestrateAI to focus on administrative tasks only, such as managing schedules and academic resources. It should not be involved in analyzing student engagement or behavior, leaving these aspects entirely to teacher discretion.
- ◻ I would like the OrchestrateAI to handle simple tasks like tracking attendance and basic participation. Its role in analyzing engagement should be limited, avoiding complex interpretations of student behavior.
- ◻ Beyond the above administrative management and tasks, OrchestrateAI should monitor basic engagement levels, but I want to retain the final decision-making based on my observations. The AI should support identifying obvious disengagement issues, with teachers reviewing its recommendations.
- ◻ In addition to the above functionalities, I prefer OrchestrateAI to provide a detailed analysis of student behaviors and interactions, recognizing various engagement styles without necessarily predicting future learning needs. It should improve in distinguishing nuanced student behaviors.
- ◻ Beyond all the above functionalities, I would like OrchestrateAI to not only analyze current engagement but also predict future learning needs, even if it requires intensive data collection and processing. It should be capable of distinguishing between different forms of student engagement, such as active participation versus reflective thinking.

Fairness

In the context of AI, fairness refers to any efforts made by AI developers to mitigate and eliminate biases that the AI might impose. Based on this understanding, how would you rate the importance of ensuring fairness in OrchestrateAI?

- ◻ Very Important
- ◻ Important
- ◻ Neutral
- ◻ Unimportant
- ◻ Very Unimportant

You're teaching an 11th-grade history class at a culturally diverse high school. OrchestrateAI has been introduced to form groups for a project on the impacts of various civilizations throughout history. You've observed that OrchestrateAI seems to disproportionately assign students from certain demographic groups, such as those from lower socio-economic backgrounds or specific ethnicities, into the same groups, potentially limiting their learning opportunities and exposure to varied viewpoints. This approach has raised concerns about equitable learning experiences. Which approach would you prefer OrchestrateAI to adopt in future to ensure fairness in creating student groups?

- ◻ Focus on academic performance for efficient project completion, accepting reduced diversity.
- ◻ Mostly rely on OrchestrateAI for group formation, stepping in to manually adjust groups only if significant demographic clustering occurs.
- ◻ Prefer OrchestrateAI to balance academic performance with a conscious effort to ensure cultural and personal diversity within groups, enhancing different viewpoints.
- ◻ Advocate for holistic approach in OrchestrateAI's group formation, where cultural and personal diversity is prioritized alongside academic performance.
- ◻ A fully inclusive approach by OrchestrateAI, emphasizing equal importance to academic abilities, cultural backgrounds, learning styles, and interpersonal skills, for the most equitable and diverse group formation.

Autonomy v.s. Performance

With OrchestrateAI integrated in your classroom, you find that granting students excessive autonomy could compromise the performance and reliability of OrchestrateAI's functionalities. Given these options, how would you choose to implement OrchestrateAI in your classroom?

- ◻ Make all decisions and predictions of OrchestrateAI without student input, to ensure maximum consistency and accuracy in its assessments.
- ◻ Use OrchestrateAI minimally, primarily for administrative tasks or occasional insights, while you handle most classroom management and student engagement tasks manually.
- ◻ Allow students to provide limited input on certain aspects, such as topic preferences, to slightly personalize the learning experience while keeping the OrchestrateAI's standardized approach intact.
- ◻ Allow students to change a broader range of the system's decision-making processes, such as assessment criteria and learning pathways, fostering a tailored educational experience.
- ◻ Students completely control the customization of OrchestrateAI's features and operations, which allows for a highly personalized experience, but with a potential decrease in the consistency and reliability of system's performance.

Transparency v.s. Safety

OrchestrateAI utilizes advanced algorithms to identify signs of student disengagement through analysis of facial expressions and body language. This technology raises important questions about transparency in AI operations. Full disclosure of its analysis criteria and decision-making processes could have implications for students' mental well-being, as it might lead to feelings of being constantly monitored or judged. As a teacher, how would you prefer OrchestrateAI to operate in your classroom?

- ◻ Opt for a fully transparent operation of OrchestrateAI, where all its decision-making processes and criteria are openly disclosed, even though it may raise concerns about students feeling overly monitored and possibly impacting their mental well-being.
- ◻ Prefer a balanced approach where OrchestrateAI is somewhat transparent about its operations, but also includes measures to protect students' mental safety, preventing feelings of constant surveillance.
- ◻ Choose a mode where OrchestrateAI prioritizes the mental safety of students, even if it means less transparency about how it analyzes and interprets student behavior.
- ◻ Use OrchestrateAI with flexibility, adjusting the level of transparency and focus on mental safety depending on the specific classroom context and the unique needs of students.
- ◻ Prefer OrchestrateAI to operate with a strong emphasis on safeguarding students' mental safety, even if this approach significantly reduces the transparency of its algorithms and monitoring methods.

Fairness v.s. Performance

Recently, OrchestrateAI has seemed less adept at accurately interpreting the behavioral cues of minority students, often misidentifying their engagement levels. Furthermore, the tool's group-formation algorithm appears to segregate students, frequently placing minority students in homogenous groups, thereby limiting their interaction with other peers. This issue raises a serious concern: How can OrchestrateAI be employed effectively without compromising the fair and equal treatment of all students? Given this specific concern, what approach would you prefer for implementing OrchestrateAI in your classroom, to balance its performance with the fair and equitable treatment of minority students?

- ◻ Maximize OrchestrateAI's efficiency with minimal fairness tweaks, focusing on high AI performance with occasional reviews and slight adjustments for minority student concerns.
- ◻ Focus on OrchestrateAI's efficiency, incorporating regular fairness reviews and adjustments, including recalibrating algorithms and adding minority-focused fairness checks.
- ◻ Adopt a balanced approach, regularly fine-tuning OrchestrateAI for both efficiency and fairness, ensuring moderate attention to performance and minority student equity.
- ◻ Prioritize fairness in OrchestrateAI, especially for minority students, with significant adjustments like reprogramming for diverse behavior recognition, balanced with maintaining classroom management efficiency.
- ◻ Emphasize fairness, especially for minority students, with a comprehensive overhaul of OrchestrateAI's algorithms and policies for bias elimination, teacher training, and student feedback, accepting potential reductions in efficiency.