Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

Teachers' Lived Experiences with Explainable AI in Classroom Decision-Making

Dr. Nazish Andleeb
Lecturer, Department of Education, University of Gujrat. nazish.andleeb@uog.edu.pk

Dr. Muhammad Ahsan Mukhtar*
Lecturer, Department of Commerce, University of Gujrat. Corresponding Author Email: Ahsan.mukhtar@uog.edu.pk

Abstract
Since the introduction of artificial intelligence (AI) technologies into educational systems is more of a trend, explainable AI (XAI) systems are becoming a solution to assist teachers in making decisions in areas like grading, student interaction, and learning diagnostics. This paper will discuss the experiences of teachers who are secondary and post-secondary and interacted with systems that are enabled by XAI-enabled systems in their classrooms. Set in a phenomenon qualitative design, the study will rely on detailed solicitation of the teachers (n=12) in both state and private schools. Theymatic coding with NVivo was used to analyze the data and both the technical familiarity and pedagogical impact were taken into consideration in the study. The next assumptions made were the increasing role of algorithmic decision-making in classroom and the fact that AI output needs to be interpretable so as not to relegate teacher agency. It presents emotional, cognitive, and pedagogical tensions that teachers feel working with AI systems that affect teaching compared to other studies that concentrate on student data analytics. The paper concludes that developing teacher confidence and competence in XAI needs professional development and a participatory approach to AI design to achieve this. The longitudinal outcomes of XAI should be examined in future as a way of affecting teaching autonomy and student outcomes.

Introduction
Artificial intelligence (AI) has an incredibly fast changing nature on the education scenes, and schools and universities are implementing AI-based solutions to the outside world to offer personalized education, testing, and prediction of student performance. Explainable AI (XAI) is one of these innovations that has become an essential subdiscipline whose goal is to ensure that the decisions made by machines can be transparent and comprehensible to humans. XAI systems are undergoing widespread adoption in the education sector in an attempt to guide teachers in their grading-, at-risk-students-, and instruction-related decisions. XAI has a promising future because it can enhance the judgment of teachers instead of eliminating it to implement more effective and timely interventions (Holstein et al., 2019). Nevertheless, even technical accuracy of such tools does not suffice on the success, but the trust, interpretability, and usability of the teacher.

Although much research has been conducted regarding AI in education in terms of seeing this in terms of systems or a learner, very limited studies focus on the lived experience of the teacher especially how they interrelate with XAI

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

systems, which do impact pedagogical decisions. It is assumed that AI can become an easy thing to use in case systems are accurate and reliable. Nevertheless, this does not take into account the emotional, ethical, and cognitive issues that teachers experience in case their professional discretion is mediated by algorithmic prescriptions. In addition, the majority of literature in the field of AI in education is based on student analytics dashboard or administrative decision-making, rather than on the way in which teachers themselves perceive and can perform AI-related feedback in the context of real classroom settings.

The proposed research paper will shed light on the perception, interpretation, and reaction of teachers to explainable AI tools in classroom decision-making. The study has used a phenomenological methodology to develop a picture of what the life of teachers in XAI-enabled systems in secondary and tertiary educational institutions is like. The results can provide new information about the interaction between protocols and their influence on human judgment and algorithmic instructions, presenting opportunities and certain tensions of AI at a classroom level and other scales. The paper is structured in the following way: the literature review about XAI in education, description of the methodology including the recruitment of participants and the analysis of the data, and the introduction of the key themes, the critical discussion, and the concluding implications of the method in terms of ethical and pedagogically reasonable the presence of AI into the realm of education.

Literature Review

Explainable Artificial Intelligence (XAI) has gained key focus of debates regarding ethical and successful AI implementation in various sectors, not exempted in the education sector. With experienced application of AI-driven systems in classroom decision-making to predict student performance, recommend an intervention or grading testing, there is a rising concern in ensuring these systems are not only precise but interpretable and reliable to educators. This literature review integrates the existing research under four primary themes, including the emergence of AI in education, explainability purpose and principles, trust and human-AI interaction in teaching, and the pedagogical implications of the emerging algorithmic decision support.

AI Integration in Education

Artificial intelligence has been gradually making its way into the academic space over the last ten years, propelled by the promise of efficiency enhancement, increased personalization, as well as making instruction data-driven. AI uses in education are as diverse as the smart tutoring systems and automatic grading tools to predictive analytics that predict student performance and prescribe an intervention. These technologies work with big learning data, and they can provide information about students, their behavior, interactions, and learning trends (Luckin et al., 2016). Education AI usage has further increased with the emergence of Learning Analytics (LA) and Educational Data Mining (EDM).

Most AI systems applied in classrooms are designed on a learner-based view even though they are becoming more adopted. The most common way of assessing AI efficacy is through outcomes of students, including grades, engagement, or retention (Zawacki-Richter et al., 2019). Nonetheless, this

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

emphasis on outputs does not take into account the mediative position of the teacher who is in many ways the person who is expected to interpret and implement AI-based findings in the classroom. Practically, educators are required to interpret AI recommendations into their individual systems of pedagogy, classroom practice and disciplinary situations.

The technocratic version of education AI is also becoming a problem of concern. Lots of systems are developed on the premise of teaching as an unliver experience, rule-based teaching, and this may result in simplistic modelization neglecting the complexity and relational basis of teaching. The algorithms can send out flagged reports which bring to attention students who are being at-risk, but unless users can explain the reasons and have an option to appeal this classification, the system might lead to dependency or even a lack of participation. One of the studies emphasizes that without being oriented towards educators, AI tools will become a hypothetical black box that teachers will remain but will not empower them (Holstein et al., 2019).

Moreover, AI tools are often implemented in classrooms without proper training and assistance in most circumstances. The teachers have been expected to have these tools in their daily schedules as well as handle the conventional roles of teaching. This has brought researchers to what they referred to as algorithmic burden, the burden on an educator to perform the cognitive and emotional work of deciphering the AI output with a small number of specifics they are not used to (Williamson and Eynon, 2020). Notably, the literature indicates a notion that AI integration should be developed along with teachers and not merely adapted by the end users. There are participatory design methods where educators are involved in the development procedure and these methods result in tools that would be more contextually pertinent, believed in, and utilized fruitfully. It is also more likely to detect and address the problems of algorithmic bias, fairness, and non-conformity to pedagogical values with this approach.

To conclude, though AI could be very helpful in enhancing the educational practice, it should be contextually, transparent, and teacher-informed when implemented in the classroom institutional environment. In the absence of this, AI has the potential to worsen the notorious educational disparities or curtail professional agency of teachers.

Explainability and Interpretability in Artificial Intelligence

With increasingly larger infiltration of AI systems in decision-making processes in industries such as education the demand to explain these decisions has become an essential design demand. The term explainable AI (XAI) can be described as the models, whose predictions or decisions can be presented using an easy to grasp rationale, and people can interpret it and query it, as well as take an appropriate action regarding the outputs of the algorithm (Miller, 2019). The question of interpretability is especially acute in the domain of education where the recommendations of AI can impact the process of grading, intervention planning, or the personalization of the curriculum.

XAI faces special difficulties in the education sector due to the fact that teaching decisions are never only linear and with pure data. Secondary school teaching is complex, contextual, and relational as opposed to subjects like finance or medicine in which the outcome is more easily measured. Consequently,

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

teachers can not always be satisfied with an output given by a system that they should learn how the student was identified as disengaged, how the risk score was determined, or which were the features that contributed the most to produce the result. Such transparency is crucial without which even accurate recommendations may be useless, or may lead to mistrust on the part of educators (Doshi-Velez and Kim, 2017).

Various categories of explainability tools have been introduced such as feature attribution techniques (e.g., SHAP, LIME), saliency maps, confidence scores, and natural language descriptions. This research has, however been conducted on technical level as little is done to determine how non technical users such as teachers understand and react to these explanations. Research indicates that the teachers tend to act based on AI outputs more when the explanations are placed in pedagogical and not mathematical contexts (Holstein et al., 2019).

The other significant difference between the literature is in global and local interpretability. Global interpretability is an understanding of the overall process of a model but on a global scale, but local interpretability looks at explaining the decisions made. Local explanations of instance, why a certain student was predicted to struggle and what exactly went into that output are the main things that teachers need to be exposed to when given individual alerts or suggestions. More importantly, interpretability is not enough. Usability and relevance of explanations is as important as technical transparency. When explanations are not clear enough, when they are too complex, when they are not actionable, teachers can chose not to pay attention to them and/or they might give them wrong interpretations. The possibility of over-trust can also exist when explanations give an illusion of knowledge when there is none of a given phenomenon known in cognitive science literature as explanation satisfaction (Miller, 2019).

Lastly, explainability touches on ethical aspects including fairness, accountability, and teacher autonomy. Educators should be in a position to question algorithms-based decisions, particularly where it is opposed to professional judgment or appears to be out of touch with classroom realities. The lack of meaningful interpretability in AI systems threatens to turn the system into an authoritative black box, which will destroy teacher confidence and pedagogical ownership. Altogether, the literature is in favor of explainable AI integration into education but it is important to note that the explanation should be pedagogically useful, cognitively feasible, and manageable by instructors. An increasing agreement is that XAI tools should be explicitly created with teacher involvement and assessed not only according to technical performance, but also with regards to their ability to facilitate practice-based teaching decisions.

Human-AI Interaction, Ethics, and Trust

Trust is a major issue of apprehension of successful implementation of artificial intelligence in education as more decision-making processes become automated. To teachers, the question of trust in AI systems lies not in the range of whether or not a given system is technically correct, but rather in whether it can be understood, whether it is in accordance with pedagogical values, and whether it is sensitive to human judgment. Even in educational settings where judgment is highly situational, value-infused, and the human-AI rapport frequently is high-

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

stakes, such an interaction should be established via ethical and relational values (Suresh and Guttag, 2021).

Research evidence in other industries, including healthcare and finances, demonstrates that users tend to place more trust in AI product when the system presents the explanation, user-reflexivity, and acts consistently across contexts (Doshi-Velez and Kim, 2017). Trust also poses a problem in education since teaching requires a relational approach to teaching is not solely a data user but a custodian of student development and well-being. Trust is easily destroyed when AI systems give recommendation that disagrees with knowledge on the behavior or history of a student held by a teacher. Educators can doubt the validity of the system or they suffer the sense that professional autonomy is being compromised (Williamson and Eynon, 2020).

The other essential element of trust is the perceived fairness and transparency of recommendations by AI. Teachers should be less likely to support or even want to use outputs of a system suspected by the teacher to be biased e.g. in disproportionately raising flags about students due to their socioeconomic status, their language proficiency, or their patterns of attendance. Accountability is another ethical issue: in the case of a low-quality decision (such as the institution sending the wrong student to the wrong course) made by an AI-driven suggestion, who should be held accountable, the teacher or the developer of such suggestions or the institution? This is an ambiguity that can leave the appearance of reluctance or unwillingness to apply AI tools (Holstein et al., 2019).

Additionally, there is an emotional aspect of education that makes the interaction between humans and AI more complicated. Teachers incorporate intuition, empathy and situation-based understanding in their judgments. His AI on the other hand works on patterns and probabilities of data. Such an epistemological difference is a tension especially where systems are viewed as being mechanistic or dehumanizing. Teachers can have an algorithmic discomfort when they feel that such alerts created by the system are forcing them to take a course of action that does not fit their own views (Zawacki-Richter et al., 2019).

Trust-construction, thus, also cannot be done without technical precision, but it entails co-design, moral regulation and continuous professional discourse. The more the teachers are engaged in the development of AI systems, the logic behind the results, and the empowerment they get to overrule or challenge the advice, the more they will trust the AI systems. There is also a building of trust based on experience; longitudinal research indicates that when the support of AI can also deliver positive outcomes, and when the explanations are also consistent and useful, the confidence level in the system increases with time (Miller, 2019). Overall, trust and ethics do not play a marginal role in AI application in educational settings they form the core. Unless paying close attention to these human-based aspects, even the most sophisticated XAI tools may be discredited or even abused or they may not create any significant effect. The literature demands open, equitable and empowering systems of teachers where AI is seen as a collaborator and not a substitute in teaching.

Pedagogical Impact/Teacher Agency

The implementation of AI into learning institutions has significant effects on

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

teacher agency and pedagogy. Although the introduction of AI-driven tools is commonly provided as a decision-support system, they are bound to have a direct impact on teacher cognition, planning, and classroom behavior. The literature also indicates the progressive worry that even in the absence of a purposeful design and professional growth, these systems can recategorize the pedagogical priorities to the point to suppress the autonomy of the teacher and change the locus of control to OE toward algorithms (Holstein et al., 2019).

A key concern is that most AI systems are built based on the optimization and efficiency incentive focusing on student performance, time of engagement or dropout potential. Although these parameters are useful, they only give a partial picture of learning, and they might not be consistent with the larger pedagogical purposes on the side of the teachers including critical thinking, creativity, or social-emotional development. The teachers might be compelled to adhere to algorithmic reasoning instead of pedagogic intuition when they are stimulated or even forced to implement a change based on system-generated suggestions with no opportunity to make any adjustments (Williamson and Eynon, 2020).

Another danger that has been produced by the literature is the de-skilling of educators, especially in the case where AI tools (e.g., grading, content order, or intervention choice) automate or otherwise take control of some instructional choices. There is a possibility that automation may decrease the administrative workload, but may also entirely decrease teacher engagement in reflective practice when the system has to outsource the basis of making a decision. In the long term, it can decline the confidence of teachers and their skills since they will have to rely more on AI and less on the possibility of a crucial decision (Perrotta et al., 2020). Nonetheless, a number of works indicate the possible positive impact of AI on pedagogy, but contingent on the systems being created as a collaborative resource instead of an authoritative substitute. As an example, AI systems that include formative observations like finding patterns in student misconception or can offer evidence-based instructions can be used to inform teacher reflexions and differentiate instructions. With a well-developed professional learning communities and continuous feedback loops, these tools can, in fact, result in more in-depth pedagogical work and innovation (Roll and Wylie, 2016). The amount of control and customization of AI tools which educators have also influences teacher agency. On the other hand, locked or centrally administered systems will tend to disempower teachers and cause a compliance culture as opposed to an inquisitive culture (Holstein et al., 2019).

Last, the symbolic use of AI in institutions may reflect on the way teachers view themselves as authoritative. At institutions of learning where artificial intelligence is a central aspect of innovation or quality control, teachers will experience a sense of implicit evaluation or assessment through the degree to which they follow AI-driven recommendations. This relationship may cause tensions between professionalism of the teacher and institutional requirements that will only make the process of integration more complex. It is advocated in the literature that AI systems must be designed to be pedagogically compatible, context sensitive, and an equal partnership with teachers in such a manner that a collaborative technological development does not inflate the teaching profession.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

Methodology

Methodological Approach

This research problem was to understand how teachers think and feel about the application of Explainable AI (XAI) in classroom decision-making. A qualitative phenomenological approach was preferred to get the subjective, fine and contextual nature of these experiences. In this case, the integration of the algorithmic decision-support systems into the educational practice, phenomenology is appropriate to study the experiences that people have and understand how they create meaning of that phenomenon in their lives. The research utilized primary data; it was descriptive and exploratory and concerned the reflections, emotional as well as the interpretations of participants.

Data Collection Methods

The data were gathered using semi-structured interviews with 12 educators (8 secondary, 4 post-secondary) who had practical experience in using AI-driven classroom tools which provided explainable functionality (predictive analytics dashboard, grading assistants, or a learning intervention platform).

- The participants were selected by means of purposive sampling, according to the experience with the XAI-enabled systems during the last two years.
- The interviews took place in Zoom and lasted between 45-70 min and were conducted around the themes of how digital decision-making works, the level of trust in AI, the perception of accuracy and transparency, ethical issues, and the adjustment of instructional practice.
- To complement the interviews, the topical artifacts like the screenshots of AI interfaces, teacher notes, or institutional guidelines were consulted to get a background idea.
- All interviews were transcribed verbatim and anonymized with pseudonyms, informed consent was received, and all received information was anonymized.

Data Analysis Methods

Braun and Clarke (2006) six step process was used to conduct the thematic analysis. The transcripts were entered into NVivo 12 and open coding revealed the first themes such as trust, transparency, conflict with professional judgment and pedagogical alignment. Introduced next, the themes were systematized in broader categories in order to reflect the lived experience of using XAI in practice by employing the method of axis coding. The analysis has also focused on in-vivo coding so as to save the language of the participants especially in instances of expressing emotions, doubts, or insights. The 10 th interview was a data saturation point after which the other two were utilised to validate and refine themes.

Evaluation and Justification

The specified phenomenological approach was suitable in the studies of the root-based perceptions and interpretations of some new technological intervention. Semi-structured interviews were used to enable flexibility to delve into personal narratives, and thematic analysis was used to aid the detection of patterns of various teaching settings.

The small sample size and possible self-selection bias are the limitations because

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

they might be more inclined to volunteer the ones who had strong opinions and were more digitally literate. As well, due to the variety of AI tools utilized in schools, there is no direct comparability. The limitations were alleviated by member checking, using various sources of data, and triangulating artifacts and institutional documents. The analytic rigour and validity were also improved with the help of researcher reflexivity and peer debriefing.

Results

The results of the thematic analysis of interviews with 12 educators consist of four significant themes that describe the varied and multiple experiences of teachers working with Explainable AI (XAI) tools to make classroom decisions. The results are discussed below based on the individual themes along with quotations of some participants.

Interpretability and Cognitive Overload

Ten of 12 teachers claimed they had a problem in following the explanations given by the systems.

"The system informed me that the student was at risk of 85 percent, but not what that number actually implied in reality, whether this was attendance, engagement, and so on. Participant T5

Educators reported explanations being either over reductive or vaguely worded, which leads to the intellectually surcharging experiences of attempting to interpret the information in the process of instructional planning.

Reliance and Confidence to Human Judgment

But whereas objectivity of AI outputs was valued by some teachers, most (9 of 12) of the teachers tended to go with their own judgement as opposed to algorithmic recommendations, particularly when the recommendations appeared to be contrary to the observations.

I have noticed that the tool has indicated that one of the students is disengaged, but she attends all the classes regularly. I just didn't trust it." Participant T3

Under some conditions, some teachers referred to the use of AI recommendations as the second opinion, not a directive. They placed greater importance on their professional intuitions than data-driven prompts in case there was a conflict.

Student Perceptions and Ethical Concerns

The ethical issues of bias, labeling, and student privacy were raised by seven participants. The crucial reason was feared to reinforce stereotypes or damage the student-teacher relationships.

I would also be concerned that students would find out they are being labeled alongside an algorithm, and trust would be harmed. Participant T1

The use of the AI-generated classifications (e.g., the low-performing) in a manner that would influence their expectations or develop unconscious bias was especially uncomfortable to teachers.

Influence on Pedagogical practice and Emotional Response.

Eight out of 12 teachers stated that they had mixed feelings regarding the impact of AI tools on the planning of the instruction. There would be those who liked the

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

time-saving advantages and new knowledge, and those, who were emotionally divided, especially when they had no choice but to act on the results they did not have a complete understanding of.

"It helps but I can sometimes think that I am only responding to the instrument rather than the ways in which I do things. Participant T8

Two teachers expressed stress or anxiety in regards to the interpretation and implementation of system suggestions, especially when there is a performance data that applied to their own judgment as instructors.

Limitations Over which the Participants were Aware

The participants mentioned that there were certain practical drawbacks of the AI tools such as:

- Absence of contextual input fields
- Unresponsiveness of feedback formats.
- Inadequate training in system functionality and reading of data.

They frequently resulted in the underutilization or conservative usage of the tools despite the perceived benefits apparent to the participants.

Discussion

This paper intends to look at the use of Explainable AI (XAI) in helping teachers in secondary and post-secondary schools to make classroom decisions. There were four overall themes, including challenges in understanding AI explanations, biased and partial trust and reliance on professional judgment, ethical issues related to bias and student perception, and ambivalent impacts on pedagogical practice.

The results show that though teachers realize the possibilities of XAI to facilitate instruction, they may tend to have cognitive and emotional dissonance when explanations fail or are not consistent with their own intuition. Ethical uneasiness especially on student labeling and privacy also influenced the willingness of teachers to adhere or implement AI suggestions. Finally, the system predicts only part of the effectiveness of XAI in classrooms, and there is a need to communicate the rationale and accommodates the context of the educator.

The results are consistent with the previous studies that highlight the importance of interpretability and usability of AI tools by non-technical users (Miller, 2019); (Doshi-Velez and Kim, 2017). Similar to Holstein et al. (2019), the given study confirms that explainability is a necessity to raise trust but has to be pedagogically meaningful and not just transparent. Ethical issues raised by teachers resonate with the alarm raised by Williamson and Eynon (2020) when it comes to the phenomenon of algorithmic governance that creates educational relationships and judgments. Notably, the emotional and professional struggles that the participants reported are similar to those mentioned in Roll and Wylie (2016), where the use of AI systems was reported to challenge instead of support teacher autonomy without close incorporation.

The small sample size and the variability in the XAI tools employed in this study hamper the study as it could introduce inconsistency that may be present across experiences. Furthermore, participants that were more digitally literate or who showed more interest in AI might have been more likely to take part which includes the selection bias. This was due to the fact that the longitudinal tracking

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

was not done and therefore any long-term changes in trust or practice could not be evaluated.

The analysis shows that there are a number of important implications to the study:

- To encourage the use of XAI to enable teachers interpret and assess outputs, professional development should accompany the adoption.
- Educators should co-design XAI systems including contextual data and having meaningful ifrater choices.
- Ethical transparency is a priority that should be considered by the policymakers and the edtech creators in order to guard the student dignity and the discretion of the teachers.
- These modifications are critical to make sure that XAI will improve and not affect teacher agency and pedagogical integrity.

There is a possibility that it is institutional dynamics or digital outbursts, rather than the technology, which makes mistrust in AI. Other teachers might have an objection towards AI based on more worldly fears of surveillance or job assessment, as opposed to any actual opposition to outputs. These psychological and structural drives could be separated in further studies.

This research was directly related to its main question:How do teachers experience and interpret XAI in classroom decision-making? It discovered that teachers are receptive to data-driven support; however, they are sceptical and suspicious when explanations lack or are ethically questionable. This subtle view adds new implications on the conditions of pedagogical empowerment of XAI.

Conclusion

The main question that was addressed in this study was as follows: How do teachers perceive and experience the use of Explainable AI (XAI) tools in the classroom decision-making? It explored the real-life experiences of teachers who are on the crossroads of human judgment and machine-generated knowledge in actual practice.

The results demonstrated the main premise of the study: XAI can potentially lead to positive changes in an educational decision but it should be designed with due care to avoid damaging teacher trust, interpretability, and autonomy. Ease and data-rich support of XAI tools were valued by teachers, whose most common problem was vague or too technical explanations. In case the outputs of AI contradicted professional intuition or appeared to be unethical, teachers reverted to their judgment pointing to the necessity of more pedagogically sound and context-responsive systems.

The study is part of the growing literature about AI in education since it puts educators, who are primarily charged with the responsibility of putting AI recommendations into practical, student-facing choices into practice, at the heart of the study. It points out that the real achievement of XAI in education is not only in the accuracy of algorithms, but also in the ability of such systems to serve, not limit, an instructional practice.

Going back to the arguments presented in the introduction, the study disproves the fact that explainability is sufficient reason to be used. Rather, it reveals that meaningful integration requires teacher agency, training, and ethical trust to take place. Such results are particularly applicable due to the continued integration of AI into fundamental pedagogical processes in K-12 and.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 3 No. 9 (September) (2025)

So what? This study highlights the fact that XAI should not just be created to teachers, but to teachers. There should be co-design processes, contextualizing customization, and support frameworks that enable the educators and not only inform them, in the future. The future of AI-enhanced education can be spoiled by distrust, abuse, or non-engagement without such measures in place.

Concluding, longitudinal influences of XAI in the teaching identity, student learning outcomes, and institutional culture should be examined in the future so that the innovation of the AI will truly benefit the human aspect of learning.

References

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. https://arxiv.org/abs/1702.08608

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (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* (pp. 1–16). https://doi.org/10.1145/3290605.3300830

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. *Pearson Education*. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/Intelligence-Unleashed-Publication.pdf

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Perrotta, C., Selwyn, N., & Player-Koro, C. (2020). Automation and the teaching profession: Understanding the impact of AI and robot technologies on education. *Teaching and Teacher Education*, 96, 103182. https://doi.org/10.1016/j.tate.2020.103182

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Suresh, H., & Guttag, J. V. (2021). A framework for understanding unintended consequences of machine learning. *Communications of the ACM*, 64(3), 62–71. https://doi.org/10.1145/3430368

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0