QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

DESIGNING A STUDENT-CENTRIC FRAMEWORK FOR EXPLAINABLE AI IN ADAPTIVE LEARNING SYSTEMS: TOWARDS TRANSPARENT, TRUSTWORTHY, AND ETHICAL AI IN EDUCATION

Hareem Arif
Visiting Faculty Member, Department of Arts and Social Sciences, University of Education.
Main Course Tutor, CELTA (Department of Cambridge Assessment English), University of Cambridge, UK. Email: hareem.aarif@gmail.com

Abstract
The integration of Artificial Intelligence (AI) technologies into educational environments—particularly through Adaptive Learning Systems (ALS)—has fundamentally transformed the delivery of instruction and personalized learning. These intelligent systems leverage machine learning algorithms to continuously analyze learner data and adapt content, pacing, and pedagogical strategies to individual student needs, leading to enhanced engagement and improved learning outcomes. However, this rapid digital evolution has surfaced critical concerns about the opacity, fairness, and accountability of AI-driven decisions. Often referred to as "black-box" systems, many AI models offer little to no insight into how they arrive at recommendations or interventions, thereby posing challenges related to transparency, interpretability, trust, and ethical governance. This paper argues that to responsibly integrate AI in education, especially in contexts where student data and futures are at stake, it is essential to move beyond system performance and embrace Explainable AI (XAI)—a design philosophy and technical paradigm focused on making AI decision-making understandable to humans. We propose a student-centric framework for XAI in ALS, which places learners and educators at the center of AI design and deployment processes. The framework is informed by interdisciplinary literature and practical case studies in educational technology and AI ethics. It emphasizes three foundational pillars: (1) Human-centered design, which ensures that AI interfaces and feedback mechanisms are accessible and usable by students and educators alike; (2) Stakeholder collaboration, which fosters ongoing dialogue between developers, teachers, learners, ethicists, and policymakers; and (3) Ethical integration, which embeds principles such as fairness, inclusivity, data privacy, and algorithmic accountability into the development lifecycle of adaptive learning tools. The framework also outlines actionable strategies to ensure that explanations provided by AI systems are contextually meaningful, developmentally appropriate, and pedagogically aligned. It considers the differentiated cognitive needs of various learner groups and promotes educational transparency as a right, not a privilege. Ultimately, this student-centric XAI framework aims to cultivate trustworthy, ethical, and pedagogically sound AI systems that not only support learning but also uphold the rights, agency, and dignity of learners in the digital age. By advocating for transparency and ethical accountability, this work contributes to the growing body of research on responsible AI in education and offers practical guidance for institutions aiming to deploy adaptive technologies in equitable and explainable ways.

Keywords: Explainable Artificial Intelligence (XAI), Adaptive Learning Systems (ALS), Human-Centered Design, Educational Ethics, AI Transparency in Education

1. Introduction
The integration of Artificial Intelligence (AI) into education marks one of the most significant paradigm shifts in the digital transformation of teaching and learning. Among AI applications, Adaptive Learning Systems (ALS) have garnered particular attention for their ability to personalize instruction by dynamically adjusting content, feedback, and learning pathways based on student behavior, knowledge levels, and learning pace (Kerr & Chung, 2021). These systems rely on advanced data analytics and machine learning algorithms to continuously assess a learner's progress and tailor instructional delivery to meet individual needs—offering the promise of more equitable and efficient education at scale. However, as the use of ALS grows, so too do concerns over algorithmic opacity, often described as the "black-box" problem. Many current AI models, particularly those employing deep learning techniques,

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

operate in ways that are not easily interpretable by end-users—students, teachers, or even developers (Adadi & Berrada, 2018). This opacity undermines the transparency and accountability of AI-driven decisions in educational contexts, where outcomes may impact student placement, progression, and performance evaluations. Without a clear understanding of how recommendations or interventions are generated, students may feel disempowered, and teachers may be hesitant to rely on system outputs for high-stakes decisions (Liu et al., 2021). The lack of interpretability also poses ethical challenges. AI systems can inadvertently perpetuate biases, marginalize certain learner groups, or operate with assumptions that may not be pedagogically sound (Scherer, Zhang, & Müller, 2021). For example, algorithms trained on historically biased datasets may reproduce systemic inequalities in access to resources or feedback quality. Moreover, students from linguistically or culturally diverse backgrounds may not benefit equitably from adaptive recommendations that do not consider their specific learning contexts (Holstein et al., 2019).To address these issues, the educational technology community has begun exploring Explainable AI (XAI)—a field of AI research focused on developing models and interfaces that make AI decision-making understandable and transparent to humans. XAI is particularly critical in domains like education, where trust, interpretability, and learner agency are essential components of ethical system design (Miller, 2019). Effective XAI not only facilitates informed user engagement but also enables the identification of errors, biases, and misalignments between system behavior and instructional goals. Yet, most current XAI efforts in education are system-centered, focusing on improving transparency for developers or system administrators, with less attention paid to student-facing explanations. This gap has led to calls for a student-centric approach to XAI that recognizes learners as primary stakeholders and information consumers. According to Luckin (2022), truly responsible AI in education must be human-centered, prioritizing user experience, ethical inclusivity, and developmental appropriateness in its design philosophy. The present study contributes to that growing discourse by proposing a comprehensive student-centric framework for Explainable AI in Adaptive Learning Systems. Grounded in interdisciplinary research across education, AI ethics, learning sciences, and HCI, the framework promotes three foundational principles:

1. Human-centered design, which ensures that explanations are usable, understandable, and meaningful for learners and educators;
2. Stakeholder collaboration, which fosters inclusive participation from developers, teachers, students, policymakers, and ethicists throughout the AI development lifecycle; and
3. Ethical integration, which embeds fairness, transparency, data privacy, and algorithmic accountability into system architecture from the outset.

The proposed framework not only supports technical and pedagogical alignment but also emphasizes the importance of giving learners a voice in how educational technologies are built and used. It recognizes that educational transparency is not merely a design feature but a fundamental right, especially in digital learning environments where AI systems increasingly mediate access, assessment, and achievement.

2. Literature Review

2.1. Adaptive Learning Systems in Education: Adaptive Learning Systems (ALS) represent a major advancement in the use of Artificial Intelligence (AI) within educational environments. These systems leverage data-driven approaches and machine learning algorithms to dynamically adjust instructional content, pace, and feedback according to the needs and performance of individual learners (Kerr & Chung, 2021). By collecting and analyzing a wide

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

array of learner data—including quiz results, time spent on tasks, engagement metrics, and behavioral patterns—ALS personalize the learning journey, making it more responsive and targeted. In higher education, ALS are increasingly used to support large-scale personalized instruction, where instructors cannot manually adapt materials for every student. For example, platforms like ALEKS and Knewton use real-time diagnostics and intelligent content recommendations to guide students through mastery-based progressions in mathematics and science (Chen et al., 2020). Similarly, in K–12 education, adaptive learning platforms help bridge learning gaps by adjusting difficulty levels based on formative assessments (Yudelson et al., 2017). Research suggests that well-implemented ALS can improve learner outcomes, engagement, and retention (Pane et al., 2017). However, the effectiveness of such systems often depends on the quality of the underlying algorithms, the relevance of the adaptation mechanisms, and the transparency of system decisions.

2.2. Explainable AI (XAI) and its Relevance in Education: Explainable AI (XAI) refers to methods and techniques in AI that make the decision-making processes of algorithms transparent and interpretable to human users (Doshi-Velez & Kim, 2017). In education, the demand for explainability is especially pressing due to the high-stakes nature of decisions—such as assessments, recommendations, and learning interventions—made by AI systems. Learners and educators must understand not only *what* the system recommends, but *why* it does so, in order to act meaningfully on the information. Studies have shown that when students are provided with understandable explanations for system recommendations, they are more likely to trust and engage with AI-powered learning tools (Liu et al., 2021). For example, if a system recommends reviewing a particular topic, a transparent explanation such as "You scored below 70% on three quizzes in this area" supports learner autonomy and metacognition. Instructors, too, benefit from explainable models, as they can better interpret student progress reports and modify pedagogical strategies accordingly. Moreover, explainability enhances system accountability by allowing educators to identify when the AI may be making flawed or biased recommendations (Holstein et al., 2019). Despite its importance, explainability in educational AI remains underdeveloped, many ALS rely on complex deep learning models whose internal logic is inaccessible to most users. Bridging this gap requires a shift toward learner-facing XAI strategies that are developmentally appropriate and pedagogically aligned (Miller, 2019).

2.3. Ethical Concerns in AI-based Education: The use of AI in educational contexts raises profound ethical questions, particularly when algorithms influence student assessment, access to learning resources, or engagement strategies. Without deliberate ethical design, ALS can amplify existing inequalities and unintentionally marginalize vulnerable learner populations (Scherer et al., 2021). Key ethical concerns include:

1. Bias in data and algorithms, which can disadvantage students from underrepresented backgrounds;
2. Lack of transparency, which obscures how decisions are made;
3. Inadequate consent mechanisms, where learners may not be fully aware of how their data is used;
4. Surveillance concerns, where constant data monitoring affects student autonomy and psychological safety (Selwyn, 2020).

Scholars argue that educational AI systems must be built on principles of fairness, inclusivity, and data privacy. This includes conducting bias audits, enabling user control over data sharing, and embedding algorithmic accountability from design to deployment (Luckin, 2022). Furthermore, the ethical use of AI in education necessitates a participatory design approach—

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

one that involves students, educators, and ethicists in shaping system behavior and governance structures (Holmes et al., 2022). This is particularly vital as AI becomes more deeply embedded in learning environments, influencing not just outcomes but the broader experience of education itself.

3. The Need for a Student-Centric Framework

The evolution of Explainable Artificial Intelligence (XAI) in education has predominantly been driven by technical and developer-oriented concerns. Much of the research and development in XAI focuses on algorithmic transparency for engineers, researchers, or administrators, rather than for the students who are most directly impacted by the system's decisions (Adadi & Berrada, 2018). This gap in design and deployment neglects a vital truth: students are not mere data generators or passive recipients of AI feedback—they are active participants whose understanding, trust, and autonomy must be prioritized for ethical and effective implementation of educational AI systems (Liu et al., 2021). A student-centric XAI framework acknowledges the unique role of learners and reorients the design of AI systems to prioritize four essential dimensions: accessibility of explanations, pedagogical alignment, cognitive appropriateness, and the protection of learner agency and dignity.

3.1 Accessibility of Explanations: Most current AI explanations are framed in technical terms or statistical metrics, which are often incomprehensible to non-expert users, including students. For example, explanations like "This recommendation was made with 83% confidence based on logistic regression weights" offer little meaning to a high school student navigating an adaptive math platform. Instead, explanations should be presented in language and formats that are age-appropriate and linguistically inclusive (Miller, 2019). Accessible explanations may include visual feedback, storytelling metaphors, or simplified natural language summaries. Studies in Human-Computer Interaction show that students are more likely to trust and engage with systems when they receive clear, comprehensible, and actionable explanations (Kulesza et al., 2015).

3.2 Pedagogical Alignment: For explanations to be meaningful in education, they must not only clarify AI behavior but also support learning goals. This means aligning AI explanations with instructional strategies, curriculum objectives, and pedagogical intent. When students understand why the AI recommends a certain activity (e.g., revisiting fractions due to consistent low scores), they are better equipped to **self-regulate** and take ownership of their learning process (Rosé et al., 2018). Educators also benefit from pedagogically aligned explanations, as they can contextualize AI recommendations within broader instructional frameworks and ensure that AI actions complement—not conflict with—teacher-driven pedagogy (Holmes et al., 2022).

3.3 Age-Appropriate Cognitive Design: Children and adolescents differ from adults in terms of cognitive capacity, metacognitive awareness, and decision-making ability. Therefore, a one-size-fits-all model of explainability is insufficient. A student-centric framework should consider developmental psychology and learner diversity when designing explanatory interfaces. For younger learners, gamified feedback or animated explanations may be more effective; for older students, interactive dashboards that show progress over time may enhance comprehension and motivation (Liu et al., 2021; Luckin, 2022). Moreover, explanations should not overload students with information. Research in cognitive load theory suggests that poorly designed interfaces may overwhelm learners, reducing engagement and impairing performance (Sweller et al., 2019).

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

3.4 Protection of Learner Agency and Dignity: Perhaps the most crucial aspect of a student-centered approach is the recognition that learners have a right to transparency, autonomy, and dignity in AI-mediated environments. Explanations must empower students to question, override, or opt out of AI decisions when necessary. Embedding mechanisms for user feedback, consent, and challenge fosters an environment where learners feel respected, rather than surveilled or manipulated (Scherer et al., 2021; Selwyn, 2020). Incorporating these principles not only promotes trust and ethical engagement with AI but also ensures that adaptive technologies serve educational equity, rather than exacerbate existing divides.

4. Proposed Framework

In response to the growing concerns surrounding the opacity, ethical risks, and pedagogical misalignment of AI in education, we propose a student-centric framework for integrating Explainable AI (XAI) in Adaptive Learning Systems (ALS). This framework is built on three foundational pillars: (1) Human-Centered Design, (2) Stakeholder Collaboration, and (3) Ethical Integration. Together, these pillars provide actionable guidance to developers, educators, and policymakers seeking to create transparent, inclusive, and pedagogically aligned AI tools.

4.1. Pillar 1: Human-Centered Design: Human-centered design in AI emphasizes usability, accessibility, and contextual relevance. In educational settings, it requires that AI-generated explanations be intelligible to students, educators, and caregivers, taking into account diverse cognitive abilities and language preferences (Miller, 2019; Liu et al., 2021). Rather than focusing on model interpretability for engineers alone, human-centered design prioritizes user-facing explainability—the capacity for students to understand *why* the system recommended a specific path or identified a knowledge gap. For example, rather than presenting technical metrics like confidence intervals or feature weights, the system should use plain language, e.g., "You are recommended this reading because your last two quiz attempts on this topic scored below 60%." Studies have shown that such contextual, developmentally appropriate feedback boosts learner trust and engagement (Kulesza et al., 2015; Holstein et al., 2019). An Example is that Implement interactive visual dashboards where students can view their progress with clear, contextual explanations. For example: *"Your performance in fractions improved by 20% after completing the interactive game module. "*This approach also supports metacognition, as learners are able to reflect on their progress and the reasoning behind system suggestions (Liu et al., 2021).

4.2. Pillar 2: Stakeholder Collaboration: Educational AI systems do not operate in isolation—they impact and are influenced by a variety of stakeholders including students, teachers, school leaders, parents, developers, ethicists, and regulators. A student-centric framework must include collaborative design processes that elevate these voices, particularly those of teachers and learners, in shaping system behavior and interpretability (Luckin, 2022). Involving teachers in the co-design of XAI ensures that explanations align with curriculum goals and pedagogical strategies. For instance, if the AI recommends revision material, teachers can help ensure that the suggested topics and explanations align with national standards or assessment criteria (Holmes et al., 2022). Additionally, student feedback is crucial for refining the clarity and relevance of system-generated explanations. Participatory design empowers students by making them co-creators of their learning environment—enhancing agency and ethical responsibility (Sambasivan et al., 2021). Example strategy is to establish interdisciplinary design committees that include student representatives, educators, HCI experts, and AI developers. These teams should meet regularly during the design and

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

evaluation phases of AI system development to iteratively improve explainability and trustworthiness.

4.3. Pillar 3: Ethical Integration: Embedding ethics into the lifecycle of AI systems in education is essential to prevent unintended harm and promote equitable outcomes. A student-centric XAI framework must incorporate ethics-by-design, ensuring that fairness, privacy, and accountability are foundational rather than optional features (Scherer et al., 2021). Key components of ethical integration include:

1. Bias mitigation: AI systems must be audited for biases related to gender, ethnicity, language proficiency, and socio-economic background. Studies have shown that unchecked biases in ALS can disproportionately disadvantage underrepresented groups (Cowgill et al., 2021).
2. Data minimization and protection: Educational AI systems often rely on sensitive personal data. Adhering to data minimization principles ensures that only essential data is collected, processed, and stored. This is in line with global data protection frameworks such as GDPR and supports student dignity and privacy (Selwyn, 2020).
3. Accountability mechanisms: Systems should include channels for challenging or reviewing AI decisions, especially in high-stakes contexts such as grading or course placement. Ethical oversight bodies or internal review panels can provide governance structures for monitoring AI impact over time (Floridi et al., 2018).

Example Strategy for this is to Conduct periodic algorithmic audits to detect and correct biases in recommendation models. For instance, if the system disproportionately flags girls for remedial math modules compared to boys with similar scores, developers must revise the model inputs and retrain the system. These audits should be documented and reviewed by an external ethics board. Together, these three pillars form a robust and actionable foundation for designing explainable AI systems that serve educational goals without compromising learner trust, agency, or fairness. By centering design around the needs of students and involving a broad network of stakeholders, educational institutions and ed-tech developers can build AI tools that are not only intelligent—but also just, inclusive, and trustworthy.

5. Implementation Strategies

Operationalizing a student-centric Explainable AI (XAI) framework in Adaptive Learning Systems (ALS) requires actionable and context-specific strategies. While the foundational principles—human-centered design, stakeholder collaboration, and ethical integration—form the philosophical backbone, implementation must address the day-to-day interaction between users and the AI system. This section outlines three practical strategies to bring XAI into effective, equitable, and pedagogically sound use in real-world learning environments which are layered explanations, contextual and timely feedback, and personalization of explanations.

5.1. Layered Explanations: A one-size-fits-all approach to explain fails to meet the needs of the diverse range of users in educational environments. Teachers, students, school administrators, parents, and developers each engage with adaptive systems for different reasons and require distinct levels of detail from AI-generated explanations (Liu et al., 2021). A layered explanation strategy—sometimes referred to as "tiered explainability"—provides multiple levels of explanation appropriate to each user's role, cognitive capacity, and decision-making responsibility.

1. For students, explanations should be concise, actionable, and delivered in developmentally appropriate formats. For instance, "You're recommended this activity because you had

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

difficulty with similar problems yesterday" is more understandable and motivating for a student than statistical confidence scores.

2. For teachers, a deeper layer of analytics is often needed. This might include visualizations of student progress over time, heat maps of areas where students struggle, or summaries of engagement trends, all contextualized within the curriculum. These explanations allow teachers to identify where AI outputs align—or misalign—with classroom assessments and intervene accordingly (Wang et al., 2019).
3. For developers and researchers, the system should generate comprehensive technical logs, including feature weightings, algorithmic pathways, and model drift reports. These layers help with model debugging, compliance auditing, and continual improvement of the AI system (Guidotti et al., 2018).

An ALS platform could present a simple interface to students, with expandable layers accessible by teachers and administrators, each layer containing progressively more granular information. Such modular architectures enable customized user experiences without sacrificing interpretability at any level.

5.2. Contextual and Timely Feedback: Explainability must be integrated within the learning process, rather than appended as a post-hoc justification. Research in learning sciences emphasizes the importance of timely, formative feedback in supporting student motivation, metacognition, and learning outcomes (Rosé et al., 2018). When explanations are delayed or overly abstract, their value diminishes significantly—students may not remember the action that prompted the feedback, or may fail to connect it to their goals. Timely explanations also align with the principles of Just-In-Time Learning (JITL), which suggest that feedback should be presented at the moment when the learner is most cognitively receptive (Kaplan et al., 2022). For example, immediately after a quiz, an explanation such as "You missed this question because you haven't reviewed the concept of equivalent fractions" can reinforce conceptual learning while motivation is still high. Contextual feedback is also key—it should be tailored to what the student was doing, their learning history, and the current learning objective. Generic or vague explanations ("You need more practice") may erode trust or lead to disengagement. Integrate pop-up explanations into learning platforms that activate at key decision points—e.g., quiz submission, topic selection, or course branching—so that learners can immediately understand the rationale behind each system decision.

5.3. Personalization of XAI: Just as ALS tailor instructional content based on learner data, XAI must adapt the form and delivery of explanations to individual learner preferences, backgrounds, and cognitive needs. This is particularly critical in diverse classrooms where students vary in age, linguistic proficiency, neurodiversity, and learning strategies. Research by Khosravi et al. (2022) in intelligent tutoring systems shows that students are more likely to benefit from explanations that match their preferred learning style—whether visual, auditory, textual, or interactive. Similarly, non-native speakers may require translated or simplified explanations to ensure comprehension without cognitive overload (Holstein et al., 2019). Personalization in XAI can be implemented through Learner profiles that store preferences for explanation style (e.g., text-based, video summaries, charts). Dynamic adaptation based on student responses, engagement patterns, or confusion indicators and Gamification elements that embed explanations within reward-based progress systems. Ethical personalization also involves allowing students to set boundaries around how much explanation they want and when, thus respecting learner autonomy (Scherer et al., 2021). An XAI interface could offer three explanation modes: basic, detailed, and visual. Students could choose their preferred

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

mode at any point, and the system would adapt explanations accordingly. Additionally, the system could adjust explanation complexity based on ongoing assessments of the student's cognitive load or feedback. For Explainable AI to be effective in adaptive learning environments, it must not only be accurate and fair but also usable, relevant, and personal. The strategies outlined above—layered explanation delivery, real-time contextual feedback, and explanation personalization—bridge the gap between technical transparency and educational efficacy. These strategies ensure that all stakeholders, particularly students, are equipped to engage meaningfully with AI tools, fostering a sense of agency, understanding, and trust in their learning journey.

6. Challenges and Future Directions

While the incorporation of Explainable AI (XAI) into Adaptive Learning Systems (ALS) holds significant potential to enhance trust, fairness, and pedagogical alignment, its integration is not without critical challenges. These challenges stem from technical, institutional, and infrastructural limitations that must be systematically addressed to realize the full benefits of XAI in education.

6.1. Challenges in Integrating XAI into ALS

6.1.1. Trade-off Between Explainability and Model Accuracy: One of the most fundamental dilemmas in XAI is the trade-off between model complexity and interpretability. Highly accurate models, such as deep neural networks, often function as "black boxes" that lack transparency, whereas more interpretable models (e.g., decision trees or linear regressions) may not achieve comparable predictive performance in complex learning environments (Carvalho, Pereira, & Cardoso, 2019). In the educational context, this trade-off becomes even more consequential, as decision-making affects student learning trajectories, assessments, and motivational pathways. Sacrificing accuracy for interpretability could lead to incorrect recommendations that affect student progress, while prioritizing accuracy over transparency risks undermining trust and ethical integrity. Resolving this tension requires innovative approaches to model-agnostic explanations and hybrid systems that combine interpretability with predictive robustness (Guidotti et al., 2018).

6.1.2. Scalability of Personalized Explanations: Another significant challenge is the scalability of personalized explanations. While personalization is central to effective XAI in education, delivering tailored, developmentally appropriate explanations to thousands or millions of learners across varied educational levels and linguistic backgrounds demands significant computational resources and design flexibility (Liu et al., 2021). Furthermore, real-time personalization must balance user preferences with system responsiveness. Excessive personalization could lead to inconsistent messaging or increased cognitive load for students, while under-personalization may result in disengagement or mistrust. Designing scalable solutions—such as modular explanation templates, AI-assisted natural language generation, or semi-automated dashboards—remains a key area of exploration.

6.1.3. Institutional Resistance and Market Priorities: A less technical but equally important challenge stems from resistance among educational technology vendors and institutions. Many commercial ALS products prioritize performance metrics, efficiency, and market competitiveness over transparency or ethical design (Sambasivan et al., 2021). Explainability is often viewed as a secondary feature or even a liability that could expose proprietary algorithms or open the system to user scrutiny. This resistance is compounded by a lack of enforceable regulatory frameworks and standardized guidelines that mandate transparency, leaving it up to individual companies or institutions to voluntarily implement XAI. Without

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

clear incentives or requirements, many vendors may continue deploying opaque systems that meet performance benchmarks while ignoring deeper issues of accountability and equity.

6.2. Future Directions

Despite these challenges, several research and policy directions hold promise for advancing the field of XAI in education:

6.2.1. Development of Domain-Specific XAI Tools for Education: General-purpose XAI techniques, such as LIME or SHAP, were not designed with educational contexts in mind. There is a growing need to develop domain-specific XAI tools tailored for learning environments. These tools should account for pedagogical theories, curriculum structures, and learner psychology, thereby producing explanations that are not only technically valid but also instructionally meaningful (Holstein et al., 2019; Khosravi et al., 2022). For instance, an explanation module in an ALS could be designed to provide pedagogical rationale—such as Bloom's taxonomy alignment or learning objective mapping—behind each AI-driven recommendation.

6.2.2. Longitudinal Studies on XAI's Impact on Trust and Learning: Most existing studies on XAI are short-term or pilot-based. There is a pressing need for longitudinal research that examines how explainability affects student trust, motivation, and academic outcomes over time. Such studies can help determine whether XAI promotes sustained engagement or whether the novelty of explanations fades without meaningful pedagogical integration. Moreover, these studies should be inclusive of diverse learner groups, including neurodivergent students, English language learners, and students from underrepresented socio-economic backgrounds, to assess how different populations experience and benefit from XAI (Scherer et al., 2021).

6.2.3. Regulatory and Ethical Governance Frameworks: ensure responsible AI integration in education, stronger regulatory frameworks are essential. Governments and educational bodies must develop policies that mandate algorithmic transparency, student data rights, and ethical oversight. The European Commission's 2021 proposal for regulating AI, which designates educational AI systems as "high risk," is a step in this direction and may serve as a blueprint for other regions (European Commission, 2021). Institutions should also establish internal ethics review boards for educational technology procurement and deployment. These boards can evaluate AI systems for bias, fairness, and explainability before they are introduced into classrooms.

7. Conclusion

As Artificial Intelligence becomes increasingly embedded in the fabric of educational systems worldwide, it is no longer sufficient to evaluate its value solely on the basis of efficiency, personalization, or performance metrics. The true promise of AI in education lies not just in its ability to process data and adapt content, but in its capacity to support meaningful, ethical, and empowering learning experiences. In this context, transparency and ethical responsibility are not peripheral features—they are essential foundations upon which responsible educational innovation must be built. This paper has proposed a student-centric framework for Explainable Artificial Intelligence (XAI) within Adaptive Learning Systems (ALS), grounded in the principles of human-centered design, stakeholder collaboration, and ethical integration. Unlike conventional XAI approaches that focus primarily on system-level or developer-oriented explanations, this framework reorients the locus of interpretability toward the student—recognizing the learner as an active agent in the AI-mediated learning process. By emphasizing accessibility of explanations, pedagogical alignment, and age-appropriate cognitive design, the framework ensures that AI systems do not merely "personalize learning," but actually promote

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

understanding, engagement, and trust. Explanations are not just diagnostic tools; they are educational artifacts that can reinforce metacognition, autonomy, and learner agency when crafted with care and empathy. Furthermore, the inclusion of multi-layered explanations enables differentiated access to information for students, teachers, developers, and policymakers, ensuring that every stakeholder is equipped with contextually relevant insights. When paired with timely, real-time feedback and personalized delivery mechanisms, XAI becomes a dynamic partner in the learning journey—guiding students through challenges, reinforcing strengths, and illuminating the rationale behind each recommendation or intervention. Nevertheless, this vision is not without obstacles. As discussed, significant challenges remain in terms of balancing model accuracy with interpretability, ensuring scalability across diverse educational environments, and overcoming institutional resistance to transparency. However, these challenges are not insurmountable. With ongoing investment in domain-specific XAI research, longitudinal studies, and the establishment of regulatory frameworks that mandate transparency, the path toward truly ethical and equitable AI in education becomes clearer. Ultimately, the success of AI-enhanced education depends not on how smart our systems become, but on how well they align with the human values that underpin meaningful education—dignity, equity, curiosity, and the right to understand. A student-centric XAI framework represents not just a technical intervention, but a philosophical commitment to place learners at the heart of innovation. As we move forward into an increasingly data-driven educational future, we must ask not only *what AI can do*, but *what it should do*—and for whom. By embracing transparency, fostering accountability, and designing with empathy, we can build AI systems that do not obscure learning, but illuminate it—creating inclusive, empowering, and trustworthy environments for every learner.

References

1. Adadi, A., & Berrada, M. (2018). *Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)*. IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
2. Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). *Machine learning interpretability: A survey on methods and metrics*. Electronics, 8(8), 832. https://doi.org/10.3390/electronics8080832
3. Chen, G., Cheng, W., Sun, Y., & Yang, S. J. H. (2020). *Teaching analytics: Towards automatic teaching evaluation using learning analytics*. Computers in Human Behavior, 107, 105868. https://doi.org/10.1016/j.chb.2020.105868
4. Cowgill, B., Dell'Acqua, F., & Deng, S. (2021). *Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics*. Columbia Business School Research Paper. https://doi.org/10.2139/ssrn.3614776
5. Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv preprint arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608
6. European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. COM(2021) 206 final.
7. Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). *AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations*. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
8. Guidotti, R., Monreale, A., Ruggieri, S., et al. (2018). *A survey of methods for explaining black box models*. ACM Computing Surveys, 51(5), 93. https://doi.org/10.1145/3236009
9. Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
10. 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

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 02 Issue No: 02 (2025)

2019 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3290605.3300830

11. Kaplan, R., Chen, Z., & D'Mello, S. (2022). *Contextual and affect-aware feedback improves learning in real-time tutoring systems*. Computers & Education, 184, 104526. https://doi.org/10.1016/j.compedu.2022.104526
12. Kerr, B., & Chung, G. (2021). *Designing for equity in adaptive learning systems*. Computers in Human Behavior Reports, 4, 100114. https://doi.org/10.1016/j.chbr.2021.100114
13. Khosravi, H., Kitto, K., & Buckingham Shum, S. (2022). *Personalised learning analytics explainers: Towards trust and transparency in education*. British Journal of Educational Technology, 53(1), 10–28. https://doi.org/10.1111/bjet.13153
14. Kulesza, T., Burnett, M., Wong, W. K., & Stumpf, S. (2015). *Principles of explanatory debugging to personalize interactive machine learning*. In Proceedings of IUI '15. https://doi.org/10.1145/2678025.2701399
15. Liu, R., Holstein, K., Aleven, V., & Rummel, N. (2021). *Designing explainable AI interfaces for education: Insights from learning sciences*. International Journal of Artificial Intelligence in Education, 31(3), 476–517. https://doi.org/10.1007/s40593-020-00227-6
16. Luckin, R. (2022). *Aligning AI development with human values in education*. AI & Society, 37(1), 75–84. https://doi.org/10.1007/s00146-021-01162-4
17. 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
18. Pane, J. F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2017). *Informing Progress: Insights on Personalized Learning Implementation and Effects*. RAND Corporation. https://doi.org/10.7249/RR2042
19. Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., & Jordan, P. (2018). *Socially adaptive learning technologies*. In R. Sottilare & A. Graesser (Eds.), Design Recommendations for Intelligent Tutoring Systems (Vol. 6, pp. 23–34). Army Research Lab.
20. Sambasivan, N., Kapania, S., Highfill, H., et al. (2021). *"Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI*. In Proceedings of the 2021 CHI Conference on Human Factors. https://doi.org/10.1145/3411764.3445518
21. Scherer, M. U., Zhang, H., & Müller, V. C. (2021). *Ethics of AI in education: Towards a student-rights based framework*. AI & Society, 36(3), 765–777. https://doi.org/10.1007/s00146-020-01050-0
22. Selwyn, N. (2020). *Datafication of education: A critical approach to emerging analytics and AI*. Learning, Media and Technology, 45(1), 1–14. https://doi.org/10.1080/17439884.2020.1694944
23. Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). *Cognitive architecture and instructional design: 20 years later*. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
24. Wang, Y., Xin, T., & Heffernan, N. (2019). *Using student behavior patterns to improve prediction of performance and overall satisfaction in adaptive learning systems*. In Proceedings of the 12th International Conference on Educational Data Mining (EDM).
25. Yudelson, M., Koedinger, K. R., & Gordon, G. J. (2017). *Individualized Bayesian Knowledge Tracing Models*. In Proceedings of the 10th International Conference on Educational Data Mining (EDM).