ICAIMT: Explainable AI for Learning in Higher Education

Meekah Ludba Abdul Hakeem
Dept. of Statistics and Business Analytics, College of Business UAE University
Al Ain-UAE
700050595@uaeu.ac.ae

Nouf Mohammed Hasan Alobeidli
Dept. of Statistics and Business Analytics, College of Business UAE University
Al Ain-UAE
201600181@uaeu.ac.ae

Dr Ananth Chiravuri Associate Prof
Dept. of Statistics and Business Analytics, College of Business UAE University
Al Ain-UAE
ananth.chiravuri@uaeu.ac.ae

https://doi.org/10.63962/VTNO9625

***Abstract*— The objective of this study is to examine the impact of explainable AI (XAI) on student learning in higher education. Drawing on Social cognitive and Flow theories, we propose that XAI enhances trust, engagement, self-efficacy and flow experience. Findings from our survey indicate that trust in XAI significantly predicts engagement, self-efficacy and flow. Findings suggest that transparency in AI systems plays a critical role in shaping positive behavioural outcomes. Implications for higher educational institutions and AI developers are discussed.**

Keywords—Explainable Artificial Intelligence (XAI), Trust, Self-Efficacy, Engagement, Flow Experience, Higher Educations.

I. INTRODUCTION

Artificial Intelligence (AI) and its related applications have rapidly gained prominence within higher education institutions (HEI's), fundamentally transforming how students access information, interact with educational resources, and engage in learning processes. The integration of Generative AI (GenAI) tools and intelligent learning applications such as ChatGPT, Co-Pilot, Gemini has enabled increasingly personalized, adaptive, and efficient learning experiences for students, while supporting instructors through automated feedback, content generation, and learning analytics [1]. As a result, AI-driven tools and systems are increasingly being embedded in teaching practices, assessment methods and pedagogical models.

Despite these advantages, the accelerated development and widespread adoption of AI-based tools in HEI's have raised significant concerns, particularly relating to transparency, accountability, and student trust [2]. Many AI systems operate using complex algorithms and machine learning models that are not easily interpretable by users. This so-called "black box" nature of AI presents a critical challenge in educational contexts, where understanding the rationale behind decisions, recommendations, or feedback is essential for meaningful learning and developing insight. When students and other users (e.g., instructors, teachers etc) are unable to comprehend how an AI system arrives at its outputs such as grading suggestions, learning recommendations, or academic guidance, they may question the system's reliability and fairness, leading to reduced trust and engagement.

Thus, it is possible that a lack of transparency in AI systems can negatively influence students' learning experiences by diminishing confidence in AI-supported educational tools and potentially discouraging their effective use. Moreover, opaque decision-making processes may reinforce biases, limit opportunities for critical reflection, and undermine the pedagogical value of AI tools and technologies [3]. In environments that prioritise academic integrity, fairness, and learner autonomy, such limitations pose serious risks to educational outcomes and institutional credibility.

In response to these challenges, Explainable Artificial Intelligence (XAI) has emerged as a promising approach to enhance transparency and trust in AI-driven educational systems [4]. XAI focuses on making AI models more interpretable by providing clear, understandable explanations for their outputs and recommendations [5]. This enables students and educators to grasp the "how" and "why" behind AI generated decisions, thereby promoting more informed, confident and immersive learning experiences [6][7]. Consequently, the integration of explainability into AI-based educational tools represents a critical step forward in responsible and ethical AI adoption in higher education.

While XAI offers advantages over traditional black box systems, empirical research examining its educational impact remains limited. Most existing studies focus primarily on the adoption and usage of AI tools by students, whereas little is known about the specific effects of explainability on student learning outcomes [8] [9]. Accordingly, there is a need to systematically examine how XAI influences learning related concepts, forming the motivation for this study. Although many people are excited about AI as a technology including its tools in higher education, there is still insufficient understanding of whether and how XAI enhances students' learning experiences [10]. Most prior studies emphasize the technical side or focus on isolated outcomes, like trust or academic performance. Presently, very few studies have developed and empirically tested an integrated and inclusive model that examines multiple student psychological and behavioural dimensions of learning.

Although some studies have explored students' perceptions of AI in educational contexts, limited attention has been given

to how "explainability" affects interconnected learning variables such as trust, engagement, self-efficacy, and flow in a unified model. This is important because these variables are theoretically interrelated and collectively shape meaningful learning experiences. Therefore, understanding their interaction can help provide deeper insight into the mechanisms through which XAI may influence HE learning outcomes. Next, we present the literature review.

II. LITERATURE REVIEW

A. Explainable Artificial Intelligence (XAI) and Learning

Explainable AI refers to AI systems that are designed to make their decision making processes transparent and interpretable to human users. Normal "black-box" AI gives answers without explaining how it got them, but XAI shows the reasons, the data, and the steps used to make each choice. This transparency reduces information asymmetry between humans and algorithms, thereby allowing users to evaluate the logic behind the suggestions and/or recommendations.

In HEI's, this is very important because HEI students need to know not just *what* a concept means, but also develop a critical understanding as to *why* and *how* these concepts work. This is justified because Bloom's taxonomy (1956), indicates that students in HEI's need to focus more applying, analyzing, synthesizing and evaluating concepts rather than just knowledge and comprehension as done earlier [11]. Thus, explainability aligns with higher order cognitive processes required in HEI's where reasoning is critical.

Some scholars have explored XAI in institutions. They found 15 different meanings of XAI and 62 problems, grounded into seven areas: making things clear, being fair, computer parts, how people and computers work together, trust, rules, and other issues [12]. But their study noted that no one agreed to one meaning of XAI, which makes it confusing and problematic to get a consensus on. This conceptual fragmentation limits cumulate knowledge development. But even with this problem, most agree that XAI is very important for making AI easier to use and more trusted in learning. In particular, transparency has been associated with greater perceptions of reliability, accountability and fairness in algorithmic systems [13].

Some of them looked at different ways to make AI explain itself in institutions. For example, they examined tools like LIME and SHAP, which help explain any AI model, and simple models like decision trees that are easy to understand by themselves. They found that how well these tools work depends on where they are used, how much people know about technology, and what learning goals the teacher or student wants to reach. More importantly, prior research suggests that the effectiveness of explanations is contingent upon user characteristics such as knowledge, cognitive ability thereby implying that explainability must be context sensitive rather than purely technical [14]. Therefore, in HEI's, XAI should be designed and implemented to maximize its impact on trust, engagement and learning outcomes.

B. Social Cognitive Theory and AI-supported Learning

The Social Cognitive Theory (SCT) provides important theoretical foundations for understanding how students

interact with technology-supported learning environments. According to the previous papers [15], a central construct within this theory is self-efficacy, defined as an individual's belief in their ability to perform tasks successfully. Students with stronger self-efficacy beliefs tend to demonstrate higher levels of effort, resilience, and participation in academic tasks [16]. In digital learning environments, technological systems can function as environmental factors that influence learners' behaviours and beliefs by providing feedback, guidance, and learning support [17].

The transparency in XAI Learning can strengthen the students' confidence by using AI tools effectively, thereby improving the perceived self-efficacy. In addition to this, the explainable systems can promote more active learner participation by enabling students to critically evaluate AI-generated suggestions and integrate them into their learning strategies. Prior research suggests that transparent and interpretable learning technologies can enhance learner motivation and engagement by increasing the perceived control and understanding of the learning process [18][19]. Therefore, from a social cognitive perspective, the explainable AI learning may positively influence students' self-efficacy and engagement in AI-supported learning environments.

C. Flow Theory and AI-supported Learning

The Flow theory [7] suggests that the flow occurs only when the learners perceive a balance between the challenge of a task and their own skills, with clear goals and immediate feedback. In educational contexts, the experience of flow has been associated with increased motivation, engagement, and improved learning performance. Research in the digital learning environments suggests that interactive technologies and well-designed learning systems can facilitate flow experience by providing structured guidance, continuous feedback, and meaningful interaction with learning tasks [20][21].

The transparency in the XAI enables the learners to focus more effectively on the task itself rather than questioning the system's recommendations. In addition to that, the trust in AI systems may further support the development of the flow experiences by allowing students to rely on the AI-supported guidance without any hesitation. Prior research indicates that the digital learning environments that provide clear feedback and interactive engagement can enhance learners' immersion and intrinsic motivation [22][23]. Therefore, explainable AI and trust in the AI systems may facilitate flow experiences by enabling sustained attention, deeper engagement, and more immersive learning interactions in AI-supported educational environments.

Based on the relevant theories, we argue and develop the following hypotheses. As discussed, XAI leads to more transparency, predictability and increased explainability, which in turn increases the reliance of the users on such systems, leading to greater trust in the AI systems. Therefore, we posit:

H1:Perceived explainable AI positively predicts students' trust in AI recommendations.

In addition, we hypothesize that perceived XAI could directly impact the three learning outcomes because of the underlying mechanisms of better structured guidance and explicit reasoning and test for the below:

H2a: Perceived explainable AI positively predicts students' self-efficacy in using AI for learning.

H2b: Perceived explainable AI positively predicts students' learning engagement.

H2c: Perceived explainable AI positively predicts students' flow experience during AI-supported learning.

Similarly, we posit greater trust will directly impact learning outcomes as well because when students trust XAI systems, they are more willing to rely on recommendations and strengthen their belief systems, and hypothesize:

H3a: Trust in explainable AI positively predicts students' self-efficacy.

H3b: Trust in explainable AI positively predicts students' learning engagement.

H3c: Trust in explainable AI positively predicts students' flow experience.

Finally, as indicated earlier, XAI provides transparency, which fosters trust by increasing perceptions of reliability and competence. Thus, we hypothesize trust to mediate the relationship between XAI and the learning outcomes and posit

H4a: Trust mediates the relationship between perceived explainable AI and students' self-efficacy.

H4b: Trust mediates the relationship between perceived explainable AI and students' learning engagement.

H4c: Trust mediates the relationship between perceived explainable AI and students' flow experience.

III. METHODOLOGY

In this section, we describe the methodology that was followed to conduct this study.

A. Research Design

This study adopted a quantitative analysis using a survey that included a questionnaire . This approach was appropriate for looking at the relationships between perceived explainability of AI (independent variable) and other variables measuring learning (dependent variables). These dependent variables include Trust (mediator variable), Engagement, Self-Efficacy, Flow Experience. We used a survey method as it enabled efficient collection of data from many students, thereby supporting the statistical examination of the proposed relationships between XAI, trust and the dependent variables. We administered the survey online to facilitate faster data collection.

B. Data Collection

The survey was administered online using Google forms, ensuring accessibility and enabling efficient data collection. The survey link was distributed to students in HEI's through multiple

channels such as direct institutional email, announcements in learning management systems (Blackboard), social media platforms used by UAEU students, and in-class announcements by cooperative faculty members. The survey began with an informed consent statement explaining the research purpose and procedures to respondents including voluntary participation, confidentiality protections, right to withdraw, and researcher contact information. Participants indicated consent by proceeding with the survey after reviewing this information. Data collection occurred over four weeks during the academic semester, avoiding examination periods to ensure adequate response rates and minimize stress-related response bias. Till date, a total of 79 usable responses were obtained. We analyzed the data using regression as it was suitable to test for direct effects and mediation pathways.

IV. INITIAL FINDINGS

Our initial findings revealed that perceived explainability of AI significantly predicts trust in AI recommendations ($\beta = .725$, $p < .001$), supporting H1. Our findings also indicated that XAI directly predicts self-efficacy and flow. More importantly, these effects remained significant even after accounting for trust, which suggests that explainability may enhance students' sense of competence and immersion through clarity, guidance, and comprehensibility of explanations (i.e., the explanation itself reduces uncertainty and increases perceived capability), rather than via trust.

We also found that trust significantly predicted student

consider trust building mechanisms as a core design principle. Future education in AI depends not only on technological advancement but also on the integration of meaningful and context driven explanations.

When a student understands the AI's advice, their perceived self-efficacy in using AI systems goes up leading to improved learning experiences which then could create a state of flow. Transparency functions as a mechanism to build trust that enables higher engagement, greater confidence and deeper learning states [25].

A limitation of our study is that it mainly focuses on four areas: Trust, Engagement, Self-efficacy, Flow experience. Future studies can investigate other dimensions across different countries to assess cultural variability and enhance external validity. To conclude, our study indicates that XAI and students' trust in AI systems plays a central mediating role in shaping students' learning outcomes in different ways thereby supporting a more meaningful and immersive learning experience.

REFERENCES

[1] Schepman, A., & Rodway, P. (2022). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human-Computer Interaction*, *39*(13), 2724-2741.

[2] European Commission. (2019). *Ethics guidelines for trustworthy AI*. High-Level Expert Group on Artificial Intelligence, European Union.

[3] Dawson, P., Tai, J., Bearman, M., Boud, D., & Ajjawi, R. (2023). The development and validation of the assessment engagement scale. *Frontiers in Psychology*, *14*, 1136878.

engagement in their studies, but not their confidence in using AI supported technologies, and their ability to remain more deeply focused on their learning [24]. XAI increases engagement because, as seen in H1, it increases trust. Students invest time, stay active, and remain focused when they perceive recommendations as trustworthy.

Finally, our findings indicate that trust mediates the relationship between perceived XAI and engagement. However, trust did not function as a universal psychological mechanism across all outcomes, as no mediation effects were observed for self-efficacy or flow. This pattern suggests that trust plays a selective role in translating explainability into behavioural engagement, whereas cognitive and experiential outcomes may stem more directly from the clarity provided by explanations themselves.

V. CONCLUSION AND LIMITATIONS

Our findings are strategically important for universities and companies that make educational technology. They suggest that XAI is not merely a nice feature but a foundational feature that determines whether AI tools are adopted, trusted and effectively utilized in HEI's.

The improvement of student learning appears to be facilitated when students exhibit higher levels of trust in the learning platform that they use. Our study indicates that organizations developing AI based educational platforms should prioritize transparency and explainability to impact the learning outcomes and enhance user trust, which in turn may strengthen user engagement. Platform evelopers shoul

[4] 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*, 39.

[5] Gunasekara, S., & Saarela, M. (2025). Explainable AI in education: Techniques and qualitative assessment. *Applied Sciences*, *15*(3), 1239.

[6] Csikszentmihalyi, M. (1975). Beyond boredom and anxiety. Jossey-Bass.

[7] Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. Harper & Row.

[8] Pearce, J. M., Ainley, M., & Howard, S. (2005). The ebb and flow of online learning. *Computers in Human Behavior, 21*(5), 745-771.

[9] Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, *29*(4), 4785-4808.

[10] Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.*

[11] B. S. (1956b) Taxonomy of Educational Objectives; the Classification of Educational Goals, by a Committee of College and University Examiners., David McKay, New York, NY.

[12] Altukhi, A., & Pradhan, S. (2024). Systematic literature review: Explainable AI definitions and challenges in education. *arXiv preprint* arXiv:2504.02910.

[13] Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International journal of human-computer studies*, 146, 102551.

[14] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence, 267*, 1-38.

[15] Arastaman, G. (2014). Validity and reliability of student engagement scale. *Bartin University Journal of Faculty of Education, 3*(2), 390-403.

[16] Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

[17] Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

[18] Wang, Y., et al. (2023). Artificial intelligence in education and learner self-efficacy: Implications for engagement and learning outcomes. Computers in Human Behavior, 139, 107672. DOI: 10.1016/j.chb.2023.107672

[19] DiBenedetto, M. K., & Zimmerman, B. J. (2018). Self-regulation and social cognitive theory in education. Educational Psychologist, 53(1), 35–50. DOI: 10.1080/00461520.2018.1465313

[20] Rodríguez-Ardura, I., & Meseguer-Artola, A. (2017). Flow in e-learning environments: Direct and indirect effects on learning outcomes. Computers in Human Behavior, 72, 699–706. DOI: 10.1016/j.chb.2017.01.039

[21] Buil, I., Catalán, S., & Martínez, E. (2020). Understanding flow in digital learning environments. Computers & Education, 149, 103821. DOI: 10.1016/j.compedu.2020.103821

[22] Kiili, K., et al. (2021). Flow experience and engagement in digital learning environments. Educational Technology Research and Development, 69(1), 361–385. DOI: 10.1007/s11423-020-09864-8

[23] Engeser, S., & Rheinberg, F. (2018). Flow, performance and motivation. Motivation and Emotion, 42(5), 1–15. DOI: 10.1007/s11031-018-9696-2

[24] Arastaman, G. (2014). Validity and reliability of student engagement scale. *Bartin University Journal of Faculty of Education, 3*(2), 390-403.

[25] Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.