Operationalizing a Human-Centered Explainable AI Framework for Adaptive, Trust-Aware Learning Analytics

Dr. Ravi Soni[1], Dr. Sharad Salunke[2]
[1-2] Lincoln University College, 47301 Petaling Jaya, Selangor Darul Ehsan, Malaysia;
Email ID: pdf.ravi@lincoln.edu.my, pdf.sharad@lincoln.edu.my

Abstract: Incorporating Artificial Intelligence in Learning Analytics, it is being increasingly utilized to forecast student performance, pinpoint students at risk and to aid educational decision making. Explainable Artificial Intelligence can help create more transparent AI generated predictions by providing reasons for the prediction. This explanation is not necessarily appropriate for every educational stakeholder as they all have different roles, levels of trust in AI and information-processing needs. This work further extends an existing proposed Human-Centered Explainable AI (HXAI) framework and operationalizes the main constructs and an adaptive mechanism for providing explanations in Learning Analytics. We introduce explainability, cognitive load, trust, educational role and effectiveness of decision support into the proposed framework. An Adaptive Explanation Engine takes these factors into consideration when deciding on an appropriate Explanation Delivery Mode (Summary, Moderate, or Detailed). The paper describes the operation of each construct, presents initial adaptation guidelines, and summarizes a DSR-based progression toward prototype development and subsequent empirical evaluation. The study offers an implementation-driven approach to adaptive explanation delivery in Learning Analytics that follows the principles of Human-Centered XAI.

Keywords: Explainable Artificial Intelligence, Human-Centered AI, Learning Analytics, Adaptive Explanation, Cognitive Load, Trust, Educational Decision Support.

Introduction

With the advent of Artificial Intelligence, Learning Analytics has come to play a greater role in supporting student performance prediction, identifying students at risk, analyzing students' learning behaviors and making decisions on education. They can process vast amounts of education-related information but the complex algorithms can be hard for instructors and students to comprehend. Explainable AI can solve this problem by providing explanations for the factors and reasoning behind the output of the model. But, in education, there may be more too technical explainability. Khosravi et al. [1] highlighted the importance of stakeholders, how to present explanations, human-centered interfaces, and the educational context in XAI. In our previous systematic review of 108 studies, we found some gaps in LCAE, especially related to stakeholder responsibilities, the role of humans, safety, reliability and trustworthiness [2].

Empirical studies also have shed light on the significance of the human factors, as seen in recent research. Wang et al. [3] found no substantial difference in cognitive load between teachers who received explanations and those who did not. However, teachers who received explanations showed higher levels of trust and technology acceptance. Feldman-Maggor et al. [4] spoke about explainability, trust and acceptance by teachers and emphasized the importance of domain-specific explanations. Jeyasothy et al. [5] demonstrated the customization of post-hoc explanations based on user properties and Saqr and López-Pernas [6] demonstrated the benefit of individual explanations for LA and highlighted that AI explanations are not a substitute for human judgment.

The findings from these studies indicate that one could separate the process of explanation generation and explanation delivery. The first part of this research has presented a conceptual model for Human-Centered Explainable AI (HXAI) framework for Learning Analytics. The second phase concerns operationalization of that framework which is why the present study focuses on this aspect. The goal is not to create a new prediction algorithm, but to modify an existing algorithm of an AI prediction and its explanation based on user related conditions.

Related work

Educational XAI, human-centered LA, personalized explanations, trust and cognitive load have been explored in existing research. These studies offer significant foundations but not directly indicate how the quantity and kind of an explanation should be tailored to the situation of an educational user. Therefore, the implementation-oriented delivery mechanism is the focus of this study, which includes cognitive load, trust and educational role.

Study	Educational XAI	User factors	Adaptive delivery
Khosravi et al. [1]	Yes	Stakeholders	Not specified
Alfredo et al. [2]	Yes	Human-centered	Not specified
Wang et al. [3]	Yes	Trust, cognitive load	No
Jeyasothy et al. [5]	General XAI	User knowledge	Personalized
This work	Yes	Load, trust, role	Yes

Table 1. Comparison of the present framework with selected related work.

Key Contribution

The main contributions of the operational framework are

- Operational definitions of explainability, cognitive load, trust, educational role, and decision-support effectiveness.
- An Adaptive Explanation Engine for Learning Analytics.
- Three Explanation Delivery Modes: Summary, Moderate, and Detailed.
- Initial rule-based adaptation guidelines: mapping of user conditions and explanation delivery modes.
- A Design Science Research (DSR) pathway for prototype development and empirical validation.

Method, Proposed Framework and Results

This section outlines the Method, Proposed Framework and Results. The concept for the operational architecture is shown in Fig. 1. Relevant learning information is delivered to the system from LMS, assessment and student interaction mechanisms. Using an AI model, a prediction, classification, risk assessment or a recommendation is made. The framework is not dependent on a particular machine-learning algorithm and it is model agnostic. An XAI method is then used to generate information to explain the AI output, either by means of feature attribution, interpretable modeling, rules or otherwise as appropriate. The Adaptive Explanation Engine takes information related to the user and determines the type of explanation to be given. The selected explanation is shown to the user with a Human-Centered interface and is one component of the user's educational decision-making process.

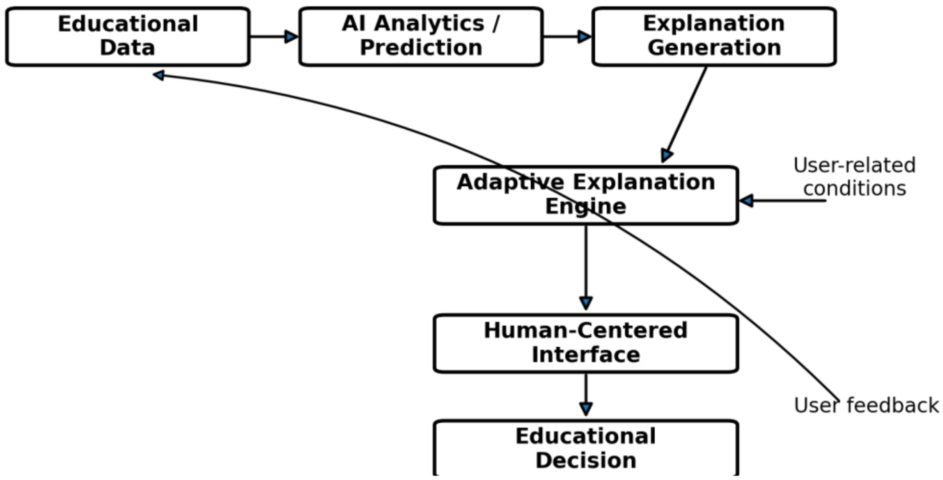

Figure 1. Operational Human-Centered XAI Architecture.

Cognitive load is added on because in learning new information, working memory is limited. Therefore, from the perspective of Cognitive Load Theory [7] the complexity of the explanation can be considered. The framework doesn't state that it is always necessary to reduce cognitive load. It considers cognitive load as one of the factors that may impact on the amount of information presented. The inclusion of trust is because factors such as user, system, and situational relate to willingness to rely on automated systems [8] and recent studies in educational XAI demonstrate that explanations have an impact on teachers' trust and acceptance [3] and [4].

The Adaptive Explanation Engine is the main processing component. It has 3 major user related inputs, which are cognitive load, trust and educational role. It next selects an Explanation Delivery Mode that is suitable for the situation. The decision mechanism proposed is depicted in Figure 2.

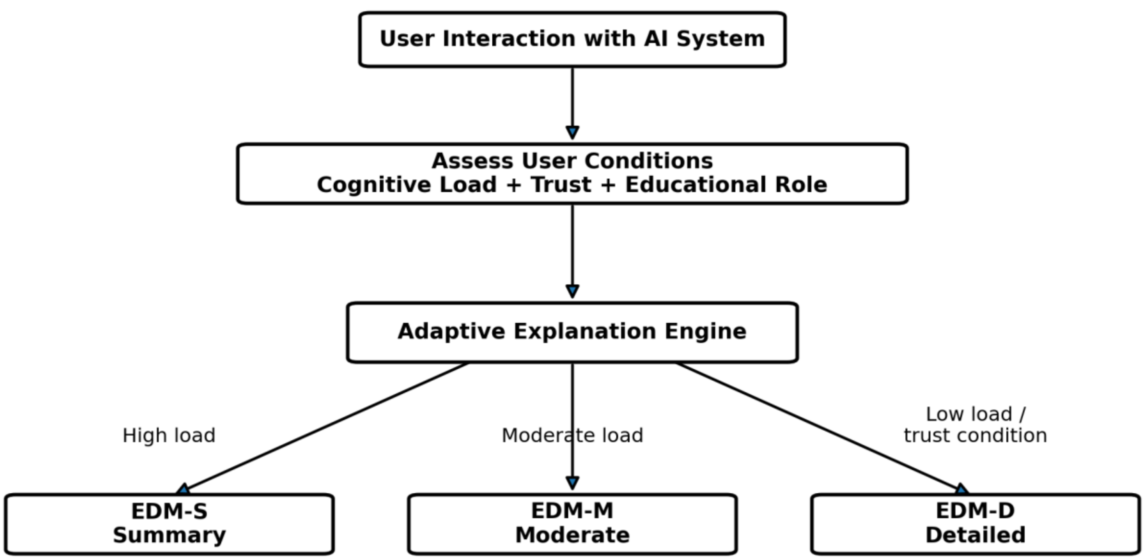

Figure 2. Adaptive Explanation Decision Mechanism.

The framework implements the key constructs with an Adaptive Explanation Delivery Mechanism with three Explanation Delivery Modes.

The operational framework considers five constructs: explainability, cognitive load, trust, educational role, and decision-support effectiveness. These constructs are connected through the adaptive explanation delivery mechanism. Explainability concerns the understandability of AI outputs. Cognitive load represents the user's information-processing demands. Trust reflects the user's willingness to rely on AI-supported recommendations. Educational role captures differences among stakeholder groups. Decision-support effectiveness refers to the extent to which an AI-supported explanation helps a user interpret information and make an informed educational decision.

- **EDM-S:** Summary Mode gives the prediction, key reason and summary implication/recommendation. It is used in situations where the amount of information to process is high or when quick interpretation of the information is needed.
- **EDM-M:** Moderate Mode will give the prediction, main contributing factors, supporting evidence, and will include some contextual information.
- **EDM-D:** For users who need further analysis details the Detailed Mode is available for them, which offers contributing factors, evidence, context and further details of analysis.

Initial Adaptation Rules

User condition	Explanation mode	Rationale
High cognitive load	EDM-S	Reduce information-processing demand
Moderate cognitive load	EDM-M	Provide balanced information
Low cognitive load + high trust	EDM-D	Support deeper analytical interpretation
Low cognitive load + low trust	EDM-M	Provide supporting information without excessive complexity
Role-specific requirement	Role-adapted content	Maintain relevance

Table 3. Initial adaptation rules.

These are not threshold limits, but design rules. Thresholds for cognitive load and trust will be developed in the prototype process and field testing.

Methodology and Planned Validation

Design Science Research is used because the present study produces an operational artifact intended to address the problem of adaptive explanation delivery [9]. The DSR process described by Peffers et al. [10] provides a structured sequence covering problem identification, objective definition, artifact development, demonstration, evaluation, and communication. The research programme is thus broken down in four phases: Phase I is a systematic review and conceptual framework; Phase II is an operational framework; Phase III is prototype development; and Phase IV is empirical validation. The first phase has been completed, and the present study addresses the second phase, namely the operationalization of the conceptual framework.

The proposed research model links explainability, user understanding, trust, and decision-support effectiveness with user-related conditions such as cognitive load and educational role. It separates the content and understandability of an AI explanation from the context in which the explanation is presented. Adaptive Explanation Delivery acts as the mechanism for selecting the appropriate level and form of explanation according to these user-related conditions. The following questions help to follow up with validation:

- **RQ1:** What are the ways of operationalizing explainability, cognitive load, trust, educational role and effectiveness of decision support in the context of Learning Analytics?
- **RQ2:** What are the ways cognitive load, trust and educational roles can be incorporated into an adaptive explanation-delivery mechanism?
- **RQ3:** Can explanation delivery using adaptive techniques enhance user understanding, trust, and decision support effectiveness over static explanation delivery?
- **RQ4:** Are there specific explanation characteristics required to facilitate effective decision making in different educational roles when using AI?

The following comparisons will be made when evaluating the planned examples: static explanation (using the same explanation structure for all examples) and adaptive explanation (using the explanation mode proposed by the mechanism). Explanation understandability, cognitive load, trust, usability, decision confidence, and decision-support effectiveness are all going to be measured by evaluation measures. There are no empirical results because the prototype is still being developed.

Discussions

The proposed framework is about the generation of an XAI explanation and the delivery of an XAI explanation to a human user. Recent research in education has demonstrated the ability of explanations to promote transparency and trust in education (XAI). Shared explanations were found to foster more teacher trust and acceptance of technology [3] and individualized explanations were found to be valuable in understanding student predictions without losing the need for human judgment [6]. The present framework introduces a layer of explicit adaptation in between the generation of an explanation and its presentation. It thus not only asks the question about whether an explanation can be produced, but what to explain or which explanation is suitable for a given user in each situation. At this stage, there are no empirical thresholds of trust or of cognitive load in the framework. Rather, those factors are seen as the user-related conditions that can affect the selection of a proper delivery mode of an explanation. They will be analyzed in the context of developing the prototype and empirical validation. The usefulness and relationships will be tested through an empirical test in the next phases.

Conclusions

- The study aims at addressing the need for an implementation-oriented mechanism for delivering adaptive explanations in Learning Analytics.
- Operationalization: Five constructs are operationalized in the framework and the approach is based on a Design Science Research path.
- The Adaptive Explanation Engine links three Explanation Delivery Modes, cognitive load, trust and educational role.
- Limits and future research: The proposed framework has not been tested with a functional prototype or an empirical user study has not been conducted. The cognitive load and trust thresholds, and the effectiveness of the adaptation rules will therefore be determined and validated in the next prototype and validation steps.

References

[1] H. Khosravi et al., “Explainable artificial intelligence in education,” Computers and Education: Artificial Intelligence, vol. 3, Art. no. 100074, 2022, doi: 10.1016/j.caeai.2022.100074.

[2] R. Alfredo, V. Echeverria, Y. Jin, L. Yan, Z. Swiecki, D. Gašević, and R. Martinez-Maldonado, “Human-centred learning analytics and AI in education: A systematic literature review,” Computers and Education: Artificial Intelligence, vol. 6, Art. no. 100215, 2024, doi: 10.1016/j.caeai.2024.100215.

[3] D. Wang, C. Bian, and G. Chen, “Using explainable AI to unravel classroom dialogue analysis: Effects of explanations on teachers’ trust, technology acceptance and cognitive load,” British Journal of Educational Technology, vol. 55, no. 6, pp. 2530–2556, 2024, doi: 10.1111/bjet.13466.
[4] Y. Feldman-Maggor, M. Cukurova, C. Kent, and G. Alexandron, “The impact of explainable AI on teachers’ trust and acceptance of AI EdTech recommendations: The power of domain-specific explanations,” International Journal of Artificial Intelligence in Education, vol. 35, pp. 2889–2922, 2025, doi: 10.1007/s40593-025-00486-6.
[5] A. Jeyasothy, T. Laugel, M.-J. Lesot, C. Marsala, and M. Detyniecki, “A general framework for personalising post hoc explanations through user knowledge integration,” International Journal of Approximate Reasoning, vol. 160, Art. no. 108944, 2023, doi: 10.1016/j.ijar.2023.108944.
[6] M. Saqr and S. López-Pernas, “Why explainable AI may not be enough: Predictions and mispredictions in decision making in education,” Smart Learning Environments, vol. 11, Art. no. 52, 2024, doi: 10.1186/s40561-024-00343-4.
[7] J. Sweller, J. J. G. van Merriënboer, and F. Paas, “Cognitive architecture and instructional design: 20 years later,” Educational Psychology Review, vol. 31, no. 2, pp. 261–292, 2019, doi: 10.1007/s10648-019-09465-5.
[8] K. A. Hoff and M. Bashir, “Trust in automation: Integrating empirical evidence on factors that influence trust,” Human Factors, vol. 57, no. 3, pp. 407–434, 2015, doi: 10.1177/0018720814547570.
[9] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” MIS Quarterly, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.
[10] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” Journal of Management Information Systems, vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.