

Evaluating the Usability of an RF-LSTM Learning Platform with Explainable AI among University Students

Diah Aryani^{1✉}, Jefry Sunupurwa Asri², Habibullah Akbar³,
Muhammad Fikri⁴, Bayu Sulistiyanto Ipung Sutejo⁵
(1,2,3,4,5) Universitas Esa Unggul, Jakarta, Indonesia

✉ Corresponding author
diah.aryani@esaunggul.ac.id

Received: 28-05-2026 | Revised: 15-08-2026 | Accepted: 28-08-2026

ABSTRACT

Educational AI platforms require usable interfaces that enable students to interpret complex predictive and explainability outputs. This study evaluated the usability of an RF-LSTM learning platform incorporating explainable artificial intelligence among university students. A multidimensional usability evaluation involved 40 Informatics Engineering and Computer Science students who completed six structured tasks covering data exploration, preprocessing, model prediction, performance comparison, SHAP interpretation, and navigation. Usability evidence was obtained from task performance, structured observations, the System Usability Scale (SUS), and open-ended feedback. Task completion ranged from 85% to 100%, while the mean SUS score of 76.7 indicated positive overall usability. Greater difficulties emerged when students compared performance metrics and interpreted SHAP visualizations, particularly regarding technical terminology, visual presentation, navigation, and system feedback. The findings suggest that algorithmic explainability does not automatically ensure comprehensibility during student interaction. Contextual explanations, clearer navigation cues, scalable SHAP visualizations, and processing feedback emerged as design priorities, although the findings remain limited to students within a higher education context.

Platform pendidikan berbasis AI memerlukan antarmuka yang mudah digunakan agar mahasiswa mampu menginterpretasikan keluaran prediktif dan explainability yang kompleks. Penelitian ini mengevaluasi usability platform pembelajaran RF-LSTM yang mengintegrasikan explainable artificial intelligence pada mahasiswa. Evaluasi usability multidimensional melibatkan 40 mahasiswa Teknik Informatika dan Ilmu Komputer yang menyelesaikan enam tugas terstruktur meliputi eksplorasi data, preprocessing, prediksi model, perbandingan performa, interpretasi SHAP, dan navigasi. Bukti usability diperoleh melalui performa tugas, observasi terstruktur, System Usability Scale (SUS), dan umpan balik terbuka. Tingkat penyelesaian tugas berkisar antara 85% hingga 100%, sedangkan rerata skor SUS sebesar 76,7 menunjukkan usability keseluruhan yang positif. Kesulitan lebih terlihat ketika mahasiswa membandingkan metrik performa dan menginterpretasikan visualisasi SHAP, terutama terkait terminologi teknis, penyajian visual, navigasi, dan umpan balik sistem. Temuan menunjukkan bahwa explainability algoritmik tidak secara otomatis menjamin keterpahaman dalam interaksi mahasiswa. Penjelasan kontekstual, penanda navigasi yang lebih jelas, visualisasi SHAP yang dapat disesuaikan, dan umpan balik pemrosesan menjadi prioritas pengembangan, meskipun temuan terbatas pada mahasiswa dalam konteks perguruan tinggi.

Keywords: explainable AI, learning platform, RF-LSTM, usability

Introduction

Machine learning has become increasingly relevant to higher education, particularly in Informatics Engineering and Computer Science, where students are expected not only to understand computational principles but also to engage directly with data-driven processes (Long & Magerko, 2020; Shapiro et al., 2018). Learning in this area increasingly involves a sequence of activities that extends from data preparation and model development to

performance evaluation and interpretation of computational outputs. Within educational contexts, machine learning has been employed to support prediction, feedback, and decision-making; however, the educational value of these systems cannot be determined solely by their computational capabilities. Students must also be able to locate relevant information, follow analytical processes, interpret model outputs, and connect those outputs with the task they are attempting to complete. Consequently, the interaction between students and machine-learning systems represents an important educational concern because technically valid information may have limited instructional value when it cannot be readily understood or used by learners (Ghimire et al., 2024; Raj & Renumol, 2022).

These demands become particularly visible when machine learning is introduced through complex analytical contexts such as renewable energy forecasting. In this setting, students encounter multiple forms of information simultaneously, including input variables, temporal patterns, preprocessing procedures, prediction outputs, and model-performance indicators. Machine learning, deep learning, and hybrid approaches have been applied to energy and environmental data because they can accommodate nonlinear relationships and temporal dependencies that characterize forecasting problems (Maarif et al., 2023; Van Poecke et al., 2024). Hybrid approaches have likewise been explored for photovoltaic power prediction, while explainability techniques can be incorporated to clarify how individual input features contribute to forecasting results (Mubarak et al., 2023). From an educational perspective, such an environment offers students opportunities to examine not only what a model predicts but also how different stages of an analytical workflow are connected. Nevertheless, the increasing technical richness of the platform also creates a pedagogical challenge: the presence of sophisticated modeling capabilities does not automatically mean that students can navigate, understand, and interpret those capabilities effectively.

This distinction between technical functionality and meaningful student interaction becomes particularly important with the incorporation of explainable artificial intelligence (XAI). XAI is intended to make the behavior of complex computational models more transparent, yet the presence of an explanation should not be treated as equivalent to the comprehensibility of that explanation. For students, an explanation must be presented in a form that can be interpreted within their existing knowledge and the learning activity in which they are engaged. Previous research has therefore emphasized that explanations need to be understandable and relevant to users rather than merely technically available (Khosravi, Buckingham Shum, et al., 2022; Saqr & López-Pernas, 2024). When appropriately designed, XAI-based features may also support feedback, reflection, and self-regulated engagement with learning activities (Afzaal et al., 2024). In this sense, explainability in an educational platform is not solely a property of the underlying algorithm; it also concerns how information is represented, accessed, and interpreted by students during their interaction with the system.

One approach to model explanation is SHapley Additive exPlanations (SHAP), which represents the contribution of individual features to model predictions (Lundberg & Lee, 2017). In educational research, SHAP has been used to interpret predictive models and to identify variables associated with student performance (El Marzki, 2025; Ghimire et al., 2024; Guan et al., 2026). Its incorporation into a learning platform can therefore provide students with opportunities to examine the relationship between model inputs and prediction outputs rather than treating predictions as opaque results. However, this potential depends on students being able to interpret the information displayed. Technical terminology, unfamiliar metrics, visually dense explanations, and the organization of information across different stages of a platform

may impose additional demands on users even when the underlying explanation method is technically sound (Bobek et al., 2025; Saqr & López-Pernas, 2024; Subramonyam & Hullman, 2023; Wang et al., 2024). The educational issue is consequently not simply whether SHAP explanations are available, but whether students can interact with those explanations in a way that allows them to complete the intended analytical activities without unnecessary interface-related difficulty.

This concern places usability at the intersection of interface design and students' engagement with technically complex learning activities. A usable learning environment should enable students to navigate between stages, identify relevant information, interpret visual outputs, and receive sufficient feedback without requiring excessive effort to understand how the interface itself operates. When interface organization, terminology, or information presentation introduces unnecessary complexity, students may need to allocate cognitive resources to managing the system rather than to the analytical task being undertaken. Such conditions may contribute to extraneous cognitive load, particularly in digital learning environments where learners must simultaneously process instructional content and interface-related information (Skulmowski & Xu, 2022). Research on student-facing learning analytics dashboards similarly indicates that visual clarity, comprehensibility, perceived usefulness, and usability influence how students interact with digital information environments (De Barba et al., 2022; Park & Jo, 2019b; Rets et al., 2021). Usability should therefore be understood here not merely as interface convenience, but as an important condition for enabling students to engage with the analytical functions made available through a learning platform.

Evaluating such usability requires more than determining whether a platform functions technically or whether users report a generally positive impression. The System Usability Scale (SUS) provides a concise and established measure of perceived usability (Castilla et al., 2024; Lewis, 2018; Vlachogianni & Tselios, 2022), but an overall score cannot by itself reveal which stages of an analytical workflow create difficulty, what kinds of errors occur, or why students encounter problems when interpreting particular outputs. Combining perceived-usability measures with task-based observations and qualitative feedback provides a more informative basis for examining navigation, information presentation, visualization, and users' experiences with digital learning environments (De Vreugd et al., 2024; Ramaswami et al., 2023). This approach is particularly relevant when actual students are the intended users because the educational usefulness of a digital system depends partly on their capacity to interpret and apply the information presented to them (Rets et al., 2021; Susnjak et al., 2022). Evaluating only technical functionality may overlook students' needs, experiences, and agency, whereas direct evaluation with student users can reveal interaction difficulties that may remain invisible during technical validation (De Vreugd et al., 2024; Kaveri et al., 2023; Wang et al., 2024).

Despite the expanding literature on machine learning, XAI, and learning analytics, an important issue remains insufficiently addressed at the intersection of these fields. Previous studies have provided substantial knowledge about predictive performance, explainability, student-performance prediction, and learning analytics dashboards, but these strands are often examined separately. Comparatively less attention has been directed toward how university students interact with an integrated learning workflow that requires them to move through data exploration, preprocessing, RF–LSTM modeling, performance evaluation, and SHAP-based interpretation within a single platform. This distinction is important because the existence of predictive and explainable features does not itself demonstrate that students can navigate those features, interpret their outputs, or use them effectively during structured learning

activities (Afzaal et al., 2024; Ramaswami et al., 2023; Saqr & López-Pernas, 2024) . Hybrid modeling and SHAP-based interpretation have been investigated in renewable energy forecasting (Maarif et al., 2023; Mubarak et al., 2023; Van Poecke et al., 2024) , yet technical evidence concerning model performance provides limited insight into the usability barriers students may encounter when the same components are presented as an integrated learning environment.

Against this background, the present study positions its contribution within educational technology rather than the development of a new forecasting algorithm or explanation method. It evaluates an RF-LSTM learning platform with XAI by examining how university students interact with an integrated RF-LSTM-SHAP workflow and by identifying usability barriers associated with interface organization, technical terminology, navigation, and explainability. The study combines task-performance measures, the System Usability Scale, task observations, and open-ended student feedback so that usability can be examined from complementary perspectives rather than represented by a single score. Specifically, the study aims to determine how effectively students complete structured activities within the platform, examine their perceived usability, and identify interaction barriers and design priorities that may inform the development of student-facing ML-XAI learning environments. The contribution is therefore confined to usability and educational design implications; the study does not claim to introduce a new forecasting architecture or XAI technique, nor does it infer improvements in students' learning outcomes.

Methods

This study employed a multidimensional usability evaluation design to examine how university students interacted with an RF-LSTM learning platform incorporating explainable artificial intelligence (XAI). The evaluation was structured around complementary dimensions of usability, including effectiveness, efficiency, perceived usability, and interaction difficulties. These dimensions were represented through task completion, completion time, task-related errors, System Usability Scale (SUS) scores, structured observation records, and open-ended student feedback. The use of multiple usability indicators was intended to provide a broader representation of system use than could be obtained from a single measure, particularly because successful task completion does not necessarily correspond to positive user perceptions or an absence of interaction difficulties (Georgsson & Staggers, 2016; Hussain et al., 2015; Lewis, 2018) . The evaluation was conducted at Universitas Esa Unggul, Jakarta, Indonesia. The target participants were students from Informatics Engineering and Computer Science programs who had completed or were currently taking courses related to machine learning, artificial intelligence, or data mining. Purposive sampling was applied to ensure that participants had sufficient academic familiarity with the concepts presented in the platform. A total of 40 students aged 19 to 23 years participated, consisting of 24 male and 16 female students. Participation was voluntary, informed consent was obtained before data collection, and participant identities were replaced with codes R01 to R40 throughout data recording, analysis, and reporting.

The research object was a web-based learning platform developed for renewable energy forecasting using Random Forest (RF), Long Short-Term Memory (LSTM), and a weighted RF-LSTM hybrid model. The platform was accessed through standard web browsers on participants' laptops and organized learning activities into five interconnected stages: data exploration, preprocessing, modeling, performance evaluation, and model explainability. Within

these stages, students could examine dataset characteristics and preprocessing procedures, generate prediction outputs from RF, LSTM, and hybrid models, compare model performance using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), and interpret SHAP-based visualizations representing the contribution of input features to model predictions. The technological specifications were described to clarify the functions with which students directly interacted during the usability evaluation rather than to evaluate the computational architecture as a new algorithmic contribution. As presented in Figure 1, the evaluation framework links the functional components of the RF-LSTM-XAI platform with structured student activities and the resulting usability evidence. This arrangement reflects the principle that educational technologies should be examined not only in terms of the information or functions they provide, but also in relation to students' ability to access, interpret, and use those functions within a learning environment (De Vreugd et al., 2024; Rets et al., 2021; Susnjak et al., 2022).

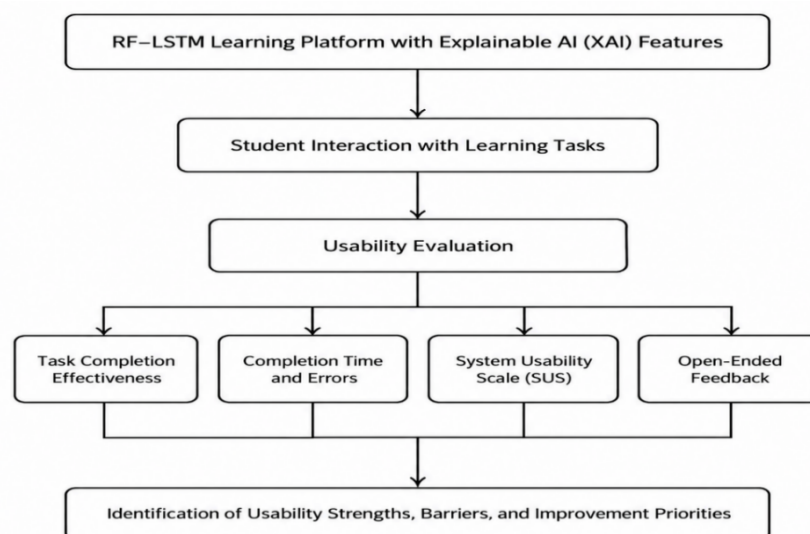


Figure 1. conceptual framework of the study

Data were collected using three complementary instruments consisting of a structured observation sheet, the 10-item Indonesian adaptation of the System Usability Scale developed by Sharfina and Santoso (2017), and two open-ended questions. SUS was selected as a standardized instrument for assessing perceived system usability and was administered using a five-point Likert response format ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) (Lewis, 2018). Because perceived usability alone cannot indicate the specific stages at which users experience operational difficulties, the SUS assessment was accompanied by structured observation of six predefined tasks. After receiving a brief orientation to the platform without demonstrations of how the tasks should be completed, each participant independently performed the activities. T1 required identification of the target variable and principal input features; T2 required identification of preprocessing transformations; T3 involved generating and locating predictions from the RF, LSTM, and hybrid models; T4 required selection of the best-performing model with justification based on RMSE, MAE, and R^2 ; T5 required interpretation of the SHAP visualization to identify the most influential features; and T6 required participants to return to a specified previous stage and locate requested information without assistance. The observation sheet was used to record task completion, completion time, observable task-related errors, and assistance required for each activity. After completing the six tasks, participants completed the SUS questionnaire and responded to the open-ended

questions concerning useful platform features, difficulties experienced during use, and suggested improvements. This combination of task-based evidence, standardized usability assessment, and user feedback was intended to capture both observable interaction performance and students' perceptions of the platform (Krabbel et al., 2025; Lewis, 2018; Reid & Drysdale, 2024) .

Data analysis was conducted according to the characteristics of each usability indicator before the findings were brought together for an integrated interpretation. Task effectiveness was calculated as the percentage of the 40 participants who successfully completed each activity according to predefined success criteria, while completion time was recorded in minutes as an indicator of interaction efficiency. Error frequency was calculated as the mean number of observable task-related errors per participant for each activity by dividing the total recorded errors by all 40 participants, including participants with no recorded errors and those who did not successfully complete the task. Task success, time on task, and errors are commonly used as complementary indicators in task-based usability assessment because they provide information about both successful goal attainment and the effort required during system interaction (Hussain et al., 2015). Wilson 95% confidence intervals were calculated for task-completion proportions to provide supplementary estimates of uncertainty around the observed completion rates (Brown et al., 2001) . SUS scores were calculated according to the established scoring procedure, in which one point was subtracted from responses to odd-numbered items, responses to even-numbered items were subtracted from five, and the adjusted scores were summed and multiplied by 2.5 to obtain an overall score ranging from 0 to 100 (Lewis, 2018) . A second researcher independently checked the SUS calculations, particularly the reverse-scored items, and the resulting score was interpreted with reference to the commonly used benchmark of 68 (Sauro & Lewis, 2016) . Open-ended responses and relevant observation notes were analyzed using inductive thematic coding. Two researchers independently coded meaningful segments of the data, grouped conceptually related codes into broader themes, and discussed coding differences until consensus was reached, following the general analytical logic of thematic analysis described by Byrne (2022) and Suhirman et al (Suhirman et al., 2026) . Finally, task-performance indicators, SUS results, observation records, and themes derived from student feedback were organized in an integrated usability evidence matrix. The matrix was used to connect observable performance with corresponding user experiences and to identify recurring usability barriers and design priorities without extending the interpretation to learning effectiveness, which was outside the scope of this study.

Result

The results are organized according to the sequence of the study objectives. The first subsection reports students' performance across the structured activities of the RF-LSTM-SHAP platform. The second describes students' overall perceptions of platform usability based on the System Usability Scale (SUS). The third presents the usability barriers identified through task observations and open-ended student feedback. The fourth consolidates the evidence from these sources to identify specific areas of the platform that emerged as priorities for design refinement. This organization maintains a direct connection between the reported findings and the usability dimensions established in the study objectives.

Task Performance across the Structured Platform Activities

The first objective of the study was to examine how effectively students completed the structured activities within the RF-LSTM-SHAP learning platform. The six tasks represented

successive forms of interaction with the platform, beginning with the identification of data characteristics and preprocessing procedures, followed by model execution and performance comparison, and continuing to SHAP interpretation and navigation between stages. Across these activities, completion rates ranged from 85.0% to 100.0%, indicating that most of the 40 participants successfully completed each task, although the observed performance differed across the workflow. The highest completion rate was recorded for T1, in which all 40 participants successfully identified the target variable and the main input features. This activity also required the shortest mean completion time, at 2.1 minutes, and produced the lowest mean error frequency, at 0.2 errors per participant. A similarly high completion rate was recorded for T2, where 39 participants successfully identified the preprocessing transformations applied within the platform, corresponding to 97.5%. The mean completion time for T2 was 3.5 minutes, with an average of 0.5 errors per participant. The Wilson 95% confidence intervals were 91.2% to 100.0% for T1 and 87.1% to 99.6% for T2. Detailed performance indicators for all six activities are presented in Table 1.

Table 1. task-completion performance

Task	Task Description	Completion Rate	95% CI	Mean Time (min)	Mean Errors per Participant
T1	Identify the target variable and main input features	100.0% (40/40)	91.2-100%	2.1	0.2
T2	Examine preprocessing steps and identify applied transformations	97.5% (39/40)	87.1-99.6%	3.5	0.5
T3	Generate RF, LSTM, and hybrid model predictions	95.0% (38/40)	83.5-98.6%	4.2	0.8
T4	Compare model performance using RMSE, MAE, and R ²	90.0% (36/40)	76.9-96.0%	3.8	1.1
T5	Use SHAP visualization to identify the most influential features	85.0% (34/40)	70.9-92.9%	5.6	1.6
T6	Return to a previous stage and locate the required result independently	92.5% (37/40)	80.1-97.4%	2.8	0.7

Performance remained high when participants progressed from data preparation to model execution, although differences in completion time and error frequency became more visible. In T3, participants were required to generate and locate prediction outputs from the RF, LSTM, and hybrid models. Thirty-eight of the 40 participants successfully completed this activity, corresponding to a completion rate of 95.0% with a Wilson 95% confidence interval of 83.5% to 98.6%. The task required an average of 4.2 minutes and produced a mean of 0.8 errors per participant. In T4, the activity shifted from generating model outputs to comparing their performance. Participants were required to identify the best-performing model using RMSE, MAE, and R² as the specified evaluation measures. Thirty-six participants completed the task successfully, resulting in a completion rate of 90.0% with a 95% confidence interval of 76.9% to 96.0%. The mean completion time was 3.8 minutes, while the mean error frequency increased to 1.1 errors per participant. Thus, although successful completion remained at or above 90.0% through T4, the recorded performance indicators show that activities involving

model comparison were accompanied by more observable errors than the earlier data-identification and preprocessing tasks.

The clearest variation in task performance appeared in the final stages of the structured activities. T5 required participants to interpret the SHAP visualization and identify the input features with the greatest influence on the displayed prediction. Thirty-four participants completed this task successfully, producing a completion rate of 85.0% with a Wilson 95% confidence interval of 70.9% to 92.9%. This was the lowest completion rate among the six activities. T5 also required the longest mean completion time, at 5.6 minutes, and produced the highest mean error frequency, at 1.6 errors per participant. Performance increased again in T6, which required participants to return to a specified previous stage and locate the requested information independently. Thirty-seven participants completed this activity successfully, corresponding to a completion rate of 92.5% with a 95% confidence interval of 80.1% to 97.4%. The mean completion time decreased to 2.8 minutes, while the mean error frequency was 0.7 errors per participant. The distribution of performance across the six structured activities is presented visually in Figure 2, allowing the differences among completion rates, time requirements, and recorded errors to be viewed across the sequence of platform interaction.

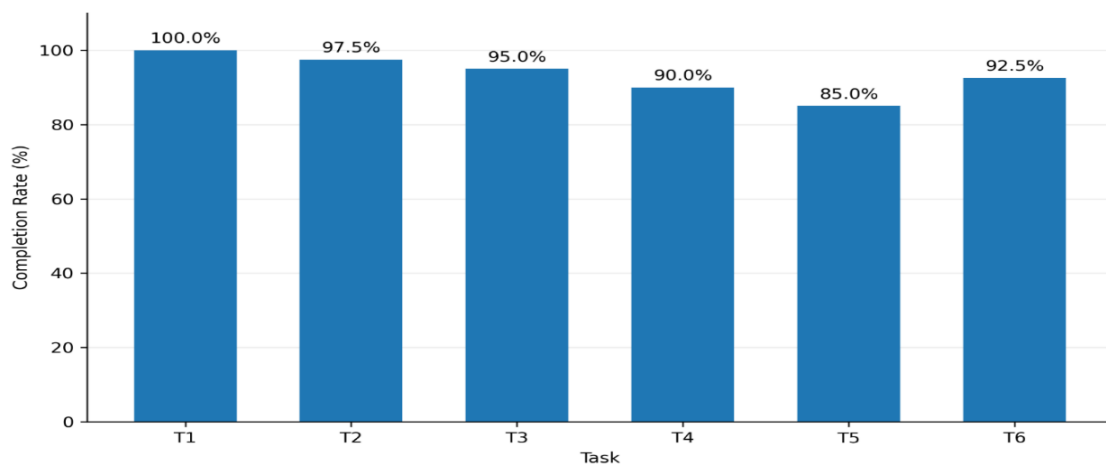


Figure 2. Task Performance across the Six Structured Activities

Viewed across the complete sequence, the results show a distinct descriptive pattern in students' interaction with different components of the platform. The first three activities, which involved identifying data components, examining preprocessing procedures, and generating model predictions, recorded completion rates between 95.0% and 100.0%, with mean error frequencies ranging from 0.2 to 0.8 errors per participant. The performance indicators became more variable when the activities required students to work with model-performance information and explainability outputs. T4 recorded a completion rate of 90.0% and a mean error frequency of 1.1, while T5 recorded the lowest completion rate at 85.0%, together with the longest completion time and the highest error frequency. The completion rate subsequently increased to 92.5% in T6, accompanied by a shorter mean completion time and fewer errors than those recorded for T5. Among the six activities, therefore, the most concentrated observable difficulty occurred during SHAP interpretation, whereas activities involving data identification, preprocessing, and model execution were completed successfully by a larger proportion of participants. These task-level differences provide the empirical basis for locating where interaction difficulties were most visible within the platform workflow.

Students' Perceptions of Platform Usability

The second objective of the study was to examine students' perceptions of the overall usability of the RF-LSTM-XAI learning platform after they had completed the six structured activities. All 40 participants completed the 10-item System Usability Scale (SUS), providing a system-level measure of perceived usability based on their experience of interacting with the platform. The resulting SUS scores produced a mean of 76.7 with a standard deviation of 12.5, while the median score was 78.0. Individual scores ranged from 45.0 to 97.5. Based on the interpretation criteria adopted in this study, the mean SUS score corresponded to the **Good** adjective rating and the **Acceptable** acceptability level. The mean score was also 8.7 points above the reference value of 68 used in the evaluation. The descriptive statistics of the SUS results are presented in Table 2.

Table 2. Descriptive Statistics of System Usability Scale Scores

Indicator	Result
Number of participants	40
Mean SUS score	76.7
Standard deviation	12.5
Median	78.0
Range	45–97.5
Minimum score	45
Maximum score	97.5
Adjective rating	Good
Acceptability	Acceptable

The distribution of SUS scores provides a more detailed description of students' perceptions than the mean value alone. The mean score of 76.7 and the median of 78.0 differed by only 1.3 points, placing the two measures of central tendency relatively close to one another. At the same time, the standard deviation of 12.5 shows that individual ratings were distributed around the group mean rather than concentrated at a single value. This variation is also visible in the difference between the lowest and highest scores. The minimum SUS score was 45.0, whereas the maximum reached 97.5, producing a difference of 52.5 points between the two observed extremes. The results therefore show that the overall group assessment fell within a positive usability category, while the individual evaluations covered a considerably broader range.

The system-level SUS results can also be considered alongside the task-performance findings reported in the preceding subsection. Across the six structured activities, completion rates ranged from 85.0% to 100.0%, meaning that a substantial proportion of participants successfully completed each activity. At the same time, performance was not identical across the platform workflow. T1 recorded 100.0% completion, followed by 97.5% for T2 and 95.0% for T3, whereas completion decreased to 90.0% for T4 and reached its lowest level of 85.0% for T5 before increasing to 92.5% for T6. The overall SUS mean of 76.7 was therefore obtained in a context in which students generally completed the platform activities successfully, but some activities showed more observable difficulty than others. In particular, the lower completion rate, longer completion time, and higher error frequency recorded for T5 were present alongside an overall perceived-usability score classified as Good and Acceptable.

Viewed together, these findings distinguish between the overall perception of the platform and the variation observed within particular interactions. SUS summarizes students' evaluations after experiencing the platform as a complete system, whereas the task-

performance indicators describe what occurred during individual activities. The positive mean SUS score therefore does not imply that every participant reported an identical experience or that every platform component produced the same pattern of interaction. This is also reflected in the individual SUS scores, which extended from 45.0 to 97.5. At the descriptive level, the findings show that the participant group generally evaluated the platform positively, while both the distribution of SUS scores and the task-level results retained evidence of variation in students' experiences. The specific areas associated with these variations are described in the following subsection through the observation records and open-ended student feedback.

Usability Barriers Identified during Platform Interaction

The third objective of the study was to identify the usability barriers experienced by students while interacting with the RF-LSTM-XAI learning platform. Evidence from the open-ended responses and observation records showed that students' experiences included both positively perceived platform features and specific difficulties encountered during task completion. Participants identified the SHAP visualization as useful for examining influential input features and regarded the model-comparison function as helpful for viewing the outputs of RF, LSTM, and the hybrid model within the same platform. The sequence of activities, beginning with data exploration and preprocessing and continuing through modeling and performance evaluation, was also generally perceived as understandable. Alongside these positive responses, recurring difficulties were identified in four areas: understanding technical terminology, recognizing the current position within the platform workflow, examining the SHAP visualization, and receiving sufficiently clear system feedback during model execution. A summary of these findings and the corresponding suggestions provided by participants is presented in Table 3.

Table 3. Summary of Usability Findings from Student Feedback and Observation

Category	Usability findings	Suggested improvement
Perceived strengths	SHAP visualization helped students identify influential features	Maintain the explainability function while improving its presentation
Perceived strengths	Model comparison supported examination of RF, LSTM, and hybrid-model outputs	Maintain the model-comparison function
Perceived strengths	The sequence from data exploration to evaluation was generally understandable	Maintain the sequential structure of the platform
Usability barrier	RMSE, MAE, R^2 , and SHAP were difficult for some students to interpret	Provide a glossary, contextual explanations, or tooltips
Usability barrier	The current stage within the platform workflow was not always clearly indicated	Add a persistent progress indicator or stage marker
Usability barrier	SHAP visualization appeared too small on laptop screens	Increase the visualization size and provide zoom functionality
Usability barrier	Hybrid-model execution involved a relatively long loading period	Improve processing feedback and make system status more visible

The difficulty associated with technical terminology was particularly visible in activities that required students to interpret information rather than merely locate or generate it. In T4, participants were required to compare RF, LSTM, and hybrid-model performance using RMSE,

MAE, and R^2 and then determine the best-performing model. Thirty-six of the 40 participants successfully completed this activity, resulting in a completion rate of 90.0%, while the mean error frequency reached 1.1 errors per participant. Student feedback identified the terminology associated with these performance measures as an area in which additional explanation would be useful. A comparable pattern appeared in T5, where participants were required to interpret the SHAP visualization and identify the most influential features. Only 34 participants successfully completed this activity, corresponding to the lowest task-completion rate of 85.0%, while the mean error frequency reached 1.6 errors per participant. In the open-ended responses, SHAP was simultaneously described as useful for identifying influential features and as a concept that was not equally easy for all students to interpret. The findings therefore show that the explainability and performance-evaluation components were available and usable for many participants, while the terminology accompanying these components remained a visible source of difficulty for some students.

A different set of barriers concerned navigation and visual presentation. Participants reported that the platform did not always provide a sufficiently visible indication of their current position within the sequence of activities. This issue was reflected in T6, where students were required to return to a specified previous stage and locate particular information without assistance. Thirty-seven participants successfully completed the task, while three were unable to do so independently, producing a completion rate of 92.5%. The associated feedback focused on the need for a clearer indication of the user's current stage, leading to suggestions for a persistent progress indicator or similar navigation marker. Visual presentation was particularly relevant to the SHAP component. Although students recognized the value of the SHAP visualization, some reported that its display was too small when viewed on a laptop screen. This observation occurred in relation to T5, which also recorded the longest mean completion time of 5.6 minutes. Participants consequently suggested increasing the size of the SHAP display and providing a zoom function to make the visual information easier to inspect.

The remaining barrier concerned system response during model execution. Participants reported that the hybrid model required a relatively long loading period during the prediction activity. This issue was associated with T3, where students generated and located prediction outputs from the RF, LSTM, and hybrid models. Despite the reported waiting period, 38 of the 40 participants completed the task successfully, corresponding to a completion rate of 95.0%. The loading issue therefore appeared in participant feedback even though successful task completion remained high. Students suggested clearer system feedback during processing so that the status of model execution could be more readily recognized. Across the observation records and open-ended responses, the usability barriers were consequently concentrated at several distinct points in the interaction process. Technical terminology was associated primarily with model evaluation and SHAP interpretation, navigation difficulties appeared when students moved between stages, visualization concerns were concentrated around the presentation of SHAP, and system-response concerns appeared during hybrid-model execution. These findings identify the specific platform components associated with students' reported difficulties without extending the results beyond the usability experience examined in this study.

Usability Evidence and Design Priorities

The final results subsection consolidates the usability evidence obtained from task performance, structured observations, SUS assessment, and open-ended student feedback to

identify the aspects of the RF-LSTM-XAI learning platform that most consistently emerged as priorities for refinement. Rather than presenting these findings as separate sources of information, the evidence was organized according to the specific platform activities in which the corresponding usability issues were observed or reported. This arrangement made it possible to examine whether difficulties recorded during task completion were also reflected in students' descriptions of their interaction with the platform. The integrated evidence showed that the most prominent correspondences occurred in four areas: SHAP interpretation, understanding model-performance metrics, navigation between platform stages, and system response during model execution. At the overall platform level, the mean SUS score of 76.7 remained within the Good and Acceptable categories adopted in this study, while the task-specific evidence identified particular components in which interaction difficulties were more visible. The relationships among these findings and the resulting design priorities are summarized in Table 4.

Table 4. Integrated Usability Evidence and Design Priorities

Usability evidence	Student-reported or observed issue	Design priority
T5: 85.0% completion, 5.6 min mean completion time, 1.6 mean errors per participant	SHAP visualization was too small on laptop screens and technical terminology was difficult for some students to interpret	Enlarge the SHAP display, provide zoom functionality, and add contextual explanations or tooltips
T4: 90.0% completion, 1.1 mean errors per participant	RMSE, MAE, and R^2 were difficult for some students to interpret during model comparison	Provide a glossary and concise contextual explanations of evaluation metrics
T6: 92.5% completion, with three participants unable to complete the task independently	Current position within the platform workflow was not always clearly indicated	Add a persistent progress indicator or breadcrumb-style navigation
T3: 95.0% completion, with a reported loading concern	Hybrid-model execution involved a relatively long loading period	Improve processing feedback and provide a clearer indication of system status
SUS: mean 76.7, Good, Acceptable, and above the reference value of 68	Overall perception was positive, although individual SUS scores varied	Maintain positively perceived platform features while addressing task-specific usability barriers

The strongest correspondence between task performance and student-reported experience was found in T5. This activity required students to interpret the SHAP visualization and identify the features with the greatest influence on the prediction. Thirty-four of the 40 participants successfully completed the activity, resulting in a completion rate of 85.0%, the lowest among all six tasks. T5 also required the longest mean completion time at 5.6 minutes and generated the highest mean error frequency at 1.6 errors per participant. The student feedback concerning the same activity identified two related issues: the SHAP visualization appeared too small when viewed on laptop screens, and the technical terminology associated with the explainability component was not equally easy for all students to interpret. The convergence of these observations identified the presentation of SHAP information as a specific area for refinement. The corresponding design priorities derived from the reported

experience included enlarging the visualization, providing zoom functionality, and making contextual explanations or tooltips available alongside technical information. A related pattern was observed in T4. The model-comparison activity recorded a completion rate of 90.0% and a mean of 1.1 errors per participant, while students specifically identified RMSE, MAE, and R^2 as terms requiring clearer explanation. Consequently, concise descriptions or contextual support for the evaluation metrics emerged as an additional design priority associated with model-performance comparison.

A second group of correspondences concerned the organization of interaction within the platform and the visibility of system status. In T6, 37 of the 40 participants successfully returned to a specified previous stage and located the requested information independently, resulting in a completion rate of 92.5%. Three participants, however, were unable to complete the activity without assistance. The observation records and open-ended feedback concerning navigation similarly indicated that the platform did not always make the user's current position within the sequence of activities sufficiently visible. The evidence associated with T6 therefore identified a persistent progress indicator, stage marker, or breadcrumb-style navigation as a corresponding design priority. A different pattern appeared in T3. Thirty-eight participants successfully generated and located prediction outputs from the RF, LSTM, and hybrid models, resulting in a completion rate of 95.0%. Despite this high level of task completion, participants reported that execution of the hybrid model involved a relatively long loading period. The corresponding priority was therefore not associated with the availability of the model output itself, but with clearer feedback while processing was underway. A visible processing indicator or clearer system-status information emerged as the refinement most directly connected with this reported experience.

At the broader platform level, the task-specific findings were considered alongside the overall SUS result. The mean SUS score of 76.7 placed the platform within the Good adjective category and the Acceptable usability level adopted in this study, and the mean remained above the reference value of 68. At the same time, individual SUS scores ranged from 45.0 to 97.5, while the task-level findings showed that difficulties were concentrated in particular activities rather than distributed uniformly across the workflow. The integrated evidence therefore identified four principal areas for refinement: clearer contextual support for technical terminology and performance metrics, more persistent information about the user's position within the platform sequence, more accessible presentation of SHAP visualizations, and clearer feedback during computational processing. These priorities were derived directly from the correspondence among task-performance indicators, observation records, SUS results, and student feedback. Accordingly, they represent specific usability priorities identified through the present evaluation and do not extend the findings to claims about improvements in students' learning achievement or the instructional effectiveness of the platform.

Discussion

The findings of this study provide a more nuanced understanding of the usability of a learning platform that integrates machine learning-based prediction with explainable artificial intelligence (XAI). Overall, students were able to interact successfully with the core functions of the RF-LSTM-XAI platform and expressed positive perceptions of its usability; however, the ease of interaction was not uniform across the sequence of activities. Students encountered relatively fewer difficulties when identifying data characteristics, examining preprocessing procedures, and executing the models than when comparing model-performance measures and

interpreting explainability outputs. This pattern suggests that the demands of using the platform became more pronounced when students moved beyond locating or generating information and were required to understand the meaning of technical outputs and use them to make analytical judgments. Successful operation of a system function and meaningful interpretation of the information produced by that function should therefore be regarded as related but distinct dimensions of the user experience. Park and Jo (2019) found that the visual characteristics and usability of a learning analytics dashboard were associated with students' understanding and perceived usefulness. Rets et al (2021) similarly showed that students did not respond uniformly to individual dashboard components and that some graphical representations required additional explanation to be adequately understood. This perspective is further supported by de Vreugd et al. (de Vreugd et al., 2024), who demonstrated that students could perceive a learning analytics dashboard as useful while still requiring reference frames and interface refinements to facilitate interpretation. The present study is consistent with these findings while extending them to a learning environment integrating RF, LSTM, a hybrid model, and XAI. Students' ability to complete most activities does not imply that all stages imposed equivalent interpretive demands. The findings therefore address the first objective of the study by demonstrating that the platform was generally usable for completing the intended activities while also revealing an increasing need for support as interaction shifted from accessing information to analytical interpretation. From an educational perspective, AI-based learning platforms should therefore do more than provide access to computational outputs; they should also support students in transforming those outputs into information that can be understood and meaningfully used within learning activities (Clow, 2012; de Vreugd et al., 2024; Khosravi, Shum, et al., 2022; Park & Jo, 2019b; Rets et al., 2021).

The distinction between accessing information and interpreting it was most apparent in students' interaction with SHAP. Students regarded the SHAP visualization as useful for identifying influential features, yet interpreting SHAP simultaneously emerged as the most demanding activity within the evaluated workflow. These findings are not inherently contradictory. Rather, they reveal an important distinction between making a model explainable and making its explanation comprehensible to users. Through the XAI-ED framework, Khosravi et al. (2022) position explainability in education as an issue that extends beyond the algorithmic techniques used to generate explanations. Their framework also considers stakeholder characteristics, approaches to presenting explanations, human-centred interface design, and potential challenges that may arise when explanations are communicated to users. Shin (Shin, 2022) provide a compatible perspective through XDesign by positioning interface design, user needs, and task context as important components of XAI education and development. At the same time, applications of SHAP in educational predictive modelling demonstrate that the method retains substantial analytical value because it can expose the relative contribution of individual features and increase the transparency of machine-learning predictions. The difficulties observed in the present study should therefore not be interpreted as evidence that SHAP is unsuitable for educational applications. Instead, they indicate that information that is transparent at the algorithmic level may not automatically be transparent at the user level. This distinction becomes particularly important when considered alongside the concerns raised by Hooshyar and Yang (2024) regarding the interpretability and fidelity of post-hoc explanation methods in educational predictive modelling. The present findings consequently suggest a need to distinguish model-level interpretability from student-level comprehensibility. Although SHAP information was available and perceived as useful, its visual representation and

accompanying technical terminology still required greater interpretive effort than several other platform activities. This finding addresses the study objective concerning usability barriers by showing that the central issue was not the absence of explanations, but the distance between computationally generated explanations and explanations that students could meaningfully understand. The educational value of XAI therefore depends not only on a model's capacity to explain its predictions, but also on the capacity of the interface and learning design to translate those explanations into forms that students can access, interpret, and use (Hooshyar & Yang, 2024; Khosravi, Buckingham Shum, et al., 2022; Shin, 2022) .

The overall usability assessment further reinforces the importance of distinguishing system-level perceptions from activity-level experiences. The SUS results indicated a positive perception of usability at the group level, while variation in individual evaluations and the persistence of difficulties in particular activities suggest that the user experience was not entirely homogeneous. This distinction is methodologically important because an aggregate usability score provides an overall indication of how users perceive a system but does not directly reveal which components facilitate interaction or create difficulties. De Vreugd et al. (2024) addressed a comparable methodological issue when evaluating a learning analytics dashboard by combining SUS with think-aloud usability testing and interviews. This approach enabled overall evaluations of the system to be considered alongside specific difficulties that emerged during interaction. Rets et al. (2021) similarly demonstrated that the usefulness of individual components was not experienced uniformly, as students responded differently to the information and visualizations presented through the dashboard. Other studies may report SUS scores that are either higher or lower than those obtained in the present evaluation, but such numerical differences should not be interpreted as direct contradictions because the technologies, participant characteristics, task structures, system purposes, and evaluation contexts differ. Instead, these variations reinforce the contextual nature of usability and demonstrate the limitations of interpreting it through a single aggregate measure. In the present study, combining SUS with task completion, completion time, errors, structured observations, and open-ended feedback enabled usability to be examined as a multilayered experience. A positive system-level perception could coexist with interaction barriers concentrated in particular activities. Methodologically, this finding supports a multidimensional approach to usability evaluation in educational AI. SUS can characterize users' overall perceptions of a system, while task-based and experiential evidence can identify more precisely where interaction barriers emerge. Evaluations of AI-based educational platforms should therefore move beyond asking whether a system is generally perceived as easy to use and should also examine which activities are most demanding, what forms of difficulty occur, and what forms of design support are needed to address them (De Vreugd et al., 2024; Rets et al., 2021) .

The barriers related to technical terminology, visual presentation, navigation, and system feedback further indicate that difficulties in using the platform emerged through an interaction between conceptual complexity and the way that complexity was represented through the interface. Students were not merely interacting with software; they were simultaneously required to read and interpret disciplinary concepts such as RMSE, MAE, R^2 , and SHAP while completing the activities. Usability challenges therefore cannot be attributed exclusively either to the technical characteristics of the system or to students' capabilities. Some of the demands emerged at the point where complex disciplinary information was translated into specific visual and navigational elements. Liu et al. (2024) demonstrated that analytical visualizations can generate different cognitive demands depending on users' visualization literacy and that

representational support, including data storytelling, can facilitate the interpretation of displayed information. Park and Jo (2019) likewise demonstrated relationships among visual characteristics, usability, and users' understanding of dashboard information. More specifically, Jivet et al. (Jivet et al., 2020a) highlighted the importance of reference frames in helping students interpret information presented through learning analytics dashboards and use it to generate internal feedback. These studies provide a useful basis for understanding why students in the present study suggested contextual explanations, larger SHAP displays, zoom functionality, and clearer indicators of their position within the platform workflow. Navigation and system responsiveness form part of the same interaction experience because users need to understand both their current position within a sequence of activities and the process being performed by the system. Bodily, R., & Verbert, K (Bodily & Verbert, 2017a) similarly identified the translation of increasingly sophisticated analytical capabilities into information that learners can interpret and meaningfully use as an important challenge for learner-facing analytics. Nevertheless, the present study does not establish that technical terminology, visualization size, navigation, or loading time directly caused differences in student performance. The evidence indicates correspondence between observed or reported issues and the activities in which difficulties emerged rather than a causal relationship. A proportionate design implication is therefore to provide adequate scaffolding without removing the academic complexity that students are expected to learn. Contextual glossaries and tooltips could support technical terminology, scalable visualizations could improve access to detailed information, persistent progress indicators could clarify users' position within the workflow, and processing indicators could make ongoing computational activity more visible (Bodily & Verbert, 2017b; Jivet et al., 2020b; R. Liu et al., 2024; Park & Jo, 2019a; Susnjak et al., 2022) .

When considered collectively, the findings extend beyond the identification of individual interface problems and point toward an analytical distinction among algorithmic explainability, representational comprehensibility, and interaction usability. Algorithmic explainability concerns the capacity of a system to provide information about the factors underlying or contributing to a model's predictions. Representational comprehensibility refers to the extent to which this information is expressed through visualizations, terminology, and explanations that students can meaningfully understand. Interaction usability, in turn, concerns students' ability to access, navigate, and use those representations while performing learning activities. This distinction should not be regarded as a newly validated theoretical model because the study was not designed to empirically test relationships among these three concepts. Rather, it represents an analytical proposition derived from the pattern of findings and provides a basis for further investigation. This proposition extends the human-centred orientation of the XAI-ED framework proposed by Khosravi et al. (2022) and is compatible with Shin et al. (2022) , who position interface design as an integral component of user-oriented explanation development. Alfredo et al. (2024) , in a systematic review of 108 studies on human-centred learning analytics and AI in education, similarly identified limited meaningful involvement of end users in many development processes and emphasized the need for stronger human-centred approaches. Within this perspective, the four refinement priorities identified in the present study, contextual support for terminology and performance metrics, consistent navigation cues, more accessible SHAP visualizations, and clearer processing feedback, should not be viewed merely as cosmetic interface modifications. Rather, they translate students' interaction experiences into concrete directions for design refinement. The scientific contribution of the study therefore does not lie in proposing a new RF, LSTM, hybrid, or SHAP

algorithm, but in identifying the interaction layer between computational outputs and student interpretation as an important component of educational AI evaluation. In practical terms, the quality of an AI-based learning platform should be evaluated not only by its capacity to generate predictions and explanations, but also by whether its intended users can meaningfully understand and use that information within educational activities (Alfredo et al., 2024; Jivet et al., 2020b; Khosravi, Buckingham Shum, et al., 2022).

These contributions should nevertheless be interpreted within the boundaries of the study. The evaluation involved 40 students from a single university who had prior academic exposure to machine learning, artificial intelligence, or data mining. The usability patterns identified in this study therefore cannot be directly generalized to students from different disciplines, institutions, educational levels, or levels of prior knowledge. The study also focused specifically on usability and was not designed to evaluate learning effectiveness. Consequently, successful activity completion, positive usability perceptions, or students' recognition of SHAP as useful cannot be interpreted as evidence that the platform improved conceptual understanding, forecasting competence, or academic achievement. Such caution is consistent with Susnjak et al. (2022), who indicate that increasingly sophisticated analytical functions in learning technologies do not in themselves establish measurable educational benefits. Matcha et al. (Matcha et al., 2020) similarly showed that the relationship among learning analytics dashboard design, support for learning processes, and learning outcomes requires more systematic evaluation. In addition, the observational records and open-ended responses in the present study were able to identify recurring experiences but were not designed to determine the magnitude of the causal effect of individual usability barriers on student performance. Future research should therefore evaluate the refined platform with larger and more heterogeneous participant groups, including students with different levels of AI, data, and visualization literacy. Pre-test and post-test designs could examine whether interface refinements are associated not only with improved usability but also with changes in conceptual understanding, while think-aloud protocols, interaction logs, screen recordings, and eye-tracking could provide more detailed evidence about where students hesitate, redirect their attention, or misinterpret technical information. Comparative studies could also examine SHAP alongside alternative explanation formats to determine whether the difficulties identified in this study are associated with the particular representation implemented in the platform or reflect a broader challenge in translating XAI into educational contexts. Such an agenda could move educational XAI research beyond the relatively straightforward question of whether students can operate an explainable AI platform toward the more substantive question of how algorithmic explanations should be represented and integrated into learning environments so that computational transparency becomes genuinely comprehensible and educationally meaningful to students.

Conclusion

This study demonstrates that the RF-LSTM learning platform integrated with explainable artificial intelligence generally provided a positive usability experience for university students, although ease of interaction was not uniform across all stages of the workflow. Students were relatively able to complete activities involving data exploration, preprocessing, and the use of model outputs, whereas interaction became more demanding when they were required to compare performance metrics and interpret SHAP visualizations. Overall perceptions of usability were also positive, although task-based evidence, observations, and student feedback

identified barriers related to technical terminology, navigation, SHAP visualization, and feedback during system processing. The study therefore achieved its objectives by demonstrating that the usability of an educational AI platform should be evaluated not only in terms of users' ability to operate the system, but also in terms of their ability to understand and meaningfully use the analytical information it provides.

The contribution of this study lies in emphasizing that explainability within a learning environment requires more than the availability of algorithmic explanations; it also depends on comprehensible representations and interactions that support students' interpretation of those explanations. The findings provide practical directions for platform refinement, particularly through contextual explanations of technical terminology and performance metrics, consistent navigation cues, more accessible and adjustable SHAP visualizations, and more informative processing feedback. Future research should evaluate these refinements with larger and more diverse student populations and extend the evaluation beyond usability to include conceptual understanding and learning outcomes. Comparative investigation of alternative XAI representations would also be valuable for determining how algorithmic explanations can be presented in forms that are more comprehensible to students without reducing the scientific complexity that forms part of the learning process.

References

- Afzaal, M., Zia, A., Nouri, J., & Fors, U. (2024). Informative Feedback and Explainable AI-Based Recommendations to Support Students' Self-regulation. *Technology, Knowledge and Learning*, 29, 331–354. <https://doi.org/10.1007/s10758-023-09650-0>
- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- Bobek, S., Korycińska, P., Krakowska, M., Mozolewski, M., Rak, D., Zych, M., Wójcik, M., & Nalepa, G. J. (2025). User-centric evaluation of explainability of AI with and for humans: a comprehensive empirical study. *International Journal of Human-Computer Studies*, 103625.
- Bodily, R., & Verbert, K. (2017a). Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems. *IEEE Transactions on Learning Technologies*, 10(4), 405–418. <https://doi.org/10.1109/TLT.2017.2740172>
- Bodily, R., & Verbert, K. (2017b). Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems. *IEEE Transactions on Learning Technologies*, 10(4), 405–418. <https://doi.org/10.1109/TLT.2017.2740172>
- Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2), 101–133.
- Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, 56, 1391–1412. <https://doi.org/10.1007/s11135-021-01182-y>
- Castilla, D., Jaen, I., Suso-Ribera, C., Garcia-Soriano, G., Zaragoza, I., Breton-Lopez, J., Mira, A., Diaz-Garcia, A., & Garcia-Palacios, A. (2024). Psychometric properties of the Spanish full and short forms of the system usability scale (SUS): detecting the effect of negatively worded items. *International Journal of Human-Computer Interaction*, 40(15), 4145–4151.
- Clow, D. (2012). The learning analytics cycle. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134–138. <https://doi.org/10.1145/2330601.2330636>
- De Barba, P., Oliveira, E. A., & Hu, X. (2022). Same graph, different data: A usability study of a student-facing dashboard based on self-regulated learning theory. *ASCILITE 2022: Reconnecting Relationships through Technology*. <https://doi.org/10.14742/apubs.2022.168>

- de Vreugd, L., van Leeuwen, A., Jansen, R., & van der Schaaf, M. (2024). Learning Analytics Dashboard Design and Evaluation to Support Student Self-Regulation of Study Behaviour. *Journal of Learning Analytics*, 11(3), 249–262. <https://doi.org/10.18608/jla.2024.8529>
- De Vreugd, L., Van Leeuwen, A., Jansen, R., & Van der Schaaf, M. (2024). Learning Analytics Dashboard Design and Evaluation to Support Student Self-Regulation of Study Behaviour. *Journal of Learning Analytics*, 1–14. <https://doi.org/10.18608/jla.2024.8529>
- El Marzki, M. (2025). Predicting and Interpreting Student Academic Performance: A Deep Learning and Shapley Additive Explanations Approach. *SHS Web of Conferences*. https://www.shs-conferences.org/articles/shsconf/abs/2025/05/shsconf_cifem2024_01001/shsconf_cifem2024_01001.html
- Georgsson, M., & Stagers, N. (2016). Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. *Journal of the American Medical Informatics Association*, 23(1), 5–11. <https://doi.org/10.1093/jamia/ocv099>
- Ghimire, S., Abdulla, S., Joseph, L. P., Prasad, S., Murphy, A., Devi, A., Barua, P. D., Deo, R. C., Acharya, R., & Yaseen, Z. M. (2024). Explainable artificial intelligence-machine learning models to estimate overall scores in tertiary preparatory general science course. *Computers and Education: Artificial Intelligence*, 7, 100331. <https://doi.org/10.1016/j.caeai.2024.100331>
- Guan, Y., Wang, F., & Song, S. (2026). Interpretable machine learning for academic performance prediction: A SHAP-based analysis of key influencing factors. *Innovations in Education and Teaching International*, 63(5), 1553–1572. <https://doi.org/10.1080/14703297.2025.2532050>
- Hooshyar, D., & Yang, Y. (2024). Problems With SHAP and LIME in Interpretable AI for Education: A Comparative Study of Post-Hoc Explanations and Neural-Symbolic Rule Extraction. *IEEE Access*, 12, 137472–137490. <https://doi.org/10.1109/ACCESS.2024.3463948>
- Hussain, A., Mkpojiogu, E. O. C., & Hussain, Z. (2015). USABILITY EVALUATION OF A WEB-BASED HEALTH AWARENESS PORTAL ON SMARTPHONE DEVICES USING ISO 9241-11 MODEL. *Jurnal Teknologi*, 77(4). <https://doi.org/10.11113/jt.v77.6035>
- Jivet, I., Scheffel, M., Schmitz, M., Robbers, S., Specht, M., & Drachsler, H. (2020a). From students with love: An empirical study on learner goals, self-regulated learning and sense-making of learning analytics in higher education. *The Internet and Higher Education*, 47, 100758. <https://doi.org/10.1016/j.iheduc.2020.100758>
- Jivet, I., Scheffel, M., Schmitz, M., Robbers, S., Specht, M., & Drachsler, H. (2020b). From Students with Love: An Empirical Study on Learner Goals, Self-Regulated Learning and Sense-Making of Learning Analytics in Higher Education. *The Internet and Higher Education*, 47, 100758. <https://doi.org/10.1016/j.iheduc.2020.100758>
- Kaveri, A., Silvola, A., & Muukkonen, H. (2023). Supporting Student Agency with a Student-Facing Learning Analytics Dashboard: Perceptions of an Interdisciplinary Development Team. *Journal of Learning Analytics*, 10(2), 85–99. <https://doi.org/10.18608/jla.2023.7729>
- Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martínez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Krabbel, H., Munim, Z. H., & Bustgaard, M. (2025). Factors Affecting the System Usability of a Maritime Learning Analytics Dashboard Prototype. In *Integrating Emerging Technologies into Education and Training* (pp. 11–22). https://doi.org/10.1007/978-3-031-84170-5_2

- Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction*, 34(7), 577–590. <https://doi.org/10.1080/10447318.2018.1455307>
- Liu, L., Zhou, T., Li, C., Huang, Y., Wu, Y., & Chen, D. (2024). The mediating effect of newly graduated nurses' coping styles between personality traits and transitional shock. *Chinese Journal of Nursing*, 59(20), 2514–2521. <https://doi.org/10.3761/j.issn.0254-1769.2024.20.012>
- Liu, R., Zhang, L., Cao, Z., & Mi, J. (2024). Influence of gamification affordance on young citizens' motivation and learning performance toward digital civic education. *Journal of Chinese Governance*. <https://doi.org/10.1080/23812346.2024.2327735>
- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3313831.3376727>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 4765–4774. <https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions>
- Maarif, M. R., Saleh, A. R., Habibi, M., Fitriyani, N. L., & Syafrudin, M. (2023). Energy Usage Forecasting Model Based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI). *Information*, 14(5), 265. <https://doi.org/10.3390/info14050265>
- Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226–245.
- Mubarak, H., Hammoudeh, A., Ahmad, S., Abdellatif, A., Mekhilef, S., Mokhlis, H., & Dupont, S. (2023). A hybrid machine learning method with explicit time encoding for improved Malaysian photovoltaic power prediction. *Journal of Cleaner Production*, 382, 134979. <https://doi.org/10.1016/j.jclepro.2022.134979>
- Park, Y., & Jo, I.-H. (2019a). Development of the Learning Analytics Dashboard to Support Students' Learning Performance. *Journal of Universal Computer Science*, 21(1), 110–133.
- Park, Y., & Jo, I.-H. (2019b). Factors that affect the success of learning analytics dashboards. *Educational Technology Research and Development*, 67, 1547–1571. <https://doi.org/10.1007/s11423-019-09693-0>
- Raj, N. S., & Renumol, V. G. (2022). Early prediction of student engagement in virtual learning environments using machine learning techniques. *E-Learning and Digital Media*, 19(6), 537–554. <https://doi.org/10.1177/20427530221108027>
- Ramaswami, G., Susnjak, T., & Mathrani, A. (2023). Effectiveness of a Learning Analytics Dashboard for Increasing Student Engagement Levels. *Journal of Learning Analytics*, 10(3), 115–134. <https://doi.org/10.18608/jla.2023.7935>
- Reid, D., & Drysdale, T. (2024). Student-facing learning analytics dashboard for remote lab practical work. *IEEE Transactions on Learning Technologies*, 17, 1037–1050. <https://doi.org/10.1109/TLT.2024.3354128>
- Rets, I., Herodotou, C., Bayer, V., Hlosta, M., & Rienties, B. (2021). Exploring critical factors of the perceived usefulness of a learning analytics dashboard for distance university students. *International Journal of Educational Technology in Higher Education*, 18, 46. <https://doi.org/10.1186/s41239-021-00284-9>
- Saqr, M., & López-Pernas, S. (2024). Why explainable AI may not be enough: predictions and mispredictions in decision making in education. *Smart Learning Environments*, 11, 52. <https://doi.org/10.1186/s40561-024-00343-4>
- Sauro, J., & Lewis, J. R. (2016). *Quantifying the User Experience: Practical Statistics for User Research* (2nd ed.). Morgan Kaufmann. <https://www.sciencedirect.com/book/monograph/9780128023082/quantifying-the-user-experience>

- Shapiro, R. B., Fiebrink, R., & Norvig, P. (2018). How Machine Learning Impacts the Undergraduate Computing Curriculum. *Communications of the ACM*, 61(11), 27–29. <https://cacm.acm.org/magazines/2018/11/232197-how-machine-learning-impacts-the-undergraduate-computing-curriculum>
- Shin, W. (2022). *Parental Digital Literacy and Mediation*.
- Skulmowski, A., & Xu, K. M. (2022). Understanding Cognitive Load in Digital and Online Learning: a New Perspective on Extraneous Cognitive Load. *Educational Psychology Review*, 34, 171–196. <https://doi.org/10.1007/s10648-021-09624-7>
- Subramonyam, H., & Hullman, J. (2023). Are we closing the loop yet? gaps in the generalizability of vis4ml research. *IEEE Transactions on Visualization and Computer Graphics*, 30(1), 672–682.
- Suhrman, L., Nugroho, A. Y., Dhyanaputri, I. G. A. S., Adawiyah, R., Soraya, A., Dzaky, M. I., Alunaza, H., Akbar, S. A., Hafizi, M. Z., & Junida, D. S. (2026). *Dasar Metodologi Penelitian*. CV. Askara Sastra Media.
- Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: a tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, 19, 12. <https://doi.org/10.1186/s41239-021-00313-7>
- Van Poecke, A., Tabari, H., & Hellinckx, P. (2024). Unveiling the backbone of the renewable energy forecasting process: Exploring direct and indirect methods and their applications. *Energy Reports*, 11, 544–557. <https://doi.org/10.1016/j.egy.2023.12.031>
- Vlachogianni, P., & Tselios, N. (2022). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education*, 54(3), 392–409.
- Wang, J., Liu, S., & Zhang, W. (2024). Visual analytics for machine learning: A data perspective survey. *IEEE Transactions on Visualization and Computer Graphics*, 30(12), 7637–7656.