p-ISSN 2407-4489
e-ISSN 2528-7303

Designing Explainable and Actionable AI for At-Risk Students: A Learning Analytics Design Science Approach

Sahrul Amin[a], Vera Irma Delianti[b]
[ab]Faculty of Engineering, Universitas Negeri Padang, Padang, Indonesia
email: [a]sahrulamin.2019@gmail.com, [b]vera5339@ft.unp.ac.id

ABSTRACT

The increasing adoption of Learning Management Systems (LMS) in higher education generates learning traces that can support early detection of at-risk students. However, many academic early warning systems still present risk scores without explaining risk drivers or translating them into feasible instructor actions. This study designs an explainable and actionable artificial intelligence framework using a Design Science Research approach. The artifact integrates LMS data, learning feature engineering, risk prediction, SHAP-based explanation, constrained Large Language Model-based recommendation, and human-in-the-loop validation. Its novelty lies in connecting prediction, explanation, pedagogical recommendation, instructor validation, and post-intervention feedback within one ethically constrained Learning Analytics workflow. The design outputs include requirements, input-process-output architecture, risk-threshold logic, SHAP-LLM workflow, intervention taxonomy, recommendation template, and evaluation roadmap. The framework remains conceptual; future studies should instantiate it with real LMS data and evaluate predictive performance, explanation quality, recommendation usefulness, fairness, usability, and intervention impact.

Keywords: Learning Analytics, At-Risk Students, Explainable AI, SHAP, Large Language Models

p-ISSN 2407-4489
e-ISSN 2528-7303

INTRODUCTION

The increasing use of Learning Management Systems (LMS) has transformed higher education into a data-rich learning environment. Student interactions with materials, quizzes, assignments, forums, and feedback activities are recorded as digital traces that can support earlier identification of academic risk. In contexts where LMS adoption is growing but learning support practices remain uneven, these traces are valuable only when they can be translated into understandable and timely instructor action.

Learning Analytics and Educational Data Mining have demonstrated strong potential for predicting student performance, course completion, and dropout risk using demographic data, assessment records, LMS behavior, and engagement indicators (Alyahyan & Düştegör, 2020; Jovanović et al., 2021; Kabathova & Drlik, 2021). Recent studies have improved predictive performance and interpretability through ensemble learning, deep learning, and explainable learning analytics (Chen et al., 2022; Ghimire et al., 2024; Abukader et al., 2025). Recent systematic and bibliometric reviews also show that Learning Analytics is increasingly moving toward explainable AI and generative AI integration in higher education (Rodríguez-Ortiz et al., 2025; Talmoudi & Choukir, 2026).

A persistent limitation is that many early warning systems produce probability scores or risk labels without explaining why a student is categorized as at risk. A lecturer may know that a student has a high-risk score, but not whether the risk is mainly related to late submission, declining LMS access, weak quiz performance, limited participation, or a combination of indicators. Predictive accuracy alone is therefore insufficient for educational decision-making.

Explainable Artificial Intelligence (XAI) helps make model predictions more transparent. Techniques such as SHAP and LIME can identify learning indicators that contribute to global model behavior and individual predictions (Alamri & Alharbi, 2021; Khosravi et al., 2022; Nagy & Molontay, 2024). This emphasis on interpretability is consistent with recent reviews of explainability and model interpretability in Educational Data Mining and Learning Analytics (Gunasekara & Saarela, 2024; Carvalho et al., 2026). Nevertheless, technical explanation is not automatically pedagogical action. Instructors still need recommendations that are specific, feasible, non-stigmatizing, and connected to the identified risk factors.

Large Language Models (LLMs) can help bridge technical explanation and instructor-facing recommendation by transforming structured evidence into readable academic support language. Recent studies show that LLMs can support feedback generation, assessment, and intelligent education systems (Dai et al., 2024; Borchers et al., 2026; Ivanova & Terzieva, 2026). Their role must be controlled because LLMs should not infer motivation, psychological condition, or personal circumstances from LMS traces. In this study, the LLM is positioned as a constrained language-generation component that works from SHAP evidence and intervention rules, not as an autonomous decision-maker.

The research gap addressed in this study is the lack of an integrated framework that connects early risk prediction, explainable risk factors, controlled LLM-based recommendation, human validation, and post-intervention monitoring in one coherent workflow. Existing studies commonly emphasize one component prediction, interpretability, or feedback generation whereas educational implementation requires these components to operate together under clear ethical and pedagogical boundaries.

This study is guided by the following research questions: (1) How can an explainable and actionable AI framework be designed to support early detection of at-risk students in Learning Analytics? (2) What design requirements are needed to ensure that risk prediction, SHAP-based explanation, LLM-based recommendation, and human review operate as a responsible decision-support workflow? (3) How can instructor-facing recommendations be made operationally actionable without allowing AI to replace human educational judgment?

The specific contributions are threefold. First, the study formulates a DSR-grounded workflow that links risk prediction, local explanation, controlled recommendation, human validation, and feedback logging. Second, it operationalizes actionable AI through design requirements, risk-threshold logic, an

p-ISSN 2407-4489
e-ISSN 2528-7303

intervention taxonomy, recommendation quality criteria, and an evaluation roadmap. Third, it clarifies the ethical boundary among prediction, explanation, recommendation, and human educational judgment.

Table 1. Novelty mapping of the proposed framework against related studies

Study focus	Prediction	XAI / SHAP	LLM recommendation	Human review	Limitation of prior studies / unaddressed issue
Academic success prediction	Yes	Limited or absent	No	Limited	Strong prediction but weak intervention guidance
XAI-based dropout prediction	Yes	Yes	No	Some user study	Explains risk but does not formalize LLM-based action generation
AI feedback generation	No or indirect	No	Yes	Varies	Generates feedback but is not grounded in student risk explanations
Proposed framework	Designed	Designed	Designed	Required	Integrates prediction, explanation, action, validation, and feedback loop

Table 1 positions the novelty of the proposed framework: the integration of risk prediction, SHAP-based explanation, controlled LLM recommendation, human validation, and post-intervention monitoring into one responsible Learning Analytics workflow.

Theoretical Background and Related Work

Learning Analytics and Educational Data Mining provide the analytical foundation of this study. Learning Analytics emphasizes the collection and interpretation of learner data to improve learning processes, while Educational Data Mining focuses on computational methods for identifying patterns in educational datasets. In this paper, both fields are used to justify the transformation of LMS traces into early warning indicators for academic support.

SHAP is selected as the principal XAI method because it can support both global and local interpretation. Global SHAP values help identify which learning features are generally influential across a course, while local SHAP values explain why a specific student-course record receives a particular risk score. This dual capability makes SHAP more suitable for linking model interpretation with individual intervention planning than methods that only provide aggregate importance or less stable local explanations.

LLMs are selected as the recommendation layer because rule-based recommendation alone may be too rigid when instructors need context-sensitive, supportive, and readable intervention suggestions. In this framework, rule-based constraints and an intervention taxonomy define the permitted action space, while the LLM transforms structured SHAP evidence into a coherent instructor-facing message with an appropriate tone and follow-up indicator. Thus, the LLM remains a constrained text-generation component rather than an autonomous pedagogical decision-maker.

An at-risk student is operationally defined in this framework as a student-course record that shows a predicted probability of unsatisfactory course outcome or persistent learning disengagement based on observable LMS and assessment indicators. The term does not define the student as academically deficient; it only indicates that additional instructional attention may be needed within a particular course context.

p-ISSN 2407-4489
e-ISSN 2528-7303

RESEARCH METHOD

This study employed a Design Science Research (DSR) approach because the aim was to design an artifact that responds to a practical problem in educational technology: early warning systems that are predictive but not sufficiently explainable or actionable. DSR is suitable for this objective because it emphasizes the creation of purposeful artifacts such as constructs, models, methods, frameworks, architectures, and instantiations (Hevner et al., 2004). The methodological flow was informed by Peffers et al. (2007), covering problem identification, objective definition, design and development, demonstration, evaluation planning, and communication.

Table 2. Mapping of Design Science Research stages and study activities

DSR stage	Activity in this study	Output in the article
Problem identification	Identify score-only early warning limitations.	Problem statement and research gap.
Objective definition	Define explainable, actionable, and human-controlled AI goals.	Research questions and design objectives.
Design and development	Construct the framework, IPO model, and SHAP-LLM workflow.	Conceptual artifact and workflow architecture.
Demonstration	Illustrate a student-course scenario and controlled recommendation.	Mini case and LLM example.
Ex-ante evaluation planning	Specify criteria for future empirical testing.	Evaluation rubric and future study design.
Communication	Present the DSR artifact for learning analytics audiences.	Journal manuscript.

Within the DSR process, the artifact specification was developed by translating the identified educational problem into design requirements, component boundaries, workflow interactions, and evaluation criteria for future instantiation using LMS data.

The unit of analysis in the proposed framework is the student-course level. This means that risk is assessed for individual students within a specific course during an ongoing semester. This level is selected because instructors typically need course-specific information to decide whether a student requires remedial material, consultation, feedback, assignment recovery, or monitoring. The framework can later be extended to semester-level or institutional-level monitoring, but the present design focuses on supporting course-level instructor decisions.

Table 3. Examples of raw LMS data and engineered learning features

Raw data source	Engineered feature	Educational interpretation
Login records	Number of LMS logins per week	General digital engagement within a course week
Resource access logs	Percentage of required materials accessed	Exposure to course learning resources
Assignment records	Average submission delay in days	Punctuality and task management behavior
Quiz results	Cumulative quiz score and score trend	Early mastery of course content
Forum or discussion logs	Number of meaningful posts or replies	Participation in collaborative learning activities
Weekly activity history	Declining activity over consecutive weeks	Possible disengagement requiring instructor review

The early detection window is designed as a weekly prediction process beginning from the third week of a course. Weeks one and two are treated as an initial data collection period, while week three

p-ISSN 2407-4489
e-ISSN 2528-7303

is the earliest recommended point for generating an initial risk signal. The prediction can then be repeated weekly until mid-semester and beyond. This design allows instructors to act while there is still sufficient time for academic support, rather than waiting until final grades reveal failure.

The framework is designed for periodic weekly monitoring rather than real-time surveillance. This design is more realistic for higher education because instructors usually review student progress at weekly or module-based intervals. Periodic monitoring also reduces unnecessary intrusion into student activity while still allowing timely academic support.

Design Requirements

The design requirements were selected through a problem-to-requirement logic. The literature gap indicates that prediction-only systems lack transparency and intervention guidance; therefore, the framework requires early risk detection, local explainability, actionable recommendation, human review, ethical safeguards, post-intervention feedback, and evaluation readiness. Each requirement is linked to a concrete educational risk: late detection, opaque prediction, generic advice, automated judgment, privacy or fairness harm, lack of learning from intervention outcomes, and weak empirical validation.

These requirements also function as design constraints. A component is included in the framework only when it helps transform LMS traces into responsible instructor support. For example, SHAP is included to make risk factors visible, the LLM is constrained to convert those factors into readable recommendation drafts, and human-in-the-loop validation is required to prevent automatic action on students without contextual review.

Table 4. Design requirements of the proposed framework

Code	Design requirement	Operational meaning
DR1	Early risk detection	Generate weekly risk signals from the early semester.
DR2	Local explainability	Provide student-specific SHAP-based risk factors.
DR3	Actionable recommendation	Produce specific, feasible, time-bound, and risk-aligned recommendations.
DR4	Human-in-the-loop	Require instructor validation before any student-facing action.
DR5	Ethical data use	Apply data minimization, privacy safeguards, fairness checks, and support-oriented language.
DR6	Post-intervention feedback	Log actions and outcomes to support monitoring and continuous improvement.
DR7	Evaluation readiness	Define prototype validation metrics.

Input-Process-Output Specification

The proposed framework is structured using an input-process-output logic. Input refers to data sources that can be ethically collected from LMS and academic systems. Process refers to computational and pedagogical mechanisms that transform the data into interpretable risk information. Output refers to decision-support artifacts reviewed by instructors. Table 5 shown Input-process-output specification of the framework.

Risk Threshold and Prediction Error Consideration

The framework requires transparent risk thresholds. In future implementation, thresholds may be determined using historical data, institutional policy, or validation against actual course outcomes. As an initial design, low risk may be assigned when the predicted probability of failure is below 0.40, medium risk when it is between 0.40 and 0.69, and high risk when it is 0.70 or above (Table 6). These thresholds should not be treated as universal; they must be calibrated for each institutional context.

p-ISSN 2407-4489
e-ISSN 2528-7303

Table 5. Input-process-output specification of the framework

Layer	Input	Process	Output
Data acquisition	LMS clicks, resource access, forum activity, quiz scores, assignments, academic status	Ethical data selection and minimization	Course-level learning analytics dataset
Feature engineering	Raw learning traces	Weekly aggregation, trend detection, punctuality measures, assessment progress indicators	Pedagogically meaningful features
Risk prediction	Engineered features	Logistic Regression, Random Forest, XGBoost, or LSTM depending on data structure	Risk probability and risk category
Explanation	Model output and features	SHAP global and local explanation	Top risk factors and protective factors
Recommendation	Risk category, SHAP factors, feature values, prompt constraints	Controlled LLM generation and intervention taxonomy mapping	Instructor-facing explanation and intervention options
Human review	AI-generated recommendation	Accept, edit, reject, request additional information, and log decision	Validated intervention and monitoring record

Prediction errors must also be considered. A false positive occurs when the system flags a student as at risk although the student is not actually at risk. This may lead to unnecessary concern or stigmatization if the recommendation is not handled carefully. A false negative occurs when the system fails to identify a student who genuinely needs support. This is more serious in early warning contexts because the opportunity for intervention may be missed. For this reason, recall and F1-score should be prioritized in future empirical testing, while all outputs should be framed as support signals rather than definitive judgments.

Table 6. Risk threshold and response design

Risk level	Indicative probability	Recommended response	Decision rule
Low	< 0.40	Routine monitoring and regular feedback	No additional intervention unless instructor observes concern
Medium	0.40-0.69	Targeted reminder, formative feedback, and short learning support	Instructor reviews risk factors before action
High	>= 0.70	Personalized message, consultation invitation, remedial task, and weekly follow-up	Instructor validation is required before contacting student

Controlled LLM Recommendation Design

The framework selects SHAP as the explanation bridge because it generates feature-level evidence that can be directly converted into recommendation constraints. For example, if the highest SHAP factors are late submission and low quiz performance, the recommendation layer can be instructed to focus only on time-management support and remedial learning, rather than producing broad or unsupported advice.

The LLM is used because the intended user is not a data scientist but an instructor or academic advisor who needs clear pedagogical language. However, the LLM is deliberately constrained through

p-ISSN 2407-4489
e-ISSN 2528-7303

structured input, prompt rules, non-stigmatizing language requirements, and human validation. This design prevents the LLM from becoming a free-form decision-maker.

The LLM component is designed as a constrained language-generation layer. It does not determine student risk, diagnose motivation, or infer psychological condition. It only receives structured information from the prediction and SHAP layers and transforms it into an instructor-facing recommendation. The prompt must explicitly prohibit unsupported inference and require supportive, non-stigmatizing language.

Table 7. Example of controlled LLM input and expected output

Element	Example
Structured input	Risk level: High; risk probability: 0.76; top SHAP factors: late assignment submission, declining LMS access, low quiz score; unit: student-course; week: 4
Prompt constraint	Use only the provided indicators. Do not infer laziness, motivation, mental health, family problems, or personal causes. Use supportive academic language.
Expected explanation	The student may need academic support because recent LMS activity has declined, one assignment was submitted late, and the quiz score is below the course expectation.
Expected intervention	Send a supportive message within two days, provide one remedial exercise related to the quiz topic, invite the student to a 15-minute consultation, and monitor LMS activity and assignment submission next week.
Monitoring indicator	Improved LMS access, remedial completion, next quiz score, assignment punctuality, and instructor decision log.

Demonstration and Ex-Ante Evaluation Plan

The demonstration stage illustrates how the framework would operate in a course. In week four, the system aggregates LMS and assessment data, produces a risk score, identifies local SHAP factors, generates a constrained recommendation, and presents it to the instructor. The instructor may accept, edit, reject, or delay the recommendation. After the intervention, the outcome is recorded and used for future evaluation.

Future empirical evaluation should be conducted in staged form rather than through a single accuracy test. The first stage should validate data readiness and feature construction using historical LMS records. The second stage should train and compare baseline prediction models, calibrate risk thresholds, and examine false positive and false negative patterns. The third stage should evaluate SHAP explanations and LLM-based recommendation drafts through instructor review. The fourth stage should test usability, workload, privacy safeguards, and fairness across relevant student groups. The final stage should examine whether validated interventions improve engagement, assignment completion, and course outcomes.

This staged evaluation is necessary because the artifact has multiple claims: it predicts risk, explains model behavior, supports pedagogical action, protects students, and enables follow-up monitoring. A framework that performs well on ROC-AUC but produces vague recommendations, biased errors, or excessive lecturer workload would not satisfy the design objective. Therefore, evaluation criteria must include technical, pedagogical, ethical, and operational dimensions.

Table 8 summarizes the proposed ex-ante and empirical evaluation criteria. In the present study, these criteria are presented as a roadmap for future implementation, not as completed empirical findings. They provide the basis for testing the framework with real LMS datasets, instructor validation, and intervention outcome analysis.

p-ISSN 2407-4489
e-ISSN 2528-7303

Table 8. Evaluation plan for future prototype and empirical validation

Evaluation dimension	Criteria or metrics	Evidence source
Predictive performance	Recall, F1-score, ROC-AUC, confusion matrix, calibration, and threshold sensitivity.	Historical LMS records with course outcomes.
Explanation quality	Instructor-rated clarity, relevance, usefulness, and consistency of local SHAP factors.	Expert review, instructor questionnaire, and case inspection.
Recommendation quality	Specificity, SHAP alignment, feasibility, supportive tone, time frame, and monitorability.	Rubric-based evaluation of LLM recommendation drafts.
Usability and workload	Ease of interpretation, decision time, perceived usefulness, and lecturer workload.	Prototype testing, SUS-style survey, and interview notes.
Fairness and privacy	Subgroup error rates, data minimization, access control, audit logs, and non-stigmatizing language.	Fairness audit and data-governance checklist.
Intervention impact	Changes in LMS access, remedial completion, assignment punctuality, quiz recovery, and course outcome.	Pre-post course data and validated instructor decision logs.

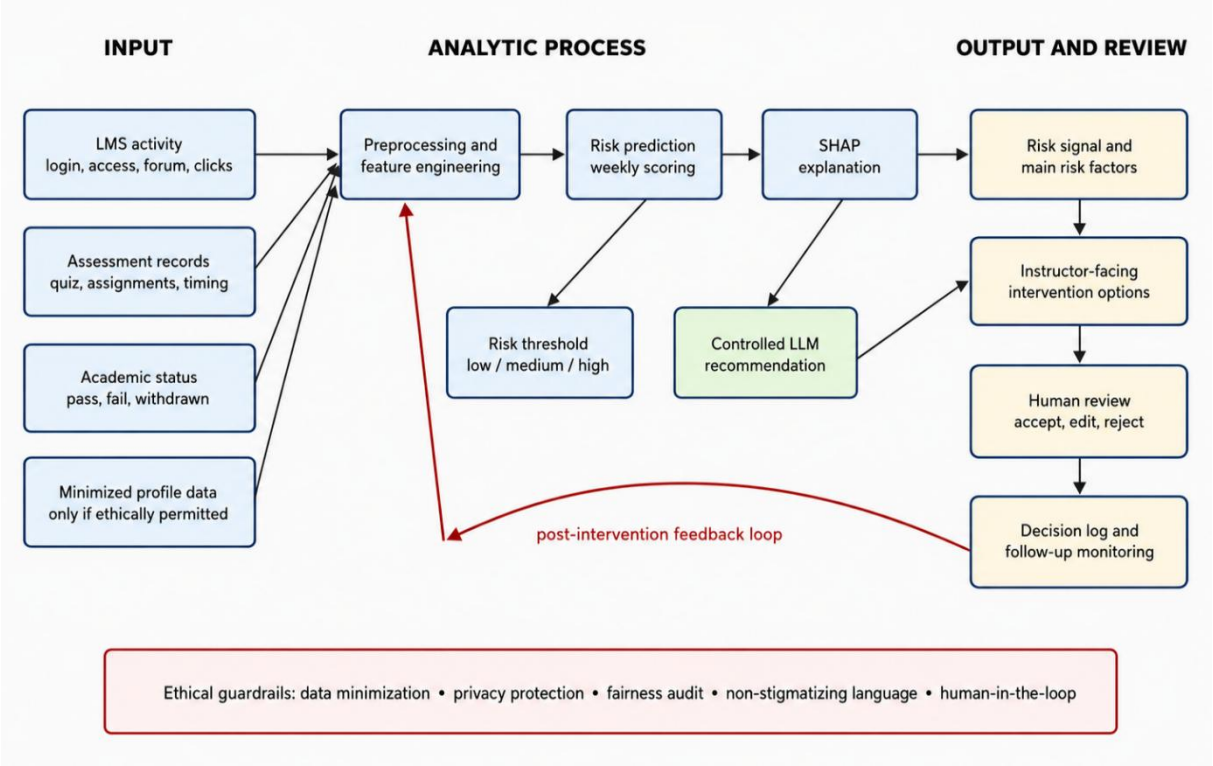


Fig 1. Input-process-output architecture of the proposed framework, showing the flow from LMS data, risk prediction, SHAP explanation, and controlled LLM recommendation to instructor validation and post-intervention feedback.

p-ISSN 2407-4489
e-ISSN 2528-7303

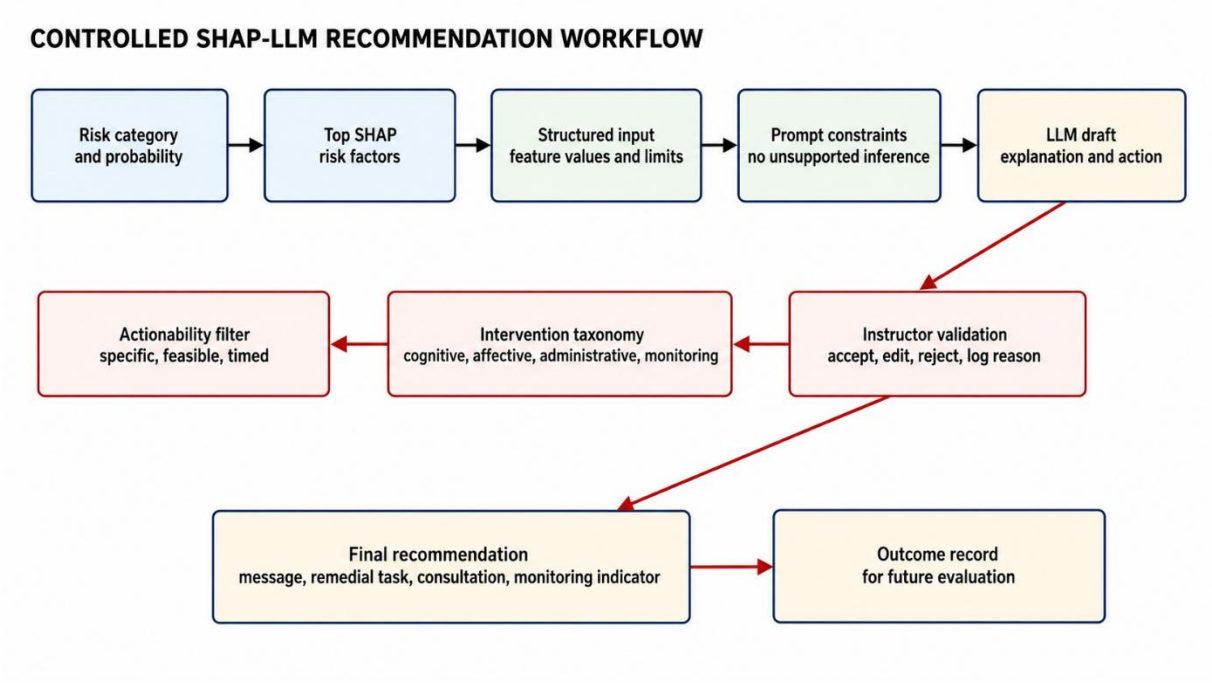


Fig 2. Controlled SHAP-LLM recommendation workflow, showing how risk factors are transformed into constrained recommendations, filtered through intervention taxonomy, validated by instructors, and logged for follow-up evaluation.

RESULT AND DISCUSSION

Framework Design Results

The first result is the DSR-grounded artifact definition and analytical justification: a workflow architecture that transforms LMS data into risk signals, SHAP explanations, constrained recommendations, instructor decisions, and post-intervention monitoring records.

The second result is the set of design requirements shown in Table 4. These requirements strengthen the design logic by showing why each component is needed and what educational problem it addresses. The framework therefore does not appear as an arbitrary combination of technologies, but as a response to limitations in score-only early warning systems.

The third result is the input-process-output specification shown in Table 5 and visualized in Figure 1. This specification makes clear the boundary between data input, analytic processing, and decision-support output. Such boundary is important because an AI system should not be allowed to move directly from raw data to educational action without explanation and human review.

The fourth result is the controlled SHAP-LLM workflow shown in Figure 2 and Table 7. This workflow is important because it limits the LLM to language transformation based on structured model evidence. The LLM is not permitted to infer personal causes beyond the available learning analytics indicators, and every recommendation must pass through instructor validation before use.

Mini Demonstration Scenario

To make the framework more concrete, Table 9 presents an illustrative student-course case. The example does not represent an empirical result; it is a demonstration of how risk indicators, SHAP explanations, LLM constraints, and instructor validation may operate in one workflow.

p-ISSN 2407-4489
e-ISSN 2528-7303

Table 9. Illustrative mini scenario of framework output

Component	Illustrative content
Student-course profile	Week 4 record in an introductory programming course; risk probability = 0.78; category = high risk
Main SHAP factors	Late assignment submission (+0.21), declining LMS access for two weeks (+0.18), low quiz score (+0.16)
Controlled LLM recommendation	Send a supportive message, provide a short remedial exercise on the quiz topic, invite the student to a 15-minute consultation, and review LMS activity in week 5.
Example instructor message	"I noticed that some recent learning activities may have been difficult to complete. Please review the short remedial exercise I have attached, and let me know whether you can meet briefly this week to discuss the assignment."
Monitoring indicator	Assignment resubmission completed, LMS access increases, and quiz recovery task submitted within one week

Actionable Recommendation and Intervention Taxonomy

In this study, an actionable recommendation is operationally defined as a recommendation that is specific, aligned with identified risk factors, feasible for instructors, non-stigmatizing, time-bound, and monitorable. A recommendation such as “help the student study harder” is not actionable because it is vague and difficult to monitor. In contrast, a recommendation such as “send a supportive message within two days, provide one remedial exercise related to the low quiz topic, and check whether the student completes it before the next meeting” is actionable because it specifies actor, action, time frame, and monitoring indicator.

Table 10. Taxonomy of pedagogical interventions

Intervention type	Risk indicator	Example action	Monitoring indicator
Cognitive	Low quiz score or weak assessment performance	Provide remedial material, short practice task, worked example, or mini-review session	Remedial completion, next quiz score, error reduction
Affective-supportive	Sudden decline in LMS engagement or repeated inactivity	Send supportive message and invite student to discuss learning difficulty without blame	Student response, renewed LMS activity, consultation attendance
Administrative	Late or missing assignment submission	Clarify deadline, provide recovery path, and explain submission requirement	Assignment resubmission, punctuality in next task
Monitoring	Medium risk or uncertain pattern	Observe weekly activity and compare with prior engagement trend	Weekly LMS access, forum participation, assessment progress

Table 10 strengthens the pedagogical foundation of the framework. The intervention taxonomy prevents the LLM from producing generic advice and ensures that recommendations are linked to risk indicators. It also supports instructor validation because the instructor can select, revise, or reject an intervention category based on contextual knowledge of the student and course.

Comparison with Existing Learning Analytics Frameworks

The proposed framework differs from conventional learning analytics models in the level of decision support it provides. Prediction-oriented models are useful for identifying students who may be at risk, but they often leave instructors with score interpretation and intervention design tasks. XAI-oriented models improve transparency by showing influential features, yet they may still stop at

p-ISSN 2407-4489
e-ISSN 2528-7303

explanation. AI feedback systems can generate readable messages, but they may not be grounded in individual risk explanations. The proposed framework attempts to connect these three layers while preserving human educational judgment.

This comparison is important because the value of the artifact should not be judged only by whether it predicts risk. Its practical value lies in whether the prediction is understandable, whether the explanation can be converted into a feasible intervention, and whether the intervention remains ethically and pedagogically controlled.

Table 11. Critical comparison with existing learning analytics approaches

Existing approach	Typical strength	Key limitation	Response in proposed framework
Prediction-oriented early warning	Identifies students with possible academic risk.	Often provides scores or labels without sufficient explanation or action guidance.	Adds SHAP explanation, risk-threshold logic, and instructor-facing intervention options.
XAI-based learning analytics	Explains influential features behind a model prediction.	Explanation may remain technical and may not become pedagogically actionable.	Connects local SHAP factors to an intervention taxonomy and LLM-generated recommendation draft.
AI feedback generation	Produces readable feedback or support messages.	May not be grounded in validated student-risk evidence or instructor review.	Constrains LLM output using risk level, SHAP factors, prompt rules, and human validation.
Human advisory practice	Uses contextual judgment and personal knowledge of students.	May be reactive, inconsistent, or limited by time and incomplete data.	Provides structured evidence while keeping instructors responsible for final decisions.

Table 11 shows that the proposed contribution is integrative rather than algorithmic. It does not claim superiority through a new classifier; instead, it specifies how prediction, explanation, recommendation, validation, and outcome logging can be organized into a responsible decision-support workflow.

Post-Intervention Feedback Loop

The framework includes a post-intervention feedback loop. After the instructor validates and applies an intervention, the outcome should be logged into the system. The record may include the action taken, date of intervention, whether the student responded, whether remedial work was completed, and whether LMS activity or assessment performance improved in the following week. This feedback loop is important because it transforms the framework from a one-directional warning system into a learning system that can be evaluated and improved over time.

The feedback loop also supports institutional learning. Aggregated decision logs can help identify which interventions are frequently used, which risk factors are most common, and which support strategies appear useful in specific courses. However, such records must be governed carefully to avoid stigmatizing students or creating permanent negative labels. The purpose of logging is improvement of academic support, not surveillance or punishment.

Theoretical and Practical Implications

Theoretically, this framework extends the concept of Learning Analytics from prediction-oriented analytics toward explainable and actionable decision support. It connects three research streams: academic risk prediction, XAI-based explanation, and LLM-supported feedback generation. In doing

p-ISSN 2407-4489
e-ISSN 2528-7303

so, the framework argues that educational analytics should not stop at identifying patterns but should help educators interpret those patterns responsibly and translate them into pedagogical action.

Practically, the framework can help instructors monitor students from the early weeks of a course, prioritize academic support, and provide interventions based on specific learning indicators. Academic advisors may use the framework to complement course-level information with broader academic guidance, while institutions may use aggregated insights to improve early warning policies. This practical implication is aligned with learning analytics feedback research showing that feedback usefulness depends not only on the quality of information but also on how users interpret and act on feedback (Weidlich et al., 2025). In the Indonesian higher education context, the framework may be useful for institutions seeking to strengthen LMS-based academic monitoring without removing the central role of lecturers and academic advisors

Implementation Challenges in Real Educational Environments

Several implementation challenges must be anticipated before the framework is deployed. First, LMS data quality may differ across courses because lecturers use platforms in different ways; some courses may have rich weekly activity traces, while others may only record assessment submissions. Second, early-semester data may be sparse, creating uncertainty in week-three or week-four predictions. Third, at-risk cases are often imbalanced, so a model may appear accurate while still missing students who need support.

Operational challenges are equally important. Instructors may not have time to review long dashboards, academic advisors may need role-based access rules, and institutions must define who is responsible for acting on medium- and high-risk signals. LLM use also creates risks of generic wording, hallucinated causes, or inappropriate tone if prompts are not constrained and outputs are not validated by instructors.

For these reasons, future implementation should begin with a limited pilot course or program, use only ethically permitted LMS features, calibrate thresholds with historical data, keep recommendations short and editable, maintain audit logs, and evaluate whether the system actually reduces lecturer workload rather than adding a new administrative burden.

Ethical Design and Boundary of AI Roles

The framework distinguishes prediction, recommendation, and decision. The risk prediction model only provides a probability-based signal. SHAP explains the learning indicators that contribute to that signal. The LLM converts structured information into a draft explanation and intervention option. The final decision remains with the instructor or academic advisor. This separation of roles is crucial because educational decisions affect students and must not be delegated entirely to automated systems.

Ethical design is embedded in the framework through data minimization, privacy protection, fairness audit, non-stigmatizing language, and human-in-the-loop review. Data minimization requires the system to use only indicators relevant to academic support. Fairness audit requires future empirical testing to examine whether prediction errors disproportionately affect certain student groups. Non-stigmatizing language prevents the system from labeling students as lazy, incapable, or problematic. Human review ensures that contextual judgment remains part of the intervention process.

Data privacy safeguards should be technical as well as procedural. Student records should be anonymized or pseudonymized during model development, dashboard access should be limited to authorized instructors or advisors, system access logs should be maintained, and individual recommendations should not be shared beyond parties responsible for academic support. These safeguards help ensure that learning analytics is used for student support rather than surveillance.

p-ISSN 2407-4489
e-ISSN 2528-7303

Operationally, the dashboard should be accessible only to authorized course instructors, academic advisors, and designated learning support staff. During model development, student identifiers should be pseudonymized, access logs should be stored, recommendation histories should be auditable, and no intervention message should be automatically sent to students without instructor confirmation. Risk labels should also remain temporary course-level signals rather than permanent records in the academic information system.

To prevent stigmatization, the system should avoid deficit-based labels such as "problematic student" or "failing student". Instructor-facing messages should use support-oriented language, for example "student who may need additional academic support" or "student requiring follow-up review". This communication principle is essential because risk detection in education directly affects how students are perceived and treated.

The framework also recognizes the interpretive limits of LMS data. Low LMS activity should not automatically be interpreted as low ability or low motivation because students may use external learning resources, face internet limitations, experience device problems, or encounter personal constraints that are not captured by the system. Therefore, LMS traces are treated as behavioral signals that require instructor interpretation, not as complete representations of student capability.

False Positive, False Negative, and Implementation Sustainability

Prediction errors require careful treatment. False positives may lead to unnecessary intervention or discomfort if students are labeled as at risk without sufficient reason. False negatives may be more damaging because students who need help may be overlooked. Therefore, the framework recommends using risk labels as support signals rather than final judgments. Instructors should receive both the risk level and its explanation so they can determine whether intervention is appropriate.

Sustainable implementation also requires institutional roles. LMS administrators should manage data access and system integration. Course instructors should validate recommendations and implement interventions. Academic advisors may support students whose risk persists across courses. Institutional quality assurance teams should monitor model fairness, privacy compliance, and evaluation results. Without clear role distribution, an AI framework may remain technically attractive but difficult to sustain in everyday academic practice.

Model maintenance is also necessary because learning behavior can change across semesters, cohorts, course designs, and LMS usage policies. Future implementation should include periodic model retraining, monitoring of performance drift, and re-evaluation of fairness across student groups. Without maintenance, a model that performs well in one semester may become less reliable in another context.

Overall, the proposed framework should be understood as an instructor-support mechanism rather than an automated judgment system. Its purpose is to help educators read early academic risk signals more quickly, interpret the contributing factors more transparently, and respond with evidence-based yet humane intervention strategies.

Threats to Validity

Several threats to validity should be acknowledged. Data validity may be affected because LMS traces do not always represent the full learning process; students may study offline or use external materials not captured by the system. Model validity may be affected by class imbalance, changing course design, and inconsistent assessment practices. Explanation validity may be affected when SHAP values are interpreted as causal explanations, although they only indicate model contribution. Recommendation validity may be affected when LLM outputs are too generic or insufficiently aligned with pedagogical context. Implementation validity may be limited because institutions differ in LMS usage, data governance, lecturer workload, and student support systems.

p-ISSN 2407-4489
e-ISSN 2528-7303

These validity threats do not invalidate the framework. Instead, they define the evidence required before operational deployment, including real LMS datasets, instructor evaluation, and intervention outcome analysis.

CONCLUSION

This study proposed a Design Science Research-based conceptual framework for early detection of at-risk students in Learning Analytics. The artifact is a conceptual framework and workflow architecture that integrates LMS data, learning feature engineering, risk prediction, SHAP-based explanation, controlled LLM recommendation, instructor validation, and post-intervention feedback. The framework is designed to answer three practical educational questions: who may be at risk, why the risk appears, and what intervention can be considered by instructors or academic advisors.

The main contribution lies in articulating a bridge between predictive analytics and responsible pedagogical action. Unlike prediction-only approaches, the framework explicitly connects risk identification with explanation, recommendation, human review, ethical safeguards, and outcome monitoring. The novelty is therefore the integrated design of SHAP, constrained LLM recommendation, human-in-the-loop validation, and feedback logging within one support-oriented learning analytics workflow.

The limitation of this study is that it does not yet report implemented model performance, SHAP plots, instructor validation results, or measured intervention impact. A clear future roadmap is therefore required. The next stage should build a working prototype using an institutional LMS dataset or a public dataset such as OULAD, implement data governance and pseudonymization, train and compare baseline models, generate SHAP explanations, connect the outputs to a constrained LLM recommendation module, and present the results through an instructor dashboard.

Subsequent empirical validation should include offline model evaluation, instructor evaluation of explanations and recommendations, usability testing of the dashboard, fairness and privacy audits, and a pilot intervention scenario in selected courses. The pilot should compare student engagement and assessment recovery before and after instructor-validated interventions. Through this roadmap, the proposed framework can move from a conceptual design toward a prototype that is not only predictive, but also explainable, ethically constrained, actionable, and useful for timely academic support.

ACKNOWLEDGMENT

The author acknowledges academic feedback received during the development of this conceptual framework. This study received no specific external funding.

REFERENCES

Abukader, A., Alzubi, A., & Adegboye, O. R. (2025). Intelligent system for student performance prediction: An educational data mining approach using metaheuristic-optimized LightGBM with SHAP-based learning analytics. *Applied Sciences*, *15*(20), 10875. https://doi.org/10.3390/app152010875

Alamri, R., & Alharbi, B. (2021). Explainable student performance prediction models: A systematic review. *IEEE Access*, *9*, 33132-33143. https://doi.org/10.1109/ACCESS.2021.3061368

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*, 3. https://doi.org/10.1186/s41239-020-0177-7

Borchers, C., Thomas, D. R., Lin, J., & Koedinger, K. R. (2026). Learning-aware reliability estimation for tutor skill assessment using large language models. *Journal of Learning Analytics*, 1-20. https://doi.org/10.18608/jla.2026.9133

Carvalho, C. S., Mattos, J. C. B., & Aguiar, M. S. (2026). Model interpretability in the educational data mining context: A systematic literature mapping. *Revista Brasileira de Informática na Educação*, *34*, 314-355. https://doi.org/10.5753/rbie.2026.6984

Chen, H. C., Prasetyo, E., Tseng, S. S., Putra, K. T., Prayitno, Kusumawardani, S. S., & Weng, C. E. (2022). Week-wise student performance early prediction in virtual learning environment using a deep explainable artificial intelligence. *Applied Sciences*, *12*(4), 1885. https://doi.org/10.3390/app12041885

Dai, W., Tsai, Y. S., Lin, J., Aldino, A., Jin, H., Li, T., Gašević, D., & Chen, G. (2024). Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. *Computers and Education: Artificial Intelligence*, *7*, 100299. https://doi.org/10.1016/j.caeai.2024.100299

Ghimire, S., Abdulla, S., Joseph, L. P., Prasad, S., Murphy, A., Devi, A., Barua, P. D., Deo, R. C., Acharya, U. 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

Gunasekara, S., & Saarela, M. (2024). Explainability in educational data mining and learning analytics: An umbrella review. In *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 887-892). International Educational Data Mining Society.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75-105. https://doi.org/10.2307/25148625

Ivanova, T., & Terzieva, V. (2026). Large language models in intelligent education systems: New educational perspectives: A systematic review. *Information*, *17*(5), 433. https://doi.org/10.3390/info17050433

Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, *172*, 104251. https://doi.org/10.1016/j.compedu.2021.104251

Kabathova, J., & Drlik, M. (2021). Towards predicting student dropout in university courses using different machine learning techniques. *Applied Sciences*, *11*(7), 3130. https://doi.org/10.3390/app11073130

Khosravi, H., Buckingham Shum, S., 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

Nagy, M., & Molontay, R. (2024). Interpretable dropout prediction: Towards XAI-based personalized intervention. *International Journal of Artificial Intelligence in Education*, *34*, 274-300. https://doi.org/10.1007/s40593-023-00331-8

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45-77. https://doi.org/10.2753/MIS0742-1222240302

Rodríguez-Ortiz, M. Á., Santana-Mancilla, P. C., & Anido-Rifón, L. E. (2025). Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges. *Applied Sciences*, *15*(15), 8679. https://doi.org/10.3390/app15158679

Talmoudi, R., & Choukir, J. (2026). From predictive analytics to explainable AI in higher education: A bibliometric mapping. *Qubahan Academic Journal*, *6*(2), 437-461. https://doi.org/10.48161/qaj.v6n2a2559

Weidlich, J., Fink, A., Frey, A., Jivet, I., Gombert, S., Menzel, L., Giorgashvili, T., Yau, J., & Drachsler, H. (2025). Highly informative feedback using learning analytics: How feedback literacy moderates

p-ISSN 2407-4489
e-ISSN 2528-7303

student perceptions of feedback. *International Journal of Educational Technology in Higher Education*, *22*, Article 43. https://doi.org/10.1186/s41239-025-00539-9