Received: 9 April 2025 | Revised: 4 January 2026 | Accepted: 18 June 2026 | Published online: 24 July 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–19
DOI: 10.47852/bonviewAIA62025883

BON VIEW PUBLISHING

XAI-Profile: Explainable AI for Transparent Learner Profiling in Digital Education

Abdelkader Ouared[1,*], Madeth May[1], Claudine Piau-Toffolon[1] and Nicolas Dugué[1]

[1] *LIUM Computer Science Laboratory, University of Le Mans, France*

Abstract: Trust-aware digital learning systems are essential for informed decision-making by educational stakeholders, as the growing complexity of digital learning environments increases the need for transparent and explainable learner profiles. While digital learning traces are crucial for identifying learner behaviors and categories, extracting meaningful profiles from large and complex data remains challenging due to bias, misinterpretation, and limited stakeholder guidance. Moreover, traditional learning analytics tools struggle to transform low-level traces into human-interpretable insights. In response to these shortcomings, we introduce a novel framework that exploits explainable artificial intelligence (XAI) to convert digital learning traces into meaningful and behavioral information. By integrating explainable machine learning techniques with human-centered design process, our framework connects fine-grained learning traces to high-level pedagogical interpretations. XAI-Profile is structured into three main components: (1) goal-driven requirements to decompose stakeholders' goals into measurable subgoals linked to learner data, (2) visual analytics design to interpret learner profiles through transparent analytical artifacts, and (3) trust flow to support hypothesis testing and actionable, context-aware insights. Experimental evaluation in the écri+ project demonstrates that our framework generates interpretable learner profiles and validates pedagogical hypotheses. Educator-in-the-loop assessment confirms improved transparency and trust through interactive visual analytics. This work bridges XAI outputs with pedagogical insight, establishing a scalable, human-centered foundation for learning analytics.

Keywords: learner profile, learning analytics, explainable AI, transparency, human-centered analytics

1. Introduction

Trust-aware digital learning systems enable informed decision-making by educational stakeholders, driven by the need for transparent and explainable learner profiles (LPs) and Learning Analytics Dashboards (LADs) (e.g., [1, 2]). At the same time, augmented learning analytics has gained attention for enhancing trust in artificial intelligence (AI)-driven education, addressing challenges such as system exploitation, bias, and misinterpretation. Ethical and stakeholder-related concerns further emphasize the need for fairness and transparency in educational AI systems [3]. In this context, explainable AI (XAI) offers promising solutions for reducing bias and improving interpretability in Learning Analytics applications, particularly for LP analysis across engagement, performance, and outcomes [4, 5]. From a linguistic perspective, a *profile* refers to a concise characterization of a person or entity that highlights relevant and informative attributes. In educational contexts, an LP is commonly understood as a category of learners defined by a set of features related to learning organization (e.g., persistence and effort management), engagement (e.g., active vs. inactive participation), and motivation (e.g., learners in progress, success-oriented learners, or those at risk). LADs support LP analysis through progress tracking, outcome evaluation, and visualization [6]. Student behavior analysis focuses on engagement, success or failure, and learning outcomes (e.g., [6–8]), while AI-based approaches improve LP analysis and success prediction [9–11]. Behavior-oriented LADs often rely on Deep Learning (DL)/machine learning (ML) techniques for monitoring learners [1, 12–14], but their lack of transparency limits interpretability. Recent work addresses this issue through explainable and transparent models for LP analysis [15–18]. Other studies explore adaptive analytics for self-regulated learning [19], visual analytics for student–ChatGPT interactions [20], and visual data mining for performance prediction [21]. Although interpretable prediction frameworks exist [22, 23], they provide limited user involvement and actionable guidance. Conceptual modeling and Ontological Engineering offer rich learner representations [24–28], yet bridging digital traces with conceptual insights remains challenging.

Traditional LADs offer static visualizations but lack algorithmic transparency, while AI-based systems provide predictive power without explanatory depth. Recent explainable LADs introduce transparency but often isolate explanations from a holistic pedagogical workflow, failing to keep educators actively in the loop. Crucially, existing approaches do not systematically connect stakeholder hypotheses to data features, nor do they close the feedback loop between interpretable models and educator validation. Our framework addresses these gaps by establishing an end-to-end, hypothesis-driven pipeline from pedagogical requirements through transparent feature engineering to interactive, explainable visual analytics. This integrated approach uniquely

***Corresponding author:** Abdelkader Ouared, LIUM Computer Science Laboratory, University of Le Mans, France. Email: abdelkader.ouared@univ-lemans.fr

supports the discovery, validation, and actionable interpretation of LPs, bridging low-level traces and high-level pedagogical insight within a trusted, educator-centered workflow.

Despite these limitations, our research provides a solid foundation for defining and explaining LPs for educational stakeholders. An LP involves multiple factors, including functional needs, dataset features, source entities and attributes, trust-aware representations, data indicators of task and process level, ML methods, learning achievements, and sociodemographic context. Existing studies rarely consider all these components together; however, their integration is essential for achieving explainable and trustworthy LP [29]. Although research has improved LP transparency, prior work often overlooks the reasoning behind insights, especially the end-to-end link from stakeholder objectives to interpretable ML models. There is a lack of frameworks that clearly explain LPs, categorize them, and justify their behaviors. Few tools support stakeholders in analyzing LPs through common educational concepts, leaving a gap between data level and interpretable representations of learners. These elements represent real-world observations and pedagogical practices rooted in widely accepted educational principles. To address this gap, data scientists need to ensure that ML insights correspond to stakeholder requirements by carefully selecting relevant features, emphasizing interpretability, reducing biases, and verifying that identified patterns are consistent with domain knowledge. Figure 1 illustrates three interaction modes between stakeholders and the LAD. In the first case, the LAD operates as a black box, providing aggregated data and ML-based alerts (left). In the second, outcomes are produced without explanation, limiting interpretation and stakeholder confidence (center). The third mode addresses these limitations by offering interpretable models that explain LPs, enhance transparency, and support interactive exploration, enabling stakeholders to uncover meaningful patterns and refine decisions (right).

To address the challenges of LP analysis, both academia and industry have introduced behavioral LADs, such as Learning Analytics for Moodle [30]. However, existing LADs often embed algorithmic biases and rely on predefined indicators, patterns, and visualizations that limit meaningful stakeholder interaction and generalizability across contexts. Moreover, LP analysis frequently lacks clear explanations due to weak links between educational goals, data features, and internal ML representations [30]. As a result, identifying LPs from large-scale data remains error-prone, while stakeholders struggle to formulate relevant queries or uncover meaningful patterns, reducing transparency and trust in behavioral insights.

For these reasons, a new approach is required for LP analysis, one that operates at a higher level of abstraction and connects fine-grained digital traces with conceptual representations to make learner states understandable to humans. Accordingly, this work is driven by the following research question: ***RQ***: *How can learner profiles be made more transparent and intelligible?*

To overcome the limitations related to transparency and intelligibility in LP analysis, we propose XAI-Profile (Explainable Learner Profile), a solution based on XAI to support interpretable LP analysis. As summarized in Table 1, existing LADs focus on activity monitoring, predictive explainability, or conceptual data modeling but fail to integrate reasoning, explainable models, and stakeholder involvement into an end-to-end workflow, a gap addressed by our explainable LAD framework.

XAI-Profile provides transparent explanations of LPs grounded in historical data. The framework is structured around three core components:

1) Goal-driven requirements, which capture stakeholders' needs by decomposing educational problems into subgoals and linking them to LPs and relevant data indicators.

2) Profile catalog and visual analytics design, which support the exploration and customization of LPs through structured categories of profiles and indicators and support the interpretation of LPs by associating indicators with visualization objectives, analysis types, and suitable visual representations.

3) Trust flow, which provides an interactive, dialogue-driven process that guides stakeholders in exploring, validating, and refining LPs. Through meaningful interactions, users can interpret ML model outputs, bridge the gap between educational concepts and data features, and uncover relevant or unexpected patterns, thereby supporting informed and trustworthy decision-making.

Within the XAI-Profile framework, we rely on abstract, interpretable model representations such as learned rules, influential features, predicted classes, and profile categories derived from transparent and explainable methods. These representations are presented to end users through interactive mechanisms, enabling

Figure 1
LP interaction modes between educational stakeholders and the LAD

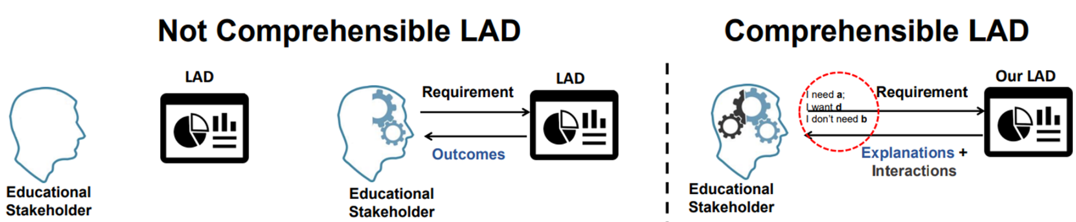


Table 1
Synthesis and positioning of LADs

LAD type	Focus	Gap
Traditional LADs [1]	Activity monitoring	Static indicators, limited reasoning
AI-based LADs [15]	Prediction and explainability	Model-centric, low user involvement
Ontology-based LADs [24]	Conceptual data modeling	Weak link to explainable models
XAI-Profile LAD	End-to-end explainable LAD	Integrates objectives, data, models, and users

a clearer understanding of the decision logic and fostering trust in the resulting LP analyses.

A case study based on real-world data from the écri+[1] project, a nationally funded ANR initiative conducted between 2018 and 2028 and involving 16 partners, including 12 universities, is used to evaluate the proposed approach. The results show that stakeholders can efficiently detect global behavioral trends and progressively explore specific LPs, supporting informed and trustworthy decision-making. Overall, XAI-Profile contributes to learning analytics by enabling AI-based profiling that is both interpretable and actionable for educational stakeholders.

The paper is structured as follows: Section 2 details the methodology, data collection, and proposed process; Section 3 presents the experimental results; and Section 4 concludes the study and outlines future research directions.

2. High-Level Architecture of Our Framework

As discussed earlier, to support educational stakeholders in validating hypotheses and exploring context-dependent explanations, we introduce XAI-Profile (Explainable Learner Profile), a framework aimed at enhancing transparency and interpretability in LP analysis. Figure 2 provides an overview of the proposed framework. XAI-Profile captures users' goals by decomposing educational requirements into subproblems through goal-oriented AND/OR refinement. This process enables the analysis of hypotheses and the discovery of relevant data features, while systematically linking these hypotheses to learner-related entities. By connecting low-level analytical elements with high-level educational objectives through ML techniques, the framework ensures a transparent understanding of the relationships underlying LP analysis.

[1] https://ecriplus.fr/

This approach supports stakeholders in monitoring different LPs and highlighting the links between input features and predicted outcomes, while enabling interactive exploration of trust-related artifacts including interpretable rule-based models and recurrent data patterns explored through visual workflows within XAI-Profile enables educational stakeholders to examine hypotheses through data-driven, incremental explanations while ensuring that explainable ML models remain aligned with their domain expertise.

To support users in the loop, XAI-Profile is designed around a set of requirements that enable a dialogue-driven approach for LP analysis:

1) **Requirement 1:** Provide formal representations of trust objects such as classification and regression models in machine-readable forms (e.g., decision rules and categories).
2) **Requirement 2:** Provide a guided flow of trust that supports users at each stage of analysis through interaction with the system.
3) **Requirement 3:** Enable stakeholders to communicate their level of confidence in LPs by expressing graded confidence levels (e.g., "highly confident," "moderately confident"), supporting iterative revision.

The following sections explain how these requirements are addressed within our framework by presenting the key components of XAI-Profile and the workflow for LP analysis.

2.1. Transparent learner profiling process

As shown in Figure 3, LP analysis begins with input data and proceeds via a complete, end-to-end process that guides users from data acquisition to the generation of LPs enriched with indicators and explanations. XAI-Profile operates through six main stages: (1) collecting data from the écri+ learning system and

Figure 2
Overview of the high-level architecture of our framework

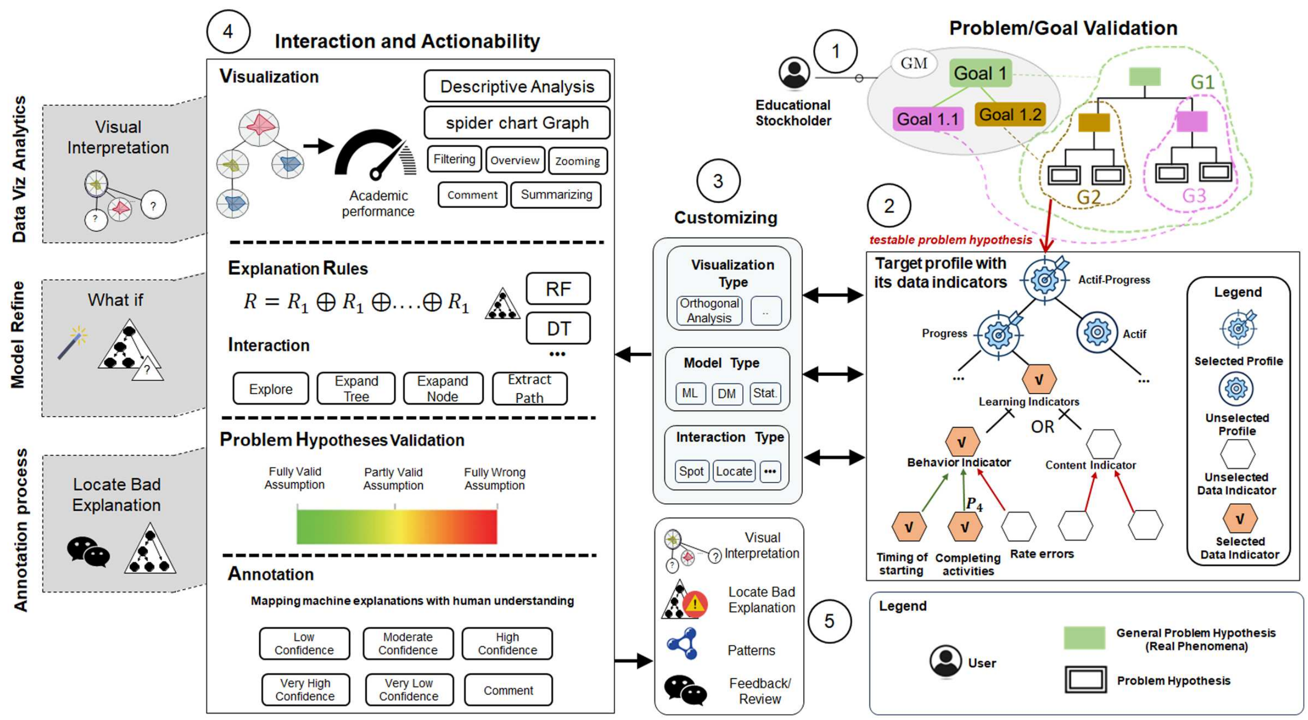
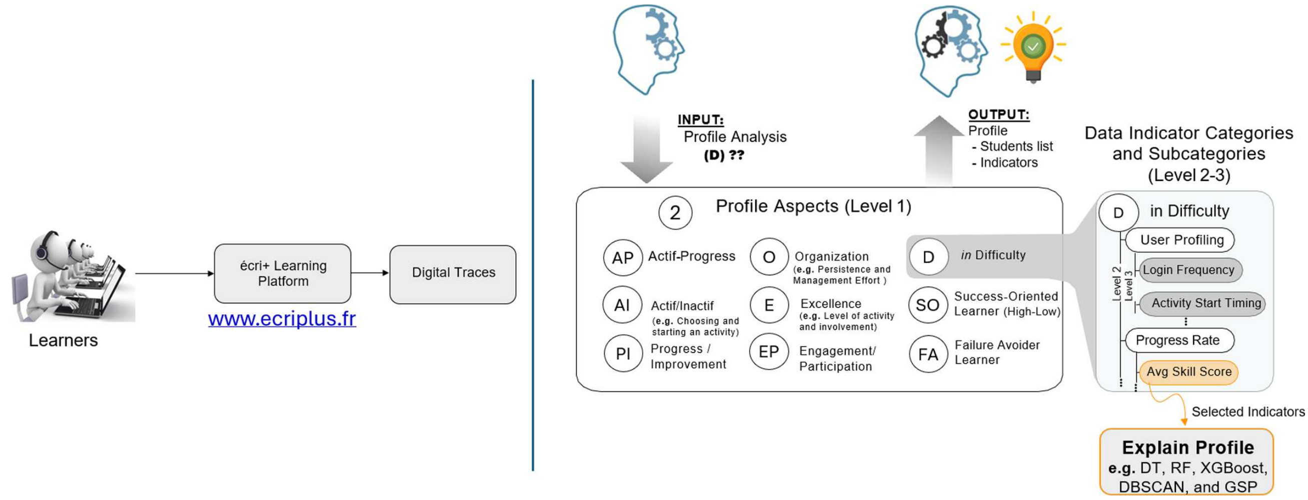


Figure 3
Process of transparent learner profiling

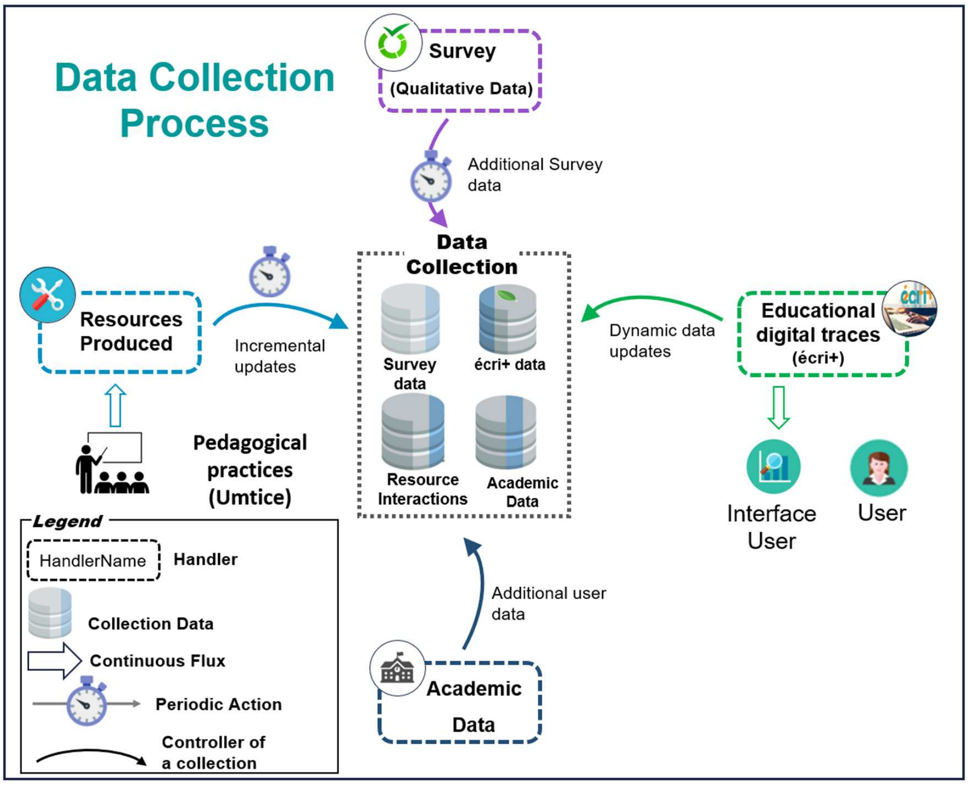


additional internal sources; (2) selecting a target LP belonging to the profile catalog; (3) configuring customized visual analytics; (4) computing profiles through preprocessing, clustering, and learning procedures; (5) visualizing results using interactive representations; and (6) supporting a trust flow that enables stakeholders to validate hypotheses by identifying, locating, and interpreting patterns related to the target profile. The following sections describe each stage in detail.

2.2. Data preparation

As depicted in Figure 4, the data preparation approach links requirements and hypotheses to database attributes, supporting feature selection and the creation of the ML dataset D. To clarify the rationale behind the data, relationships between educational concepts and data features are explicitly examined. The preparation process involves three stages: (1) data collection, (2) feature

Figure 4
Illustration of the data preparation process



extraction based on requirements, and (3) data integration and loading.

1) **Data collection:** We first gather relevant information to analyze LPs, including digital traces from the écri+ platform, qualitative survey responses, academic records, and materials generated through pedagogical activities (Figure 4).
2) **Testable factor identification:** We map the problem hypothesis to the relevant data (Figure 4) to extract a testable factor, classifying attributes as nominal, ordinal, interval, or ratio, with unmatched types labeled "unknown." Data features are then associated with relevant dataset elements, with unmatched ones labeled as "not mapped."
3) **Data integration and preprocessing:** Extracted features were merged into a unified tabular dataset using learner ID and temporal keys (Figure 4). Continuous variables were standardized (z-score normalization) for model stability, while categorical variables were one-hot encoded. To mitigate class imbalance in the target variable (e.g., "Low" vs. "High" performance groups), we applied SMOTE (Synthetic Minority Oversampling Technique) exclusively on the training folds during cross-validation, ensuring no data leakage between training and test sets. The final integrated dataset was stored in a versioned repository, accompanied by a data dictionary detailing each feature's origin, preprocessing step, and pedagogical rationale.

2.2.1. *Choosing a specific learner profile with its associated indicators*

The process begins by choosing a target profile from the profile catalog, which provides an extensible, hierarchical organization of profiles and indicators. This structure allows stakeholders to tailor LPs to specific engagement or performance aspects, customize profiles, and compose complex profiles according to their requirements. For illustration, consider the following user requirement.

"A user aims to identify students on the écri+ platform within a specific competency domain by analyzing relevant knowledge elements (e.g., reading comprehension and text construction), academic background information, and learning resource usage."

To accurately extract data indicators and retrieve measures relevant to a specific LP, we employ domain knowledge as a set of features (*Feature*) and constraints. This model guides the selection of a target LP and its associated data indicators to accurately determine the indicator types (see Figure 5).

Our framework structures indicator families (FI) for LPs. Selection is guided by rules of the form $\text{Rule}_j = \langle \text{ind}_1$, exclude | require, $\text{ind}_2\rangle$, with $Feature = \langle Feature_{and} \mid Feature_{or} \mid Feature_{alt} \mid ind_i\rangle$, where $Feature_{or}$ are mandatory, $Feature_{and}$ optional, and $Feature_{alt}$ alternative indicators. Some profile states are derived from others, reflecting variability in goal achievement and enabling analysis of alternative scenarios via OR/AND decompositions. This approach supports goal merging and prevents conflicts by directing users to a focused set of assumptions aligned with requirements.

2.2.2. *Personalizing visualization goals*

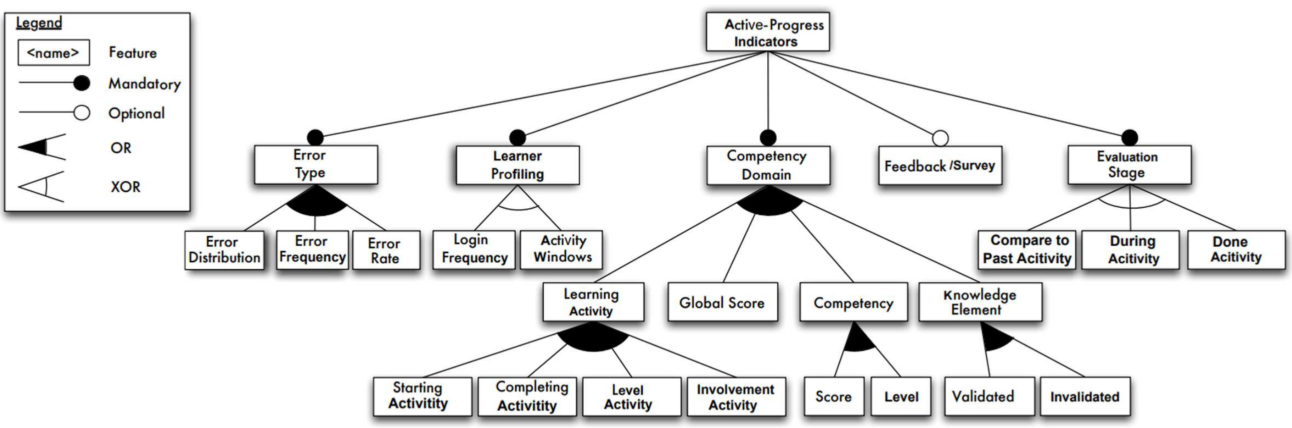
The third step focuses on customizing visualizations, covering the indicator type (vertical or orthogonal) and the modeling approach (statistical or machine learning), visualization goals (V_G), interaction types (I_t), and analysis types (A_t). Users can perform actions like filtering or overview, with each visualization linked to one or more analytical graphs and specific LP goals, such as detecting difficulties or tracking progress. Detailed metrics, for example, login/logout times, can be explored to uncover patterns, providing additional explanations and insights for interpretation.

2.2.3. *Profile calculation*

The primary objective of this step is to compute data indicators derived from events such as responses, assessments, knowledge validation, and competency evaluation. It considers multiple temporal granularities (hourly, daily, weekdays vs. weekends) and usage windows (morning, afternoon, evening, night) captured in the écri+ platform logs. In addition, XAI-Profile exploits fine-grained interaction traces expressed through xAPI verbs (e.g., "completed," "reviewed") to characterize learner actions. Usage time data were collected according to the following structure:

Activity_Record (Record_ID, Event_Type, Action_Type, Event_Timestamp, Event_Details, Tracked_Entity, Current_Phase, View_Window, Additional_Info)

Figure 5
Sample configuration of data indicators for an "active-progress" learner



In our data preprocessing pipeline, we followed two main steps to prepare raw data for analysis. First, we organized and cleaned the data by transforming it into a structured DataFrame and filtering out irrelevant entries. Learners were then grouped using a clustering algorithm based on their scores at different times of day and week to reveal learning patterns. Second, we rebalanced biased datasets to mitigate potential distortions and ensure fairness in subsequent analyses.

2.2.4. Visual analytics

Following the previous stage, users interact with a structured requirement toolkit, denoted as $\mathscr{R}(LP_i) = \langle I_d, V_o, I_m, A_m \rangle$. This toolkit links the requirement associated with a given learner profile LP_i to profile categories (e.g., difficulty, activity level, learning progress) via a set of learning indicators I_d. These indicators are coupled with visual analytics components that define visualization objectives V_o, support multiple interaction modes I_m, and enable different analysis mechanisms A_m. The visual analytics environment exposes relationships between learner behavioral data and graphical representations. Users begin by interpreting visual outputs to identify and explore behavioral patterns, relying on interactive operations such as zooming, filtering, aggregation, and exploration. Overall, the toolkit connects LP categories with trust artifacts (e.g., interpretable decision rules), visualization goals (e.g., outlier detection), and analytical strategies (e.g., clustering). For example, an LP dimension such as persistence (e.g., completion rate $\geq$ 80%) can be analyzed through cross-analysis by combining DBSCAN clustering to identify dense groups of similar behavioral trajectories with sequence mining and timeline visualizations to relate task completion patterns to assessment outcomes.

2.2.5. Flow of trust

To foster understanding and confidence in the model's decisions, these outputs need to be engaged with interactively, allowing stakeholders to explore and act upon them. Through this interactive engagement, stakeholders gain insights into the workings of the ML models. Users can select specific operations from a predefined set of action types, such as browsing, editing, spotting, or locating. This interactive process is structured around the following three key components.

1) Trust objects

Trust objects provide concise and explainable interpretations of learners' behavior represented as Behavioral Information, Main Indicators, and Techniques. Behavioral information includes effort and organization, response time and attention, work regularity, learning strategy and performance, training level, preparation posture, engagement and effectiveness, and resource interaction. Main indicators are activity frequency, timing, response time, focus state, scores, attempts, and resource access. Techniques correspond to Sequence Mining, Decision Tree (DT), Random Forest (RF), and XGBoost. The analysis relies on learning behavior indicators (LBI) (engagement, start/completion timing) and content progress indicators (completed activities, grades). The TrustObjects class captures outputs from white-box XAI models (rules, feature importance, classes, categories), ensuring transparency and reliability in learner profiling.

For example, white-box ML models generate an explicit representation, consisting of a set of rules denoted by r: *Condition* = $R_{1J} \bowtie R_{2J} \bowtie \cdots \bowtie R_{kJ}$, with $k \in$[1, n] and $j \in$ [1, J]. Alternatively, a sequential pattern is a subsequence that appears in multiple sequences with a minimum support threshold. A subsequence P of S is defined as $P = \langle p_1, p_2, \ldots, p_l \rangle$, where $p_i \subseteq s_j$, or as linear regression models derived as $LR(Pi) = p_{i1} \oplus p_{i2} \oplus \cdots \oplus p_{ik}$. These mathematical constructs need to be conveyed in a form that is understandable to educational stakeholders.

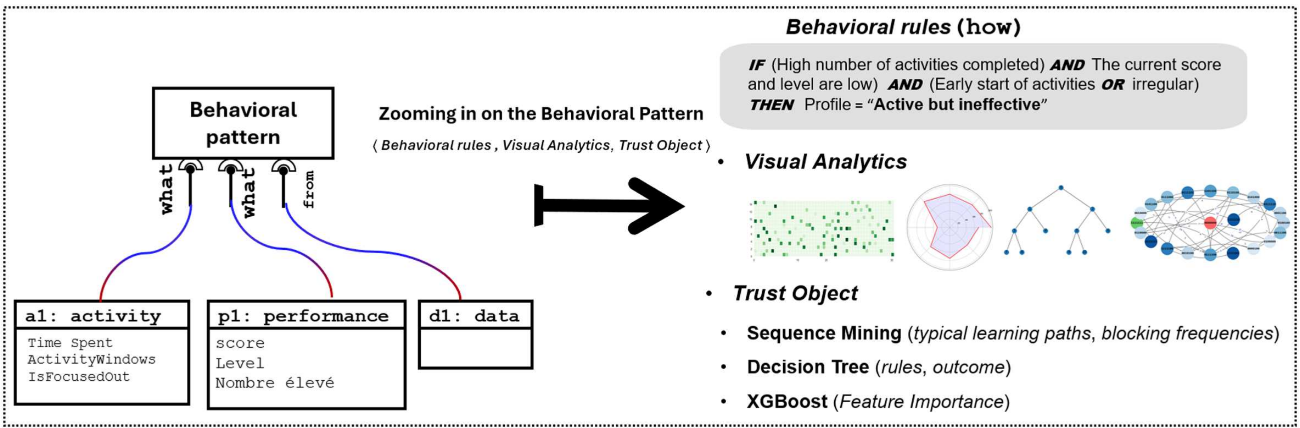
2) Interaction and actionability

Explaining how information relates to stakeholder goals helps identify and analyze patterns corresponding to a specific LP. Interaction and actionability capture the flow of trust, involving operations such as selecting, browsing, locating, filtering, clustering, and customizing, and are connected to the goals of visualization (VG), allowing mapping analytical requirements to data features. Trust objects from white-box XAI models (DT, RF) can be uploaded and refined via the interface. Experts can assign confidence using Probabilistic PCA and adjust decision paths with InteractionType operations like add, remove, expand, or restore.

In XAI-Profile, users can be engaged throughout the process loop (Figure 3). To support this, we introduce a Goal-Oriented Model (GM), which guides the definition of user objectives decomposed using AND/OR operators. This model provides an interactive interface for refining trust objects. By combining ML capabilities with expert insights, the GM fosters a synergy between computation and educational knowledge. As noted in augmented analytics literature, the effectiveness of this approach depends on specific assumptions [31].

Figure 6 shows how LPs are generated from behavioral patterns by combining activity, performance, and interaction

Figure 6
An illustrative example of LP generation from behavioral patterns using trust objects



data. Behavioral patterns are analyzed using rule-based reasoning expressed as IF–THEN rules to interpret learners' actions and outcomes. Visual analytics are used to explore and validate these patterns through graphical representations. A trust object is built from data-driven models such as sequence mining, DTs, and XGBoost to ensure explainability and reliability. Together, these elements enable the generation of meaningful, interpretable, and trustworthy LPs.

3) Learner profile revisions

Finally, we support the revision of explanations to satisfy Req.3. Domain experts can update each LP using predefined action types to express degrees of confidence (e.g., "I'm sure," "I'm fairly confident," "I'm doubtful"). Revisions involve two indicators: (1) profile data indicator relevance, annotated as NOT, RELEVANT, or VERY, where VERY signifies high-quality contributions requiring full review by others; and (2) reviewer confidence, labeled as LOW TRUST or HIGH TRUST, indicating low or high certainty in their assessment. The XAI model is mapped with stakeholder hypotheses via a feature selection analysis matrix, highlighting gaps between algorithmic explanations and expert reasoning. Strongly aligned features are retained, weaker ones are discarded, and in cases of contradiction, users manually set feature selection to ensure structured, unbiased representations.

3. System Evaluation

In this study, conducted at the Faculty of Languages, Le Mans University, France, we systematically examined the requirements of end users and the available data sources to define the LPs under investigation. Existing data were evaluated for relevance and completeness, and supplementary sources were identified as needed. To facilitate stakeholder analysis, LPs were organized into five categories based on difficulty, activity level, and engagement state, in accordance with the primary research questions (RQ1–RQ3) and the objectives of uncovering patterns and validating educational hypotheses (RQ4–RQ5), as outlined below:

1) **RQ 1:** Detecting learners facing learning challenges
2) **RQ 2:** Identify learners who employ learning strategies and manage their effort effectively
3) **RQ 3:** Identify learners demonstrating engagement within the platform
4) **RQ 4:** Identify factors and assumptions that influence learning outcomes
5) **RQ 5:** Explain how and why various factors influence learners' acquisition of specific competencies

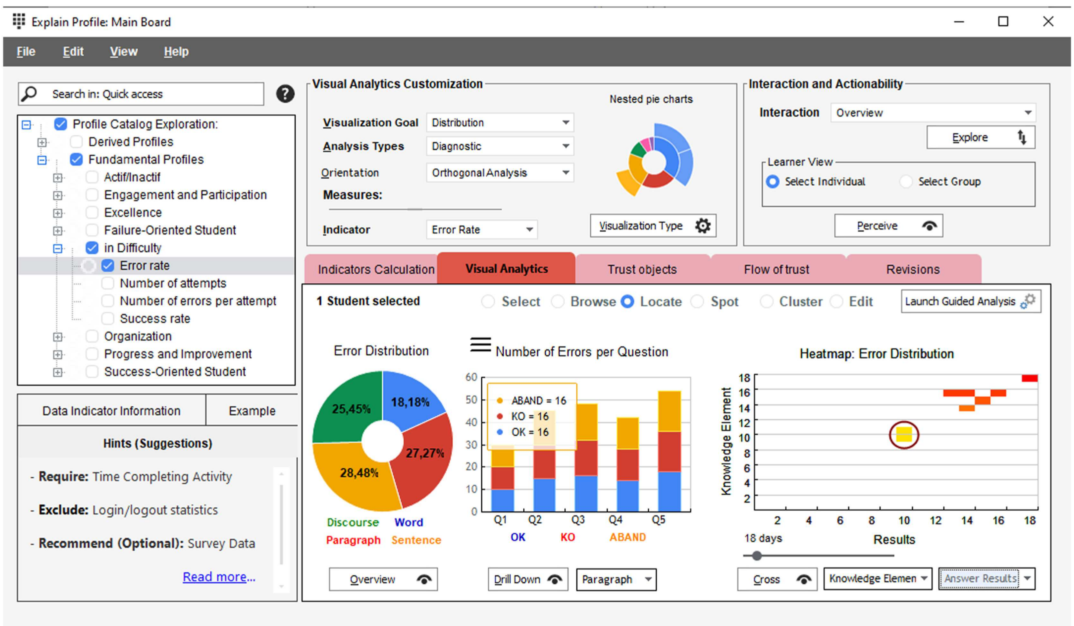
3.1. Supporting tool

The proposal has been instantiated as a prototype tool that converts XAI outputs into interactive dashboards. It translates patterns into visual formats, allowing users to explore trust objects, including rule-based models and recurring data sequences, through interactive workflows. This method links level data with high-level pedagogical insights, in line with the framework design principles. Figure 7 shows the supporting tool within the case study. A video demonstration of XAI-Profile is also available at Youtube.[2]

Users start by selecting a profile and its indicator space from a hierarchical catalog. They then customize analytical graphs, specifying visualization goals, analysis types, and orientation. The system calculates the profile based on the chosen indicators, providing details such as names, educational goals, formulas, and examples. The visual analytics panel provides graphical representations linked to indicators and goals. In the Trust Objects tab, users can explore sequence mining patterns, outcomes, and confidence scores. The Flow of Trust panel shows rules derived from white-box ML models, explaining potential learner difficulties. Within the revisions tab, experts provide annotations to models with thresholds or confidence levels. Through the combination

[2] https://www.youtube.com/watch?v=R6imULg7KoQ

Figure 7
Supporting tool of XAI-Profile



of trust objects and the trust flow, the tool allows interactive analysis of rule-based models and patterns extracted from sequence mining, bridging XAI techniques with user expertise, enhancing understanding, trust, and transparency (see Figure 8).

3.2. Data used

Learner data were gathered from écri+ digital learning traces over the course of one year in 2024, forming an initial sub-dataset. Data were organized by temporal intervals—hours, days, weeks, and months—of platform usage to capture temporal patterns. The dataset includes approximately 227,205 learners and covers Answers, Assessments, Competence Evaluations, Knowledge Elements, Tutorial Evaluations, User-Saved Tutorials, and User-Recommended Trainings. Additional data were integrated, along with responses to an initial questionnaire, to assess learner needs and expectations. The collected data are available for review in a GitHub repository.

3.3. Results and discussion

Experiments were carried out to assess XAI-Profile. Initially, we examined its support for visual analytics in mapping profile categories (RQ1–RQ3), followed by testing its capability to uncover and validate stakeholder requirements (RQ4–RQ5).

3.3.1. Analysis of students' attempts to validate competencies (RQ1)

Our framework presents an overview of each learner's activity within their LP (Figure 9), providing detailed insights into their digital behaviors and the types of actions performed during sessions. Figure 9 illustrates the graphical user interface designed for stakeholders to monitor LPs and digital practices.

Figure 9 presents the analytical method for visualizing the frequency distribution of answer outcomes (OK, KO, Abandon) for each knowledge element. Histograms display scores; some skills, such as M-ORTHO and P-GRAM, had up to 175,000 recorded attempts, reflecting challenges in validating these elements. Other competencies like **D-MCONJ** and **P-CONS** also recorded high numbers of both successful and unsuccessful attempts, while several competencies showed fewer attempts, highlighting variability in learner engagement and performance. Overall, most knowledge elements had moderate and varied validation values. This interface enables comprehensive monitoring of learners' performance relative to écri+ skills and provides stakeholders with clear, timely reports. Additionally, it lists learners with correct responses and efficient response times, facilitating the identification of high-performing learners in large cohorts. Educational stakeholders can also examine data indicators at the individual learner level. Table 2 provides a transparent assessment of a learner's profile using metrics, with well-defined formulas that promote transparency and foster confidence in XAI-Profile.

1) Typical mistake patterns

To track learners' activities during practical sessions, XAI-Profile provides information and course-specific activity reports, helping stakeholders identify common errors across the four écri+ competency domains: word, sentence, text, and discourse. Recognizing frequent mistakes allows stakeholders to pinpoint challenging material, improve explanations, and optimize educational content. The system's visual analytics support LP analysis, as illustrated in Figure 10, showing the state of difficulties across the écri+ domains. The visual analytics output is organized into four sections:

a. **Top-left panel:** Displays the errors of a selected learner to provide insight into individual behavior. Certain challenges show

Figure 8
Supporting tool: illustration of trust objects and flow of trust

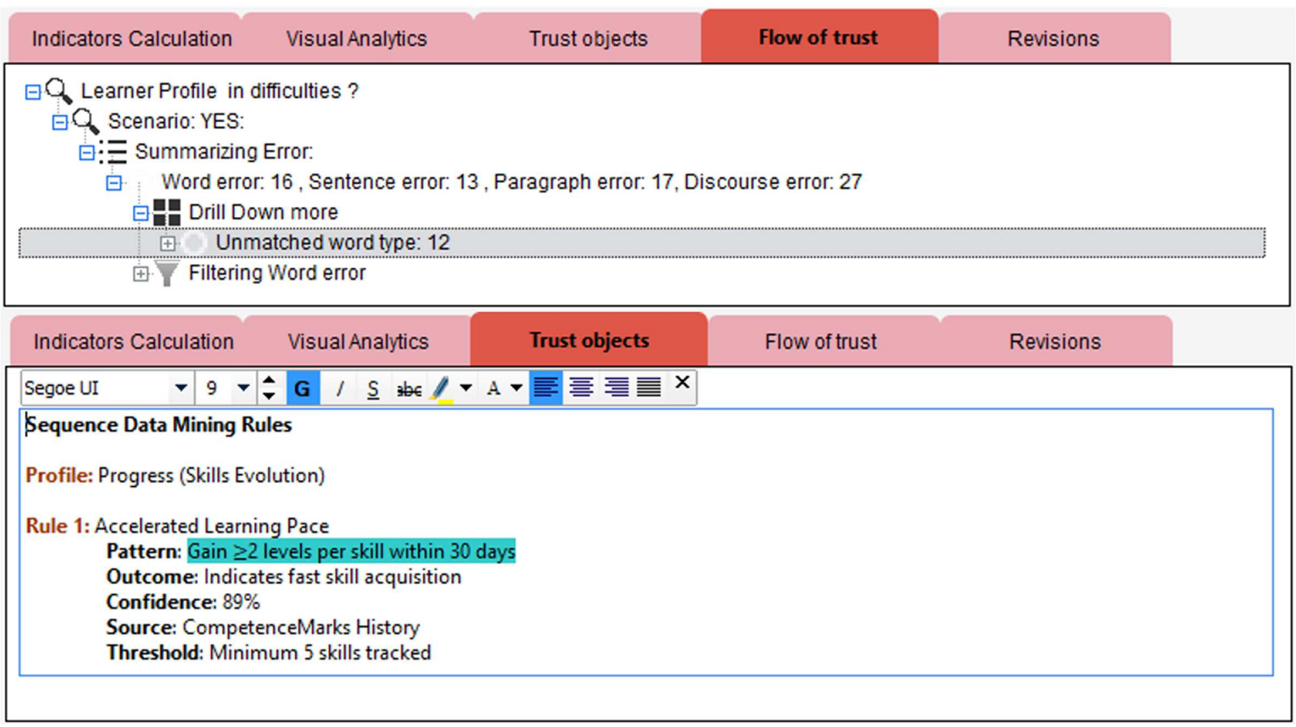


Figure 9
Analysis of learners' competency validation attempts

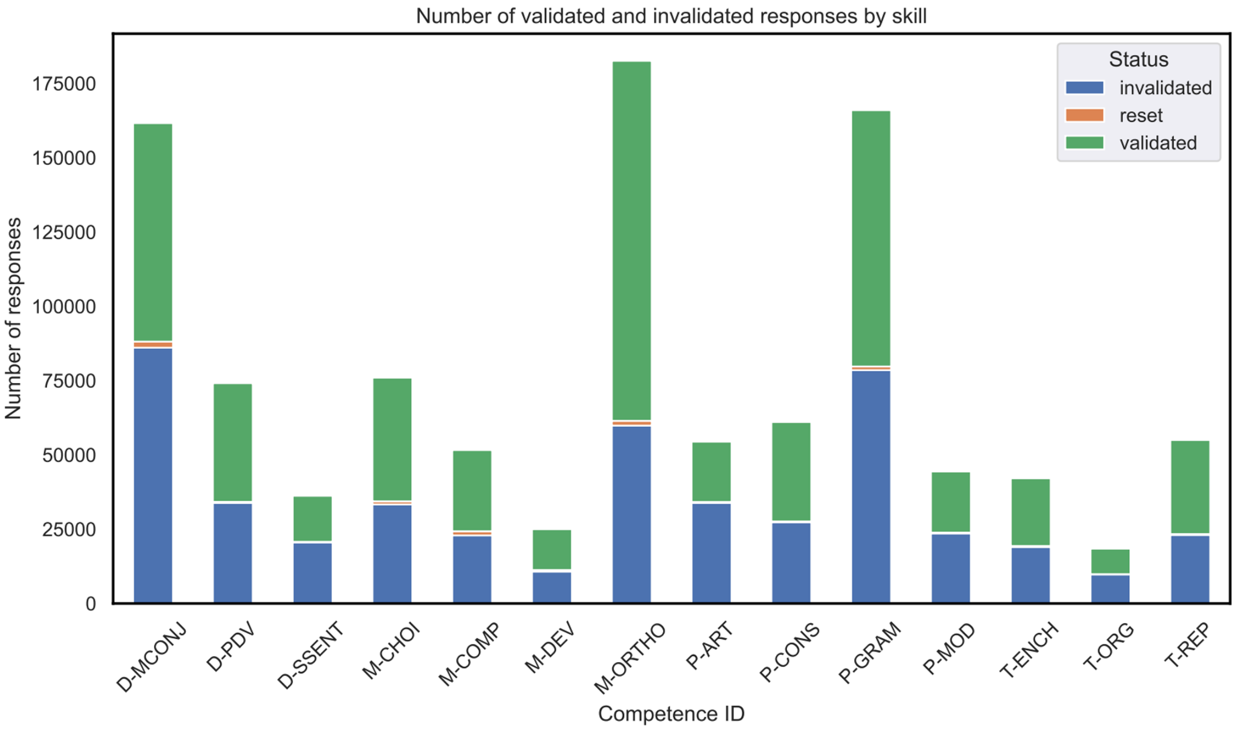


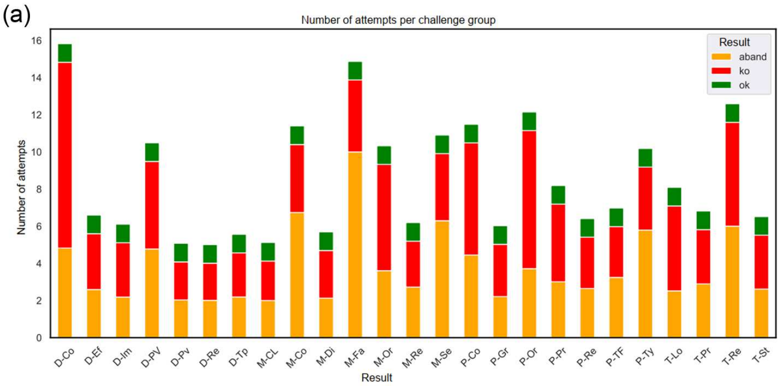
Table 2
Learner-specific indicators linked to a difficulty profile

Profiles	Data indicators
in Difficulty	Error Rate, Number of Errors per Attempt, Error Frequency, Number of Attempts, Error Correction
Active/Inactive	Startup Activities, Time to Complete Activities, Login/Logout Statistics, Students' Active Time, Active Attendance Rate, Login Frequency
Engaged	Starting Activities, Validated Competence, Competence level Average, Error Rate, Completing Activities, Resource Interactions, Performance Score, Competency Types

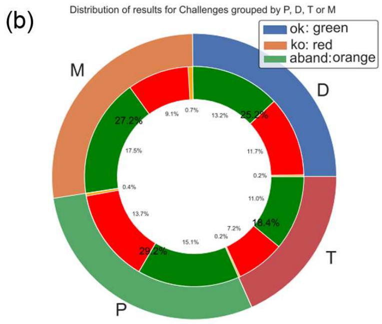
higher failure tendencies, whereas others are attempted successfully with fewer errors. This visualization helps highlight learning difficulties, inform pedagogical support, and characterize learners' weaknesses in interacting with tutorials and resources, offering a reflection of their overall progression.

b. **Top-right panel:** Shows the overall number of solution attempts (Figure 10(b)). Fast Fourier Transform (FFT) analysis is used to identify dominant frequencies in the time series, uncovering underlying patterns and periodic trends that can help predict future performance. Before applying FFT, the distribution of errors across competencies was examined, revealing that some competencies tend to have higher error rates while others consistently show lower error rates. Interestingly, some competencies with higher error rates also exhibit notable correct responses, reflecting variability in learner performance. Abandonment rates were generally low. The FFT of error trends emphasizes periodic fluctuations, offering insights to anticipate learning challenges and guide the adaptation of instructional strategies.
c. **Bottom-left panel:** Displays the total number of solution attempts (Figure 10(b)). FFT analysis identifies dominant frequencies in the time series, revealing monthly patterns useful for forecasting trends. Prior to FFT, error distributions across competencies were examined, showing that some competencies consistently had higher error rates while others performed more reliably. Interestingly, competencies with higher error rates also exhibited a notable proportion of correct responses, indicating variability in learner performance. Abandonment rates were generally low. FFT highlights periodic fluctuations in errors, linking performance variations to assessment timing or workload and supporting better anticipation of difficulties and refinement of learning strategies.
d. **Bottom-left panel:** Shows question traceability using a heatmap to depict the distribution of response outcomes across challenges (Figure 10(c)). Colors represent frequency, from blue (low) to red (high), with the x-axis displaying answer types (OK, KO, Aband). Each data point was mapped to its corresponding cell, incrementing counts for multiple points in the same cell. The heatmap highlights challenge difficulty, identifies patterns in responses, and reveals trends linked to knowledge elements or competency domains. Analysis over time shows usage patterns on écri+, including regular weekday activity, higher engagement on weekends, and lower activity on other days.

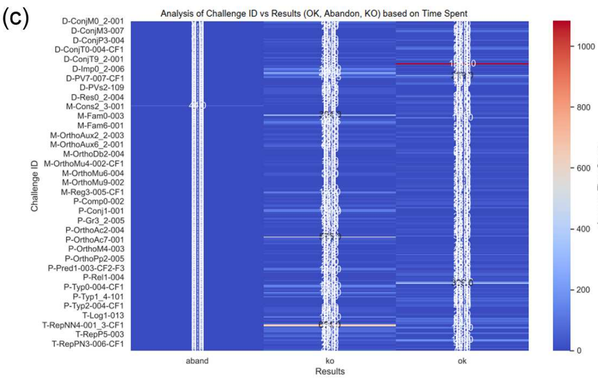
Figure 10
Frequent errors observed in learners' responses



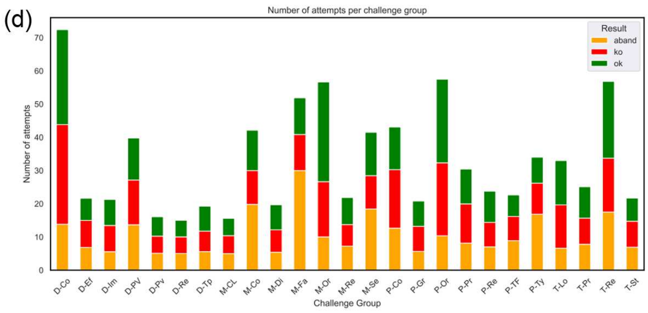
Error count per question for a selected learner.



Error distribution.



Traceability on questions with answer error distribution.



Total number of errors among questions

e. **Bottom-right panel:** Displays error counts per question through a bar chart (Figure 10(d)), illustrating the distribution of errors across categories such as word, sentence, text, and discourse from learner digital learning traces. The chart distinguishes between abandoned, incorrect, and correct attempts. Some challenges show consistently high error and abandonment rates, reflecting greater difficulty, while others have lower rates, suggesting easier tasks or lower learner engagement. These insights can guide targeted instructional improvements.

Based on the heatmap results, we examined how knowledge elements intersect with response details (Figure 11, top) and validation state (validated or not; Figure 11, bottom) for a sample of 20 students to assess whether these components are independent. Through an analysis of the distribution of competency across validation states and two conceptually orthogonal facets (Figure 11, left), we evaluated whether learners actively validate and understand the elements. Some known patterns were confirmed: crossing knowledge elements with answer details revealed that roughly 45% of items correspond to validated knowledge. For instance, D-PVs5-007 in the Discours category shows that most related knowledge elements are validated only after multiple attempts. Additionally, distribution over answer details (Figure 11, right) indicates that most knowledge elements remain validated and are frequently retried even after errors. Notably, no Discours-based knowledge element appears after validation following many attempts.

In summary, the experimental findings show that LP difficulties can be identified by examining ordered lists of typical errors to highlight recurring mistakes and biases. Shifts in LP, influenced by factors such as competence, progress, or completion, help users investigate potential causes of learner difficulties and may expose latent biases through cross-analysis of multiple visual analytics.

3.3.2. Analyzing LP active/inactive (RQ2)

1) Temporal patterns of activity start

Another crucial dimension in analyzing learner behavior relates to the moment at which the activity begins, which can highlight engagement patterns and potential biases. Activity frequency and duration act as proxies for learner effort, supporting transparent progress tracking. Repeated practice generally improves answer accuracy and leads to higher success rates in final competency validation attempts, reducing bias impact. Our system provides data on activity start times to evaluate learner engagement during sessions (Figure 12), helping to prevent misinterpretations.

XAI-Profile adopts a concept similar to the GitHub Contributions Graph to display learners' sequences of actions,

Figure 11
Visualization of knowledge elements by validation status (upper) and response breakdown (lower)

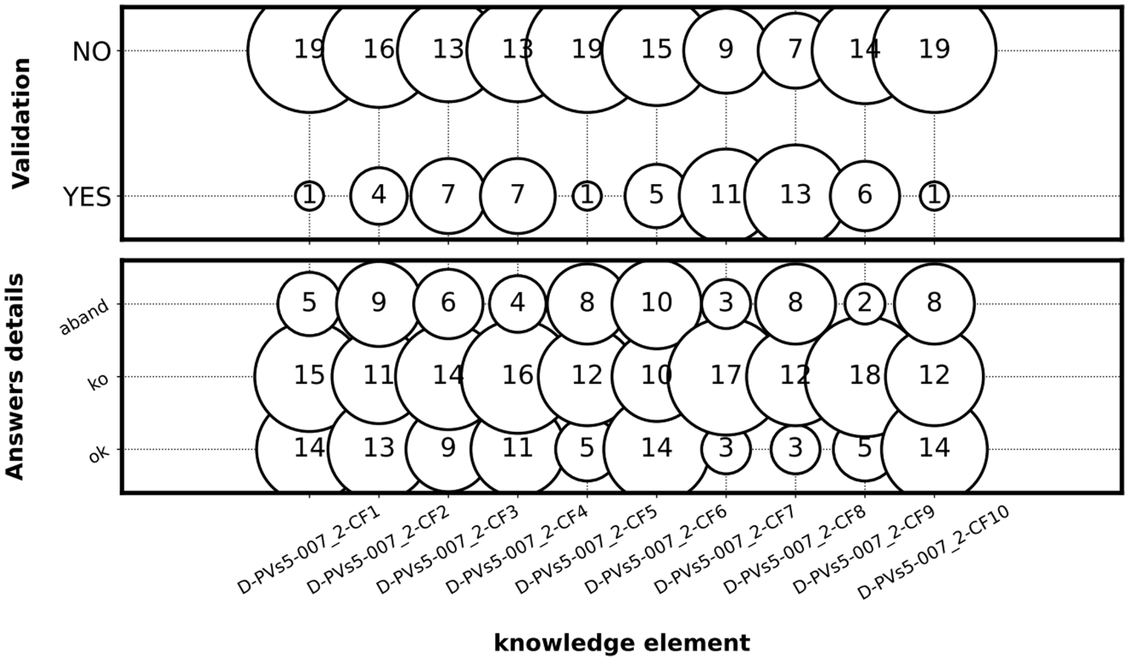
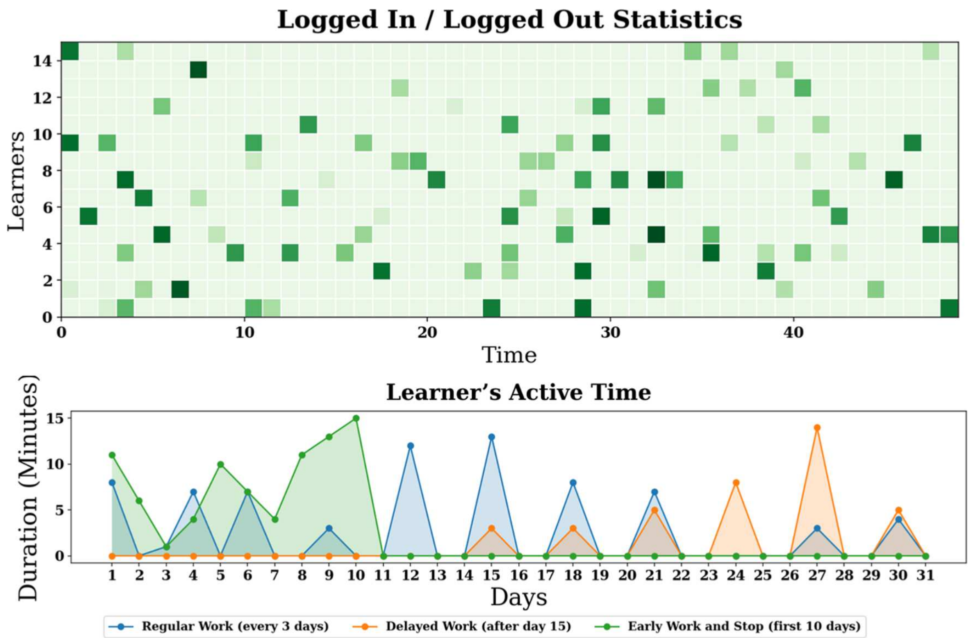


Figure 12
Activity-window–based learner profiling



improving explainability. This is reflected in the logged-in/logged-out activity statistics for each learner (Figure 12). Stakeholders can evaluate learner engagement on the écri+ platform, distinguishing active from less active users while maintaining fairness. Green squares represent activity levels, with darker shades indicating higher engagement. A balanced pattern signifies consistent activity, helping identify potential biases or participation disparities.

2) Time investment in learning tasks

This visualization helps stakeholders identify learners with effective organizational and learning strategies, though potential biases may exist. By observing sequences of learner actions during practice, stakeholders can gauge engagement and identify active versus inactive learners, supporting fair evaluation. However, "learner gaming" can distort this interpretation; for instance, repeated attempts at questions may generate high log activity without genuine learning. Conversely, inactivity or long pauses may be misinterpreted as disengagement. The écri+ platform tracks focus with Boolean variables such as IsFocusedOut and HasSeenEndTestScreenVerificationCode (true/false), mitigating exploitation of the system for high scores without real engagement. Orthogonal analysis of activity start times can further reveal correlations with performance indicators (e.g., activity completion, error rates). Stakeholders can combine visualizations to associate behavioral patterns with digital learning activities, offering insights into learners' engagement levels and overall learning states.

3.3.3. Analyzing learner profiles aligned with engagement (RQ3)

1) Success/failure-oriented learner

To gain insights from stakeholders, we present an engagement state that simplifies complex LP information into a clear and interpretable form. This abstraction converts detailed data into visually intuitive representations, making it easier to understand multiple facets of learner engagement. Engagement LPs are derived using a synthesis of task-level and process-level indicators. These are visualized as spider charts across eight dimensions: activity initiation, validated knowledge components, average competency level, frequency of errors, task completion, interactions with learning resources, performance scoring, and competency areas. These charts allow educational stakeholders to interpret learner engagement levels effectively. We also perform a comparative analysis between the engagement states generated by XAI-Profile, including metrics such as course activity completion and graded assignments, to evaluate alignment. The visual analytics of XAI-Profile demonstrate that engagement states are easily comprehensible. Figure 13 illustrates this with two examples: Figure 13(a) shows an unengaged student, while Figure 13(b) depicts an engaged learner, highlighting the relationships between LP indicators.

Figure 13 illustrates how the two LBIs—activity start timing and activity completion timing—reflect delays in initiation or punctual completion, which adversely affect the assessment of responses and, consequently, the quality and correctness of the proposed solutions.

2) Temporal dynamics of learner engagement

Figure 14 illustrates a clear change in LP engagement states, captured across eight indicators in the spider chart. Figure 14(a) shows the transition from an unengaged to an engaged LP, while Figure 14(b) depicts the reverse shift from engaged to unengaged. Color coding is used to represent these transitions: blue indicates LP behavior before the change, and red represents behavior after the shift. The visualization demonstrates that engagement status is easily interpretable for stakeholders, highlighting that changes in LP indicators effectively reflect shifts in learner engagement.

Figure 13
Representative learner engagement case

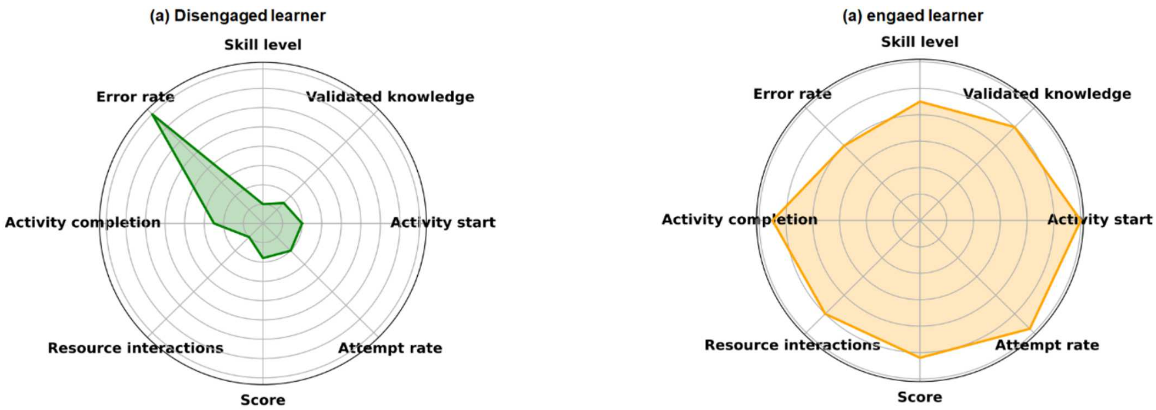
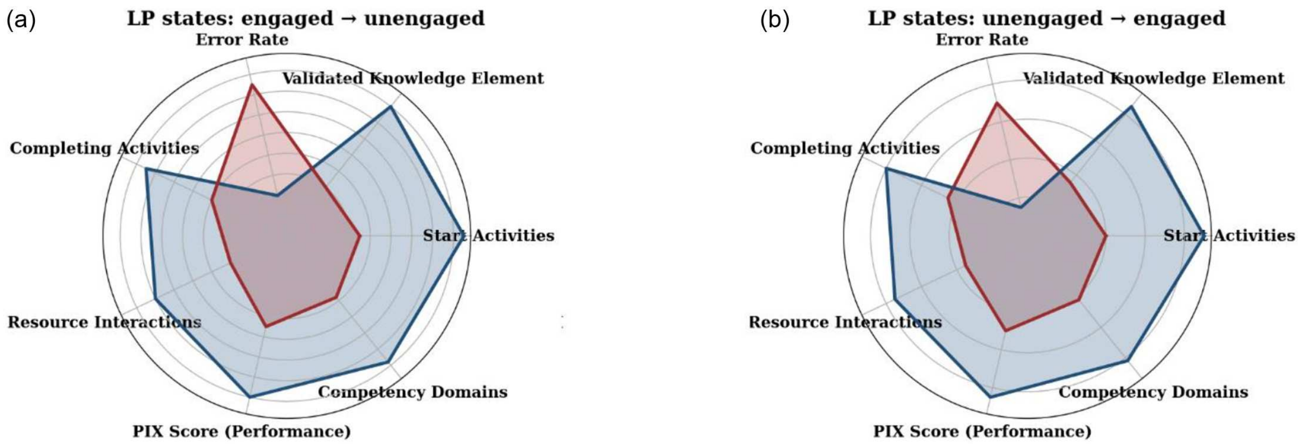


Figure 14
Temporal dynamics of learner engagement



An examination of multiple cases, each reflecting a distinct learner state inferred from data analytics and validated by educational stakeholders, reveals a consistent trend. In most observations, the engagement states identified by our framework align with expert intentions and values. Overall, a strong relationship is observed between students' self-perceived successes or difficulties and their observable engagement behaviors. This alignment allows stakeholders to intervene more accurately in engagement-related issues. Moreover, continuous engagement monitoring through XAI-Profile supports a feedback mechanism that can enhance problem-solving, as increased awareness of observation may positively influence learner behavior through the Hawthorne effect.

3.3.4. Uncovering learning patterns and evaluating stakeholders' assumptions (RQ4–RQ5)

In this section, we examine notable and surprising patterns to test our assumptions and provide explanations tailored to specific contexts. This approach ensures that AI-generated insights are both reliable and practical, supporting a human-centered perspective.

1) Discovery of temporal patterns

Guided by research questions RQ4 and RQ5, which seek to explain how learners achieve or fail to reach competency on the écri+ platform, we apply sequence mining techniques to identify recurring sources of difficulty, with a focus on the most frequent patterns observed in the data. Learner competency progressions are represented as ordered sequences that capture skill mastery at successive stages, enabling the analysis of transitions between these states. The following sections explore these transitions in detail, emphasizing both expected and atypical pathways within learners' trajectories. As learners progress through their learning paths, preprocessing is formalized as follows.

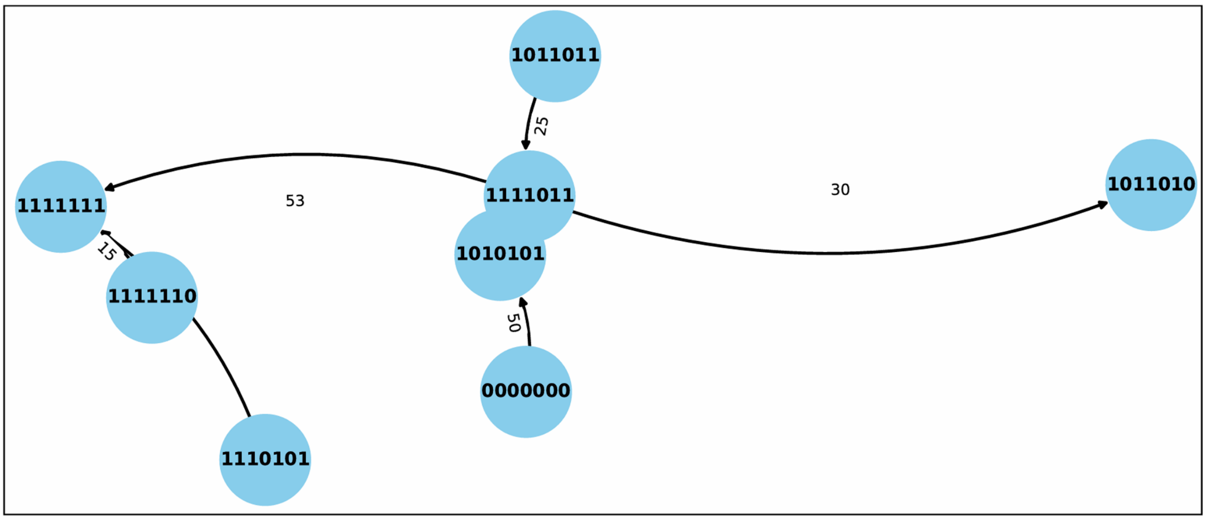
a. **Raw data:** contains TimeID, User ID, Skill ID, and Status
b. **Processed data frame:** contains TimeID, User ID, and binary representations of the Status for each Skill ID

In this study, a *sequence* is represented as an ordered binary vector, where each position corresponds to a specific skill within a competency domain. A value of 1 indicates that the skill has been validated or mastered by the learner, while a value of 0 denotes that the skill has not yet been acquired. Clustering was initially performed using the DBSCAN algorithm, considering the four competency domains of the écri+ platform—word, sentence, text, and discourse—together with the Performance Score. For illustration, the competency progression associated with the word domain follows an ordered set of skills, including spelling, conjugation, sentence construction, grammatical rules, vocabulary familiarity, discourse organization, and lexical coherence.

To analyze learning dynamics, we applied sequence mining using the Generalized Sequential Patterns (GSP) algorithm, which enables the identification of frequent learning paths while preserving human interpretability. Each learner follows a competency trajectory expressed as an ordered series of binary mastery states, starting from a state where no skills are validated and progressing toward full competency. Transitions between states capture incremental learning steps, reflecting the gradual acquisition of skills from simpler to more complex levels. A detailed examination of learner interaction data from the écri+ platform through sequence mining techniques enabled the monitoring of learners' knowledge states and the identification of recurrent difficulty points during competency progression. The analysis revealed both typical learning paths and atypical behaviors, including commonly observed transitions and irregular trajectories. Figure 15 presents the resulting Competency Transition Graph, highlighting frequent mastery patterns among learners who reached advanced competency levels while still lacking a limited number of skills.

At this stage, a subset of learners progressed to a state representing full skill mastery, while others diverged toward a configuration reflecting partial acquisition. The transition toward complete mastery was generally easy to interpret. However, the occurrence of reverse transitions, where learners moved from a more advanced state back to a less advanced one, was unexpected and suggests that some students modified earlier responses, leading to the failure of previously validated test cases. Most learners reaching advanced intermediate states continued progressing, whereas a smaller proportion experienced stagnation or regression. In some cases, learners failed only the final validation step, indicating that the last skill was more demanding than earlier ones. This finding is consistent with the heatmap analysis,

Figure 15
Patterns and deviations in learners' competency evolution on the écri+ platform



which showed generally low difficulty levels across skills, alongside frequent alternations between success and failure. Such alternating mastery patterns were common in learner responses. A further insight concerns competency trajectories: the structure of final transitions suggests that success or failure in a given competency tends to be largely independent of performance in other competency domains.

2) Examining stakeholder-driven educational assumptions

To validate educational stakeholders' hypotheses and provide complementary explanations, it is essential to align ML explainability models with human expertise. The écri+ project steering committee conducted two surveys: an initial survey aimed at identifying learners' needs, followed by a second survey designed to assess the extent to which these needs were addressed. Figure 16 presents the most influential features identified by the XGBoost model based on analyses of both the original and preprocessed datasets, highlighting their positive or negative impact on learners' scores within the ÉCRI+ platform. The feature Resources Consulted emerged as the most influential factor, followed by Language (Native/Schooling) with moderate importance. Other variables, such as Field of Study, Age, and Type of Baccalaureate, exhibited a limited impact.

Drawing on the survey outcomes, expert users are tasked with contrasting these results with the outputs of the ML explainability model to investigate possible explanations for the proposed hypotheses:

1) $\mathscr{H}_1$**:** Low learner scores could be influenced by whether French is the learner's native language or a second language.

2) $\mathscr{H}_2$**:** Low learner scores may be associated with underutilization of available resources.

Validating user hypotheses requires clarifying the explicit links between specific fine-grained issues and overarching educational objectives. To bridge analytical issues and data, stakeholders' needs are first articulated as high-level goals and subsequently refined through GM decomposition, enabling the exploration of elementary hypothesis formulations and the identification of relevant data attributes for analysis. Alternatively, tool-assisted semi-automated or fully manual mapping approaches may be applied. Data preparation relies on questionnaire-based datasets, which may require feature-specific preprocessing, as well as the construction of an ML dataset that supports hypothesis validation.

To test a problem hypothesis, we derived explicit causal relationships expressed through logical operators (e.g., AND, OR) from DT and RF models, ensuring traceability across decision rules explaining performance scores in three categories: Low, Medium, and High. The DT algorithm produces rules corresponding to these categories, with scores classified as High (≥85%), Medium (50–84%), and Low (<50%). The generated rules, presented in Table 3, allow users to understand the factors behind success or failure, fostering trust in the decisions. This approach aims to complement problem hypotheses with ML models, while acknowledging their risk of biased and incorrect results.

The overall performance of the ML models is shown in Figure 17. As observed previously, the RF model demonstrates the strongest overall accuracy, achieving a precision of 70.40%, a recall of 0.743, and an F-measure of 0.725. In comparison, the DT model attains a higher precision of 81.13% but lower recall (0.708) and F-measure (0.694), resulting in weaker overall performance.

The steering committee of the écri+ project aims to validate two primary goals: G1, which concerns unmet learner requirements and is linked to hypotheses $\mathscr{H}_1$ and $\mathscr{H}_2$, and G2, which focuses on enhancing learner success and is associated with hypotheses $\mathscr{H}_3$ and $\mathscr{H}_4$ (see Figure 18). A sample of 336 survey responses was analyzed from the CSV dataset to evaluate whether the rules that were generated by the DT and RF algorithms align with the educational stakeholders' hypotheses H1 and $\mathscr{H}_2$. Only rules containing data features relevant to the problem hypotheses were considered. The results for $\mathscr{H}_1$ and $\mathscr{H}_2$, as identified by the proposed model, to demonstrate the level of explainability achieved. The results of testing the hypotheses $\mathscr{H}_1$–$\mathscr{H}_4$ reveal distinct patterns in how different factors influence student performance. Hypothesis $\mathscr{H}_2$, which associates low performance with unreviewed learning resources, emerges

Figure 16
Important features in the surveys for the original and preprocessed datasets

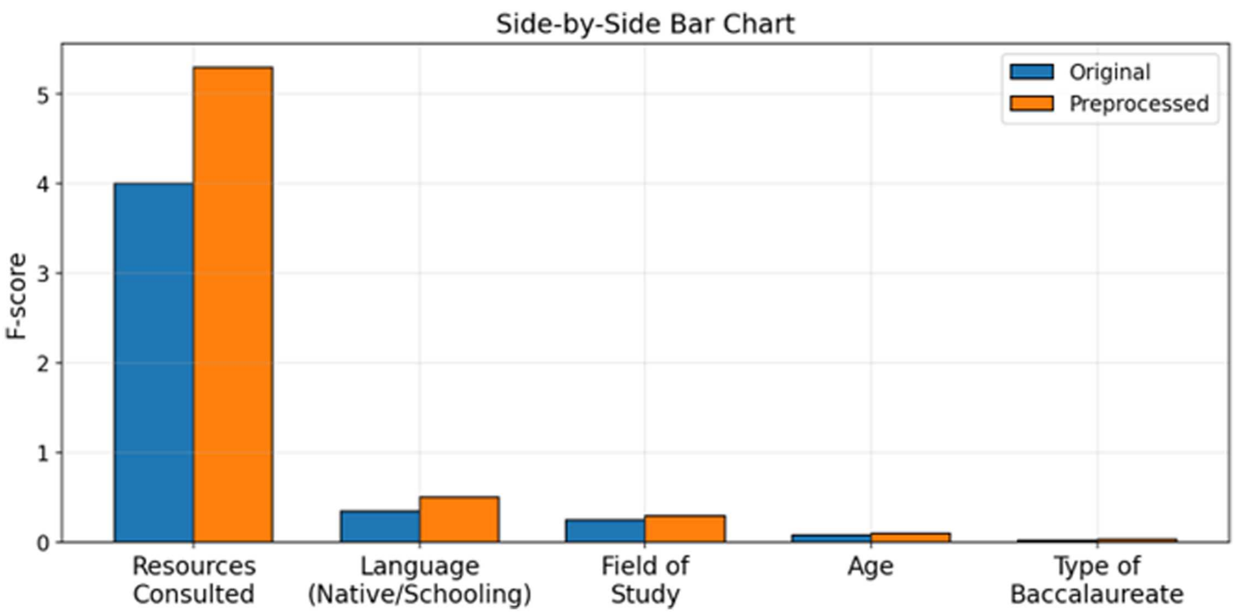
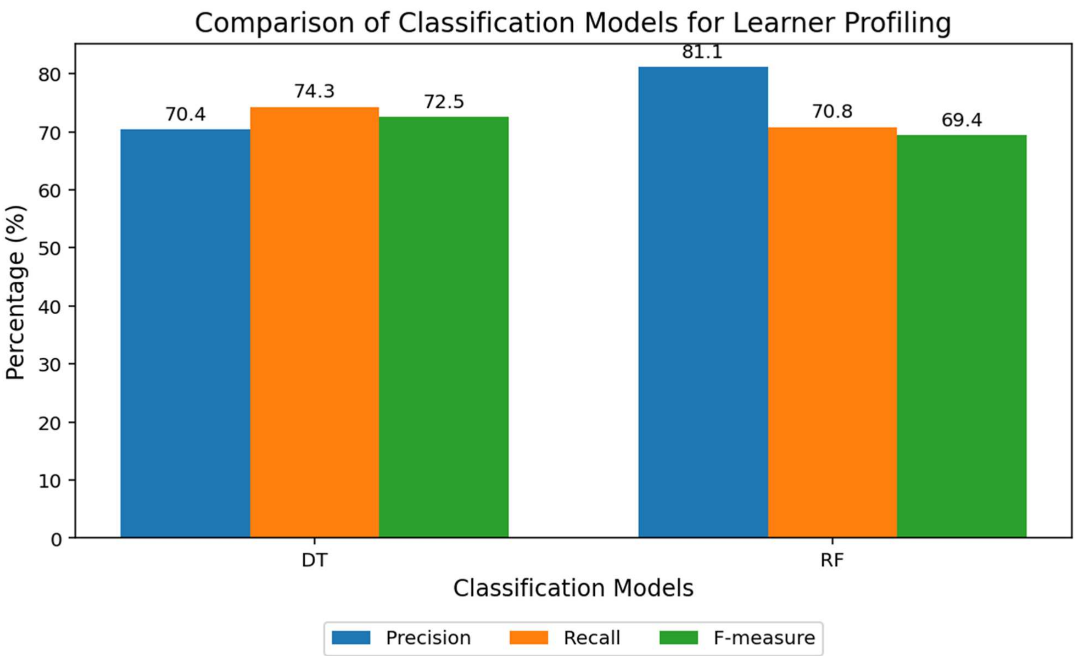


Table 3
Extracted rules based on the collected data

Line	Model	Rule	Score	Outcome (%)
1	DT	**If** All KE of Word = ”Yes” **and** All KE of Phrase = ”Yes” **and** LogFreq > 20 **and** ImpObs = ”Yes”	High	85
2	DT	**If** Some KE of Text = ”Yes” **and** Not all KE of Discourse = ”Yes” and SavedTut = ”Yes” **and** NbrSaveTut > 5	Medium	78
3	DT	**If** Not all KE of Word = ”Yes” **and** UnreviewedRes = ”Yes” and AbandExist = ”Yes”	Low	80
4	DT	**If** IsNativeFrench = ”Yes” **and** RecentChallenge = ”Discourse” **and** All KE of Discourse = ”No”	Medium	75
5	RF	**If** (All KE of Text = ”Yes” **OR** All KE of Discourse = ”Yes”) **and** Login frequency > 25 **and** NumSaveTut = ”Yes”	High	85
6	RF	**If** Some KE of Word = ”Yes” **and** Some KE of Phrase = ”Yes” **and** NbrSaveTut > 5 **and** UnreviewedRes = ”Yes”	Low	88
7	RF	**If** Last Challenge ID = ”Discourse” **and** All KE of Discourse = ”No” **and** If Abandoned < 10 = ”Yes”	Low	82
8	RF	**If** NbrSaveTut > 5 **and** UnreviewedRes = ”No” **and** ImpObs = ”Yes”	High	87

Note: KE: Knowledge Element, LogFreq: Login Frequency, ImpObs: Improvement Observed, SavedTut: Tutorial Saved, UnreviewedRes: unchecked learning resource, AbandExist: Abandonment Exists, IsNativeFrench: French as Native Language indicator.

Figure 17
Most influential survey features before and after data preprocessing



as the most significant, indicating that active engagement with materials is critical for success. Hypotheses $\mathscr{H}_1$ and $\mathscr{H}_4$, concerning the role of French (native or second language), show lower validation rates, suggesting that language proficiency has a more limited but noticeable effect on outcomes. Hypothesis $\mathscr{H}_3$, addressing issues with resource accessibility or use, demonstrates moderate support, highlighting that improving resource guidance can help bridge performance gaps. Additional insights from the analysis show that learners who frequently revisit tutorials and follow structured study strategies tend to achieve higher performance scores, reinforcing the importance of monitoring both resource usage and learning behaviors. Moreover, results from DT and RF indicate that RF captures more complex interactions between multiple factors, providing slightly higher validation rates, while DT offers clearer, interpretable rules for individual cases.

Assumption validation is represented using a color scale from green (fully valid) to red (completely invalid). For instance, LP3 marked in red indicates irrelevance to the goals and is evaluated with exclude (-) values. The SuccessRate for each goal is computed based on the contributions of the indicators from the learning paths (LP).

Figure 18 summarizes the validation results for the two hypotheses proposed by educational stakeholders. Both the DT and RF models offer interpretative insights into the factors affecting learner performance. Figure 19 provides a complementary

Figure 18
Hypothesis validation with machine explanations

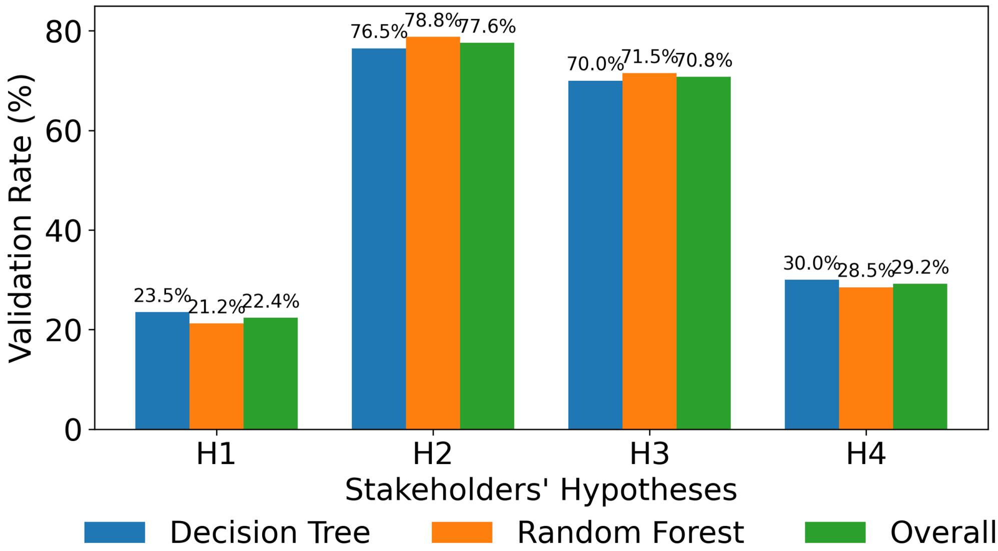
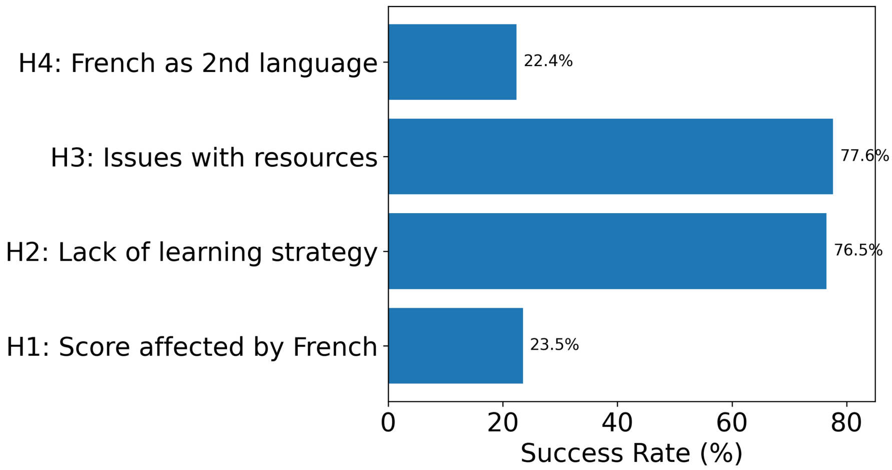


Figure 19
Hypothesis validation with machine explanations

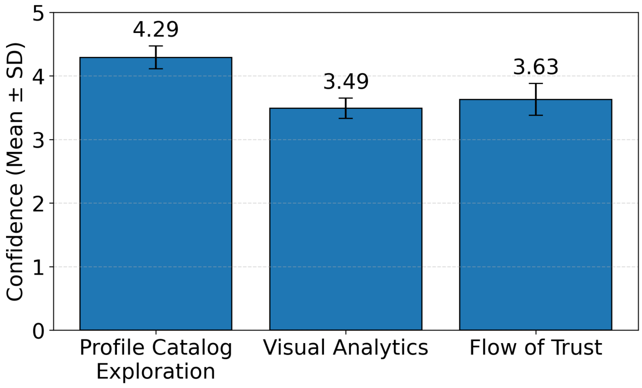


visualization of the overall validation results, highlighting the success rates associated with the stakeholder hypotheses. The first hypothesis (H_1), which examines the influence of French as a first or second language on learners' scores, receives limited empirical support. It is validated in 23.5% of cases by the DT model and 21.2% by the RF model, resulting in an average validation rate of 22.4%. These findings indicate that language background contributes to performance variability but does not represent a primary explanatory variable. In contrast, the second hypothesis (H_2), which relates lower performance to limited engagement with learning resources, is strongly supported by both models. The DT model assigns 76.5% of performance variance to this factor, while the RF model reports a higher contribution of 78.8%, leading to a combined validation rate of 77.6%. This strong convergence underscores the pivotal role of resource usage in explaining learner outcomes and highlights the importance of fostering active interaction with available learning materials. Although the effect of language proficiency is comparatively weaker, it remains non-negligible, suggesting that targeted support for learners for whom French is a second language may still be beneficial.

3.4. Integrating educators into learner profile analysis for enhanced interpretability

The educator-in-the-loop approach integrates expert feedback into LP analysis, emphasizing key components such as profile exploration, visual analytics, and trust flow (see Figure 20). Educators actively interact with the system to interpret learner data, refine decisions, and ensure that AI outputs align with pedagogical goals. Sessions included demonstrations, guided exercises, and example analyses covering learners in difficulty, varying engagement levels, and different activity states. Participants engaged with both paper-based and digital tools to examine indicators, behavioral traces, and ML outputs. Feedback

Figure 20
Educator confidence in learner profile analysis



was collected through online forms and think-aloud workshops, capturing quantitative assessments and rich qualitative insights. Results indicate that the majority of educators felt confident in selecting indicators and profiles, moderately confident in using visual analytics, and less confident in refining ML outputs. Visualizations such as DTs, heatmaps, and radar charts were particularly valuable for understanding profiles and detecting biases. Educators highlighted the need for more detailed linguistic profiling and transparency in feature weighting. The approach encourages interactive reasoning, bridging human expertise with AI outputs to improve interpretability and trust. Overall, combining structured guidance, visual analytics, and reflective engagement enhances actionable understanding of LPs and supports informed educational decision-making.

3.5. Implications of our framework for pedagogical practice

Our framework provides actionable insights for teaching and learning by integrating explainable AI with learner profiling. It allows educators to transparently identify learner states including those struggling, inactive, or exhibiting motivational profiles such as Failure Avoider or Overstriver and act on them through concrete interventions. Visual analytics and profile exploration help interpret complex behavioral and engagement data, supporting informed pedagogical decisions. Clustering and adaptive modeling enable personalized interventions aligned with learner diversity, while validation against educators' domain knowledge enhances trust and interpretability. Collaborative tools allow multiple stakeholders to refine profiles from diverse perspectives. Integration of behavioral indicators, learning outcomes, and sociodemographic data supports evidence-based, targeted guidance and self-regulated learning. Transparency in model reasoning and feature-to-decision explanations mitigates bias, promotes fairness, and informs instructional design. Overall, the framework operationalizes AI insights into concrete pedagogical actions, bridging learner data and effective, human-centered educational practices.

4. Conclusion

This study presents a transparent framework for learner profiling that uses explainable AI and interactive visual workflows to link raw data to pedagogical insights. It provides clear, actionable explanations of LPs while engaging users to validate patterns and hypotheses.

Experimental evaluation in the écri+ project validates the framework's capacity to generate interpretable LPs, empower hypothesis testing, and foster educator trust through interactive visual analytics. Results show that explainable AI outputs, when integrated with guided workflows and educator-in-the-loop feedback, successfully translate low-level data into transparent pedagogical insight, bridging computational analysis and educational decision-making. Future work will evaluate the quality of explanations using additional criteria, such as stability, considering multiple interpretations of LPs due to notation variability and differing pedagogical practices. We will address cognitive biases in interpreting cause-effect relationships within visualizations, employing methods like scoring functions and eye-tracking. Experiments will assess explanations from models such as DTs, RFs, XGBoost, DBSCAN, and GSP. Additionally, we aim to improve clarity and usability through user-centered design principles tailored to the educational context.

Acknowledgement

Our work is carried out within the French project PIA écri+ (no ANR-17-NCUN-0015, http://ecriplus.fr/).

Funding Support

This work is sponsored by ANR3 (National Research Agency) as part of a national project spanning 10 years (2018–2028), involving 16 partners, including 12 universities.

Ethical Statement

The authors declare that this study does not involve any experiments with human participants or animals. The research is limited to the design and analysis of computational models and does not require ethical approval or informed consent.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this work.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at https://github.com/ouared14/XAI-Profile.

Author Contribution Statement

Abdelkader Ouared: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Madeth May:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration. **Claudine Piau-Toffolon:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration. **Nicolas Dugué:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Visualization, Supervision.

References

[1] Duan, X., Wang, C., & Rouamba, G. (2022). Designing a learning analytics dashboard to provide students with actionable feedback and evaluating its impacts. In *Proceedings of the 14th International Conference on Computer Supported Education*, *2*, 117–127. https://doi.org/10.5220/0011116400003182

[2] Chen, J., Mokmin, N. A. M., Shen, Q., & Su, H. (2025). Leveraging AI in design education: Exploring virtual instructors and conversational techniques in flipped classroom models. *Education and Information Technologies*, *30*(12), 16441–16461. https://doi.org/10.1007/s10639-025-13458-z

[3] Christopoulos, A., & Pellas, N. (2020). Theoretical foundations of virtual and augmented reality-supported learning analytics. In *2020 11th International Conference on Information, Intelligence, Systems and Applications*, 1–4. https://doi.org/10.1109/IISA50023.2020.9284410

[4] Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, *32*(4), 1993–2020. https://doi.org/10.1007/s12525-022-00605-4

[5] Wong, B. T., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, *7*(1), 7–28. https://doi.org/10.1007/s40692-019-00143-7

[6] Park, Y., & Jo, I.-H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science*, *21*(1), 110–133.

[7] Park, Y., & Jo, I.-H. (2019). Factors that affect the success of learning analytics dashboards. *Educational Technology Research and Development*, *67*(6), 1547–1571. https://doi.org/10.1007/s11423-019-09693-0

[8] Dickler, R. (2019). An intelligent tutoring system and teacher dashboard to support mathematizing during science inquiry. In *Artificial Intelligence in Education: 20th International Conference*, 332–338. https://doi.org/10.1007/978-3-030-23207-8_61

[9] Nanavaty, S., & Khuteta, A. (2023). A deep learning dive into online learning: Predicting student success with interaction-based neural networks. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(1), 102–107.

[10] Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J., & Shimada, A. (2024). Artificial intelligence in education: Implications for policymakers, researchers, and practitioners. *Technology, Knowledge and Learning*, *29*(4), 1693–1710. https://doi.org/10.1007/s10758-024-09747-0

[11] Lexman, R. R., & Baral, R. (2024). Digital twins in MOOCs: Exploring ways to enhance interactivity. *Development and Learning in Organizations: An International Journal*, *38*(4), 23–26. https://doi.org/10.1108/DLO-04-2023-0091

[12] Huang, L., Zheng, J., Lajoie, S. P., Chen, Y., Hmelo-Silver, C. E., & Wang, M. (2024). Examining university teachers' self-regulation in using a learning analytics dashboard for online collaboration. *Education and Information Technologies*, *29*(7), 8523–8547. https://doi.org/10.1007/s10639-023-12131-7

[13] Mubarak, A. A., Cao, H., Zhang, W., & Zhang, W. (2021). Visual analytics of video-clickstream data and prediction of learners' performance using deep learning models in MOOCs' courses. *Computer Applications in Engineering Education*, *29*(4), 710–732. https://doi.org/10.1002/cae.22328

[14] Zhang, Y., & Li, J. (2025). Deep learning-based model for predicting student learning behavior: A pathway to early intervention and enhanced outcomes. *Journal of Computational Methods in Sciences and Engineering*, *25*(3), 2822–2835. https://doi.org/10.1177/14727978251322332

[15] Ouared, A., May, M., Piau-Toffolon, C., & Dugué, N. (2026). Automating transparent learner profiling through explainable AI. *Automated Software Engineering*, *33*(1), 33. https://doi.org/10.1007/s10515-025-00574-w

[16] Geethanjali, K. S., & Umashankar, N. (2025). Enhancing educational outcomes with explainable AI: Bridging transparency and trust in learning systems. In *2025 International Conference on Emerging Systems and Intelligent Computing*, 325–328. https://doi.org/10.1109/ESIC64052.2025.10962568

[17] Li, M. J., Li, S. T., Yang, A. C. M., Huang, A. Y. Q., & Yang, S. J. H. (2024). Trustworthy and explainable AI for learning analytics. In *Joint Proceedings of LAK 2024 Workshops*, *3667*, 3–12.

[18] Spinner, T., Schlegel, U., Schafer, H., & El-Assady, M. (2020). explAIner: A visual analytics framework for interactive and explainable machine learning. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 1064–1074. https://doi.org/10.1109/TVCG.2019.2934629

[19] Jivet, I., Scheffel, M., Schmitz, M., Robbers, S., Specht, M., & Drachsler, H. (2020). 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

[20] Chen, Z., Wang, J., Xia, M., Shigyo, K., Liu, D., Zhang, R., & Qu, H. (2025). StuGPTViz: A visual analytics approach to understand student-ChatGPT interactions. *IEEE Transactions on Visualization and Computer Graphics*, *31*(1), 908–918. https://doi.org/10.1109/TVCG.2024.3456363

[21] Johora, F. T., Hasan, M. N., Rajbongshi, A., Ashrafuzzaman, M., & Akter, F. (2025). An explainable AI-based approach for predicting undergraduate students academic performance. *Array*, *26*, 100384. https://doi.org/10.1016/j.array.2025.100384

[22] Chitti, M., Chitti, P., & Jayabalan, M. (2020). Need for interpretable student performance prediction. In *2020 13th International Conference on Developments in eSystems Engineering*, 269–272. https://doi.org/10.1109/DeSE51703.2020.9450735

[23] Islam, M. M., Sojib, F. H., Mihad, Md. F. H., Hasan, M., & Rahman, M. (2025). The integration of explainable AI in Educational Data Mining for student academic performance prediction and support system. *Telematics and Informatics Reports*, *18*, 100203. https://doi.org/10.1016/j.teler.2025.100203

[24] Perez-Berenguer, D., Kessler, M., & Garcia-Molina, J. (2020). A customizable and incremental processing approach for learning analytics. *IEEE Access*, *8*, 36350–36362. https://doi.org/10.1109/ACCESS.2020.2975384

[25] Nouira, A., Cheniti-Belcadhi, L., & Braham, R. (2019). An ontology-based framework of assessment analytics for massive learning. *Computer Applications in Engineering Education*, *27*(6), 1343–1360. https://doi.org/10.1002/cae.22155

[26] Costa, L. A., Pereira Sanches, L. M., Rocha Amorim, R. J., Nascimento Salvador, L. D., & Santos Souza, M. V. D. (2020). Monitoring academic performance based on learning analytics and ontology: A systematic review. *Informatics in Education*, *19*(3), 361–397. https://doi.org/10.15388/infedu.2020.17

[27] Evain, N., Exposito, E., Gueye, M. L., & Arnould, P. (2024). Ontology-driven approach for competency-oriented and student-centered engineering education. In *2024 IEEE Global Engineering Education Conference*, 1–10. https://doi.org/10.1109/EDUCON60312.2024.10578793

[28] Hamim, T., Benabbou, F., & Sael, N. (2020). Toward a generic student profile model. In *Innovations in Smart Cities Applications Edition 3: The Proceedings of the 4th International Conference on Smart City Applications*, 200–214. https://doi.org/10.1007/978-3-030-37629-1_16

[29] Cheng, N., Zhao, W., Xu, X., Liu, H., & Tao, J. (2024). The influence of learning analytics dashboard information design on cognitive load and performance. *Education and Information Technologies*, *29*(15), 19729–19752. https://doi.org/10.1007/s10639-024-12606-1

[30] Talbi, O., & Ouared, A. (2022). Goal-oriented student motivation in learning analytics: How can a requirements-driven approach help? *Education and Information Technologies*, *27*(9), 12083–12121. https://doi.org/10.1007/s10639-022-11091-8

[31] Hazzan-Bishara, A., Kol, O., & Levy, S. (2025). The factors affecting teachers' adoption of AI technologies: A unified model of external and internal determinants. *Education and Information Technologies*, *30*(11), 15043–15069. https://doi.org/10.1007/s10639-025-13393-z

How to Cite: Ouared, A., May, M., Piau-Toffolon, C., & Dugué, N. (2026). XAI-Profile: Explainable AI for Transparent Learner Profiling in Digital Education. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62025883