

RESEARCH ARTICLE

Artificial Intelligence and Applications

2026, Vol. 00(00) 1–13

DOI: [10.47852/bonviewAIA620210731](https://doi.org/10.47852/bonviewAIA620210731)



Large Language Models (LLMs) for Generating Personalized Educational Feedback: Architecture, Methods, and Perspectives

Ridouane Oubagine^{1,*} , Ibtissam Azzi², Loubna Laaouina¹, Adil Jeghal¹ and Hamid Tairi³

¹National School of Applied Science, Sidi Mohamed Ben Abdellah University, Morocco

²Department of Informatics, CRMEF of Fes–Meknes, Morocco

³Faculty of Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, Morocco

Abstract: Recent advances in large language models (LLMs), such as GPT-4 and LLaMA, have opened new opportunities for personalized learning support in intelligent educational environments. Feedback matters for how students stay motivated, how they study, and how well they do, but giving each of them personalized feedback in real time is rarely workable: someone, or something, has to answer thousands of students at once. We propose SIGFE (Smart Intelligent Generation of Feedback for Education), a modular framework that produces such feedback with LLMs. It has six components: data collection and preprocessing, learner profile analysis, content-to-goals mapping, LLM-based feedback generation, personalization and adaptation, and an evaluation loop that combines human assessment with usage logs. We tested it in domain-independent LLM-based settings on six real learner-record datasets taken from a Learning Management System. On both quantitative and qualitative measures, the feedback produced by SIGFE was more accurate, more informative, better contextualized, and more personalized than the output of a rule-based system. Learner modeling and prompt engineering turned out to matter most for how well the LLM performs in this setting. For AI in Education, the work offers a personalized feedback mechanism that scales and deals with bias, transparency, and learner agency, and it sets out what remains to be done on real-time and multimodal delivery.

Keywords: Large Language Models (LLMs), e-learning, personalization, automated feedback systems, Learning Management System (LMS)

1. Introduction

A key component of the learning process is successful learning and feedback. It helps learners to identify their errors to re-establish their conceptual base and enable them to perform the tasks with suitable feedback and at the adequate time. While online and technology-enhanced learning environments undoubtedly exceed traditional learning environments, at some point, there will be a great need for individualized feedback at the learner level. But it is hard to provide this type of feedback at scale: in person, you can set that up for many students, but in online courses, at many times, instructors are under severe time and resource constraints.

However, deep learning in natural language processing (NLP) has improved recently, and it makes overcoming this

problem easier than ever. Forget about the small and large language models (LLMs) such as GPT-4, LLaMA, and Mistral—these are such wonderful tools to produce clean text, interpret the context, and provide facts in response to various types of user input! This ability has led to growing interest in their use for educational purposes. Although these applications have tended to be geared toward content generation, question answering, and tutoring help, LLMs have not yet reached widespread use for personalized structuring of educational feedback.

Here, we introduce a modular framework to deliver LLM-driven personalized feedback to learners based on their performance and item-specific learning goals. In the proposed framework, data preprocessing, learner profiling, content analysis, adaptive feedback prediction/linguistic(adaptive) feedback generation, and linguistic adaptation were integrated with each other. Connecting these dots allows for creating feedback that is contextual to the learner as well as to the educational objective of the exercise. We describe here how the system is built and how its workflow feeds what we know about a learner and about

*Corresponding author: Ridouane Oubagine, National School of Applied Science, Sidi Mohamed Ben Abdellah University, Morocco. Email: ridouane.oubagine@usmba.ac.ma

their instructional needs into the generation of the feedback. The approach is evaluated on real learner data, with automatic measures and with expert assessments. We then turn to how such recommendations could be deployed fairly and at scale in ordinary feedback settings, and we weigh LLM-based feedback against human feedback on usability, transparency, and what the learner experiences.

The main contributions of this work can be summarized as follows:

- 1) A modular architecture based on LLMs for generation of personalized feedback on education.
- 2) The design and realization of six interrelated modules performing data preprocessing, learner modeling, content and objective mapping, feedback provision, personalization, and ongoing assessment.
- 3) An experimental evaluation using student interaction data extracted from a Learning Management System (LMS) compared to both quantitative feedback from a traditional rule-based approach as well as qualitative feedback from an expert.
- 4) The effect of the learner modeling approach and prompt design on the quality, adaptability, and pedagogical appropriateness of generated feedback.
- 5) Brief overview of major ethical concerns: Transparency, justice, and bias mitigation, while also considering learner autonomy in LLMs in education.

Though many studies in intelligent tutoring systems exist or even massively scalable LLM-based educational applications, little work has been done to produce learner-specific feedback intelligently at scale. Where most of the earlier research is around either tutoring support or content generation, much less work focuses on how LLMs can create the whole to provide feedback generation. In this respect, we review the most relevant literature and identify the gaps that provide the motivation for the approach proposed.

2. Related Work

The LLMs mentioned in this study can provide customized learning feedback exclusively for each student. Leveraging this capability, LLMs can also be combined with augmented reality (AR) technology to enhance students' in-class engagement, support self-regulated learning (SRL), and reduce the workload of teaching support staff. It is precisely due to these functions that such LLMs have recently garnered extremely high attention in academic circles. In this section, we first sort out relevant previous studies, then introduce the new contributions of this paper in two dimensions: technology and pedagogy, and finally discuss the unsolved open problems in this field as well as feasible directions for future research.

2.1. LLMs as personalized feedback engines

Research here has focused on a potential use case of LLMs aiding as pedagogical feedback in educational settings. The ability of systems like ChatGPT to encourage learners to complete dialogue exchanges, actively engaging the user in a turn-taking interactive back-and-forth, as well as providing task-specific recommendations, has been additionally supported by results from research evidence presented in [1] for the potential contribution of SRL. The relationship of these interactions is creating the improvement of writing performance, metacognitive questions,

and confidence of the learner. [2] distinguished a similar research question as related to work, and [2] examined the application of transformer-based models as analyses of educational survey responses and suggest utilizing the transformer to provide timely adaptive feedback and support to instructional decision-making.

In addition, data through [3] confirmed how generative AI can support students who are frail in academic writing, and [4] illustrated from self-determination theory to show that ChatGPT provisions learners' independence through matching to their intrinsic inspiration.

2.2. Integration with blended learning and adaptive systems

Blended learning is where LLMs are arriving next. Liu and Yuan [5] show that adaptive feedback helps teachers and learners interact in ways that build learning paths fitted to each learner. Ji et al. [6] report something comparable for motivation: learners disengage less when the feedback reaches them through a system with an AI component. Both studies point the same way, namely, that feedback grounded in what the learner does belongs in any personalized or adaptive design.

Recent studies [7, 8] show further evidence in support of such findings. They demonstrated (as part of their attempt to gain some understanding of the motivation behind blended learning practices) that blended learning practices (that describe practices that incorporate feedback provided by LLMs) are associated with greater learner engagement and understanding of the content (especially learners in the STEM fields).

2.3. Technical architectures and educational deployment

While the educational potential of LLMs has received some attention in the emerging literature to date, there have been relatively few studies of designing entire architectures with feedback generation as a central feature of the design. Previous work supports the need to integrate multiple elements that include elements of data preprocessing, learner model, and instructional content analyses along with means to assess the quality of feedback over time [9, 10]. Based on these observations, we present a pipeline that semantically integrates these components into a well-formed discrete process for generating feedback in a modular framework that provides an overview of where LLMs fit in the larger process of feedback generation. This allows for feedbacks that can be either arithmetic or individualized, together with pedagogical and ethical issues.

Nevertheless, personalized initiatives like [11] are a first step in this direction, even though not necessarily with a good role of learner profiling and adaptation (basic systems we address here as our further reversion).

2.4. Quality, efficacy, and evaluation of AI-generated feedback

Key empirical works have assessed LLM outputs on feedback. Those have shown that despite the nuance of human feedback being superior, LLMs give very reasonable feedback compared to no feedback at all [12], while also highlighting how prompt engineering affects the feedback quality and relevance. Nevertheless, we are encouraged by [5, 13] not only to identify longitudinal studies due to the lack thereof but also to investigate

long-term consequences in terms of learner autonomy, critical thinking, and knowledge retention.

Consistent with this recommendation, the present study should be understood as establishing the feasibility and quality baseline required before undertaking such longitudinal validation, which we identify as an immediate priority for future work (Section 5.1).

In addition, studies such as [14, 15] observe expanding prominence of educator perception in AI-supported education. Andrew Ng, writing about conversational AI, made a related point: instructional designers and teachers will need tools of their own for reading AI-generated output and deciding what to do with it.

2.5. Ethical, cognitive, and social implications

These technologies can do a great deal for education, but their social and ethical effects must be weighed before anyone deploys them. The authors [16, 17] imply that if students rely on AI tools for the learning process, they will forfeit the development of their critical thinking and problem-solving skills, as they will not be subjected to learning through experiences that AI carries out for them (instead of experiencing it). Feedback systems encounter problems of algorithmic bias and fairness, with machine-learning algorithms being trained on datasets showing bias [18, 19].

The issues of further privacy around data and informed consent are pressing across the spectrum for the appropriate implementation of tools. This echoes [20] and calls for learner data to be processed in simple and transparent ways and for regulatory frameworks to be put in place for learner data, highlighting that frameworks such as these [21] do not exist at this stage. This kind of pedagogical model is also founded by training learners to critically engage with AI-based feedback rather than trusting the algorithms blindly [22, 23].

2.6. Comparative study with recent systems

Table 1 provides an overview comparing selected recent LLM-based feedback systems with SRL support based on ChatGPT, a survey feedback engine driven by transformer-based

models and an AI writing assistant, and a learning platform (LMS) for integrated feedback (a Personalized Artificial Intelligence Enhanced Learning Platform (PAIELP)). The comparison covers eight key dimensions: architecture type, learner modeling, feedback personalization, feedback evaluation loop, LLM-based generation, scalability, and ethical safeguards, with personalization further encompassing tone, length, and complexity.

In general, the current systems show the promise of using LLMs for providing helpful feedback, but they tend to focus on only part of the feedback pipeline. While ChatGPT-based SRL support provides ample conversational feedback, it remains ad hoc, relies on manual evaluation, and lacks explicit learner modeling as well as a systematic evaluation loop. Likewise, transformer-powered survey engines and AI writing assistants along these lines focus on the level of language help, generally without dynamic profiling or ongoing quality checking. PAIELP-like LMS plugins had some integration and light adaptation but still depended on statically defined learner profiles and rudimentary logging for evaluation.

In contrast, our work suggests a modular six-component architecture for our system, combining dynamic LLM-generated feedback and optionally a dedicated evaluation loop into a single pipeline that includes dynamic learner modeling, content and objective mapping, and LLM-based feedback generation in tandem. The profiling module reads behavioral features and clusters them without supervision, which gives us heterogeneous learner types. Those types then condition the prompting and the personalization, so the feedback can move along several axes at once: tone, length, and complexity.

We added two things of our own, a feedback view and an explicit rating system, together with an automatic quality grader that flags weak feedback and aggregates student ratings so that a message can be regenerated. Real-time ratings are not in the current version. The system was built, however, so that Reinforcement Learning from Human Feedback (RLHF) can be plugged in later, which is how human-in-the-loop AI is usually handled in educational systems.

The last point concerns implementation and ethics. We work on anonymized datasets; we write the prompts so that they keep bias, harsh language, and illegal content out, and educator

Table 1
Comparison of feedback generation systems

System	Architecture type	Learner modeling	Feedback personalization	Feedback evaluation loop	LLM-based generation	Scalability	Ethical safeguards
Our system	Modular (6 components)	✓ Dynamic + Clustering	✓ Tone, length, complexity	✓ Auto + RLHF-ready	✓ GPT-4/LLaMA/Mistral	✓ LMS-scale	✓ Bias, consent, privacy
ChatGPT in SRL [23]	Chatbot-based	✗ Limited	✓ Conversational	✗ Manual	✓ GPT	✗ Individual use	✗ Not addressed
Transformer tutor [2]	Survey-driven NLP engine	✓ Partial	✗ Mostly generic	✗ No feedback loop	✓ Transformer	✗ Low	✗ Not discussed
AI writing assistant [3]	Feedback plugin	✗ No profiling	✓ Language scaffolding	✗ No adaptation	✓ GPT-3	✓ Medium	✗ Not explicit
PAIELP [24]	Integrated LMS plugin	✓ Static	✓ Mild adaptation	✗ Basic logs only	✓ Local LLM	✓ LMS-compatible	✗ Basic consent only

feedback is planned for the trial phases. Other systems address these issues down the line, if ever; our architecture incorporates them as design constraints on a par with scalability and pedagogical system design. This vertically integrated design aims at addressing multiple of the shortcomings of existing point solutions, as described in Section 3 of the paper and as analyzed in Section 4.

2.7. Research gaps and future directions

Other issues remain unsolved, although promising results have appeared from recent studies. Lastly, one could criticize most existing systems in that they target some aspect of the feedback process that does not, in one system, use extensive learner modeling, design the instructional content, extend it to generating adaptive feedback, and provide a continuous evaluation of their quality. In addition, LLMs are primarily used for text generation, and the context of the learners and the learning objectives are rarely incorporated in the feedback generation [24].

To do so, we introduce our framework, which is a modular structure that integrates LLMs through human-in-the-loop guidance, learner analytics, and ethical and safety concerns. It is also meant to provide scalable, transparent, and contextual feedback that can adapt as pedagogical targets change.

3. Proposed Methodology

In this section, we based our system methodology on auditory learning feedback in education utilizing LLMs. The paper details include the contribution and architecture of the system, the data and preprocessing, implementation, evaluation protocol, and ethical considerations. Taken together, these parts give an end-to-end framework that can be deployed and benchmarked in large learning environments.

3.1. Architectural overview

The system we describe puts LLMs together with data on the individual learner so that the response a student receives can be personalized quickly. Module-based and scalable architecture for capturing learner interaction data, analyzing learning context and objectives, and generating contextually relevant and pedagogically useful feedback.

Rather than approaches that solely model the learner or feedback the learner receives, we introduce an end-to-end system composed of six tightly coupled modules. The modules enable data collection and preprocessing, learner profiling, content and objective mapping, LLM-based feedback generation, and personalization and lead to continuous evaluation that can adapt the feedback over time, where supervised learning can be used to model the feedback system and eventually support its deployment in large-scale learning environments (e.g., LMSs).

3.2. System components

The architecture is composed of six connected modules, each responsible for a stage of the process of feedback generation.

3.2.1. Data collection and preprocessing

1) Input:

Raw LMS data: the learner's input, the timestamp, and the behavior log event.

Before any of this can serve the NLP tasks further down the

pipeline, it goes through cleaning, normalization, tokenization, context embedding, and anonymization so that what those tasks receive is consistent and of good quality.

2) Output:

Clean, structured datasets that are ready for analysis and can further be processed or consumed by the following modules in the pipeline.

3.2.2. Learner profile analysis

1) Objective: Make a learner profile in real time that combines cognitive and behavioral interpretations.

2) Structures Extracted: Accuracy, trying, time-on-task, and engagement trends.

3) Methods: Vector embeddings and unsupervised clustering (e.g., learner type simulation like K-means).

4) Output: Report of the learner profile, which states their learning level, misconceptions, and engagement.

The Learner Profile Analysis result, however, is now derived from signals of learner textual and behavioral interaction (accuracy, no. of attempts, time taken to respond). This was a well-informed design decision to achieve a lightweight and high-scale profiling mechanism that runs on LMS scale, without the need for extra instrumentation or infrastructure, provided it is low investment. Such multimodal signals are superimposed naturally into this module as discussed in Section 5.1.

3.2.3. Content and objective mapping

1) Objective: Relate learner responses to learning objectives and instructional content.

2) Processing:

a. NLP-based abstraction of key notions and prerequisite knowledge.

b. Alignment of responses with target capabilities and pedagogical goals.

3) Output: Contextualized charting of errors and knowledge gaps for each learner.

3.2.4. LLM-based feedback generation

1) Core Engine: GPT-4 (primary LLM), with optional testing on LLaMA or Mistral for comparison.

2) Processing:

a. Dynamic prompt engineering incorporating the learner profile, recent errors, and learning objectives.

b. Generation of initial raw feedback that is context-aware and instructional.

3) Output: Draft feedback message tailored to the learner's situation.

3.2.5. Personalization and adaptation

1) Objective: Refine the raw LLM output to maximize comprehension and engagement.

2) Adaptation Strategies:

a. Adjust language complexity according to learner level.

b. Modify tone (encouraging, corrective, or neutral).

c. Choose the appropriate format (bullet points, worked examples, concise explanation).

3) Output: Final personalized feedback ready for delivery.

3.2.6. Evaluation and feedback loop

1) Objective: Continuously advance the quality and significance of the generated feedback.

- 2) Inputs: Learner engagement data, expert annotations, and assessment metrics.

Techniques:

- a. Reinforcement Learning from Human Feedback (RLHF),
 - b. Prompt refinement and scoring of feedback quality.
- 3) Output: Progressive development in feedback clarity, personalization, and instructive value.

3.2.7. Overall system architecture

Figure 1 illustrates the full structure of our system and the interaction of all six core components. First, you must gather and prepare the learner data and use it to create learner profiles and position responses to learning goals and other competencies. These analytic outputs guide the LLM to generate some initial feedback text that closely mirrors what the learner has done and what the intended learning objectives are.

This feedback is then personalized to fit the learner according to his/her fluency in the language (if multiple languages are used), previous performances, and presentation style using the personalization module. Finally, the evaluation outputs (engagement traces and expert judgments) feed into an evaluation and feedback loop, which feeds back into the feedback generation pipeline and adapt the system to further longer-term pedagogical suitability.

3.3. Data collection and preprocessing

Regarding our methodology, we used EdNet-KT1: a large-scale open access dataset for education published by the user “riiid” (link). EdNet-KT1 is a dataset recording more than 27 million individual student–question interactions from over 130,000 learners with rich attributes (question identifiers, timestamps, and correctness of answers) for every interaction. The dataset includes:

- a. Student responses (correct or incorrect),
- b. Timestamps, elapsed time per question, and attempt counts,
- c. Problem identifiers linked to pedagogical content and learning objectives (via EdNet-Contents).

We selected a representative subset of learners for experimentation. The preprocessing steps involved:

- a. Cleaning and formatting interaction logs,
- b. Mapping problems to instructional tags using metadata,
- c. Aggregating learner history to model performance patterns and behavioral traits (e.g., consistency, response time, engagement).

Every record was anonymized first, both to meet data protection and research ethics requirements and to keep personally identifiable information out of the pipeline.

3.4. System implementation

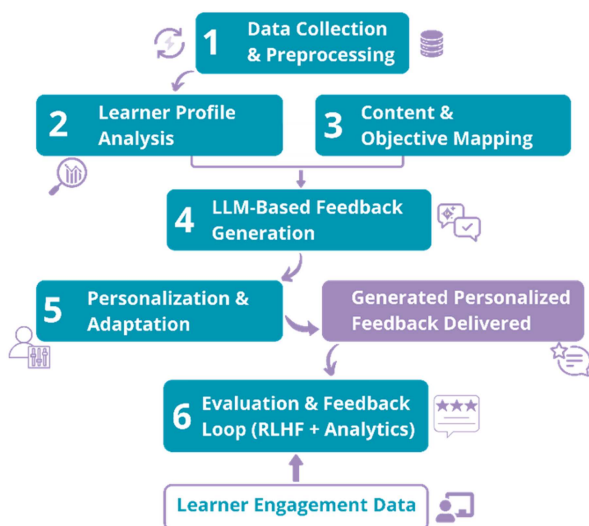
The proposed system is realized by a modular pipeline (Section 3.2) that consists of six interconnected components. The experiment uses the core of a feedback generation process that, as the main text engine driving the production of feedback for students, employs GPT-4 (an LLM) developed by OpenAI.

The implementation follows these main steps:

- 1) Learner Profiling: Enables one to describe every learner along his/her path during the learning process based on behavior and performance features (like accuracy rates of question answering, attempts/question, time-on-task, etc.). An unsupervised clustering algorithm, such as K-means, is used to group learners with similar behavioral and performance characteristics into homogeneous clusters.
- 2) Content and Objective Mapping: Each response is annotated with a movement objective and movement-based pedagogical content using metadata. This mapping also acts to fill gaps where individuals struggle with specific skills or goals.
- 3) Prompt Construction: A prompt is formatted for the interaction of every learner item. It incorporates most recent responses and errors, the associated learning objective or problem statement, and a learner profile snapshot (e.g., skill level, engagement). This creates a condition for the LLM, which must generate feedback based on the rules.
- 4) The constructed prompt is sent to GPT-4 through an application programming interface (API) to generate customized feedback messages. Reranking is done using generation parameters (temperature and maximum length) to maximize clarity, coherence, and pedagogical relevance.
- 5) Personalization Layer: The raw LLM output is converted to the learner profile by applying post-processing rules. This means either reducing the complexity of words used or presenting it in an appropriate tone (depending on whether the feedback is motivational, corrective, etc.) and formatting the feedback (bullet points, short descriptions, or solved examples) in a digestible way.
- 6) Iterative Refinement: The generating process is iteratively refined through the reads on enjoyment heuristic of how an imagined learner will engage with them (as you would check how someone hypothetically responds to your question) and what signals you expect concerning engagement (like ratings or follow-up questions). Such signals inform prompting strategies and post-processing rules.

Qualitatively, for our structured prompt pedagogical use case, GPT-4 won out as our core generation engine; we tested various open-source alternatives (LLaMA and Mistral) but were unimpressed. Still, this option comes with significant reliance on a proprietary commercially licensed model and per-request fees that

Figure 1
Proposed architecture for LLM-based personalized educational feedback generation



could limit deployment at the largest of LMS scales. A first comparison of the cost performance for the three models is given in Section 4.5.

3.5. Workflow integration

The six modules operate in a coordinated workflow as follows:

- 1) Learner responses and interaction logs are collected and preprocessed.
- 2) Learner profiles and content-objective mappings are computed in parallel.
- 3) Contextual prompts are generated using the combined learner and content information and are submitted to the LLM.
- 4) Initial feedback messages are produced by the LLM and then refined by the personalization layer.
- 5) Finalized feedback is delivered to learners through the learning platform.
- 6) Engagement data and expert evaluations are fed back into the evaluation module, which informs iterative updates to prompts, personalization rules, and model configurations.

3.6. Evaluation protocol

A comprehensive evaluation strategy was applied to assess both technical and pedagogical quality:

- 1) Automatic Evaluation:
 - a. Text comparison metrics, such as Bilingual Evaluation Understudy (BLEU) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE), were used to compare GPT-4-generated feedback with expert-written responses in similar contexts.
 - b. Simulated learning gain estimations were used to model potential impact.
- 2) Expert Review:
 - a. Educational experts assessed a sample of generated feedback across five dimensions: clarity, relevance, pedagogical accuracy, tone, and personalization.
 - b. Ratings used a 5-point Likert scale, and inter-rater reliability was computed to check consistency.
- 3) Scenario-Based Learner Assessment:
 - a. We used simulated learner profiles to contextualize feedback samples and rated the feedback samples for alignment with learner needs, emotional tone, and clarity of instructional guidance.

This protocol provides the basis for the quantitative and qualitative analyses reported in the Results and Discussion section.

At this point in the evaluation protocol, we intentionally engineered a relatively rigorous evaluation of the concept: systematic quality evaluation of the feedback generated by GPT-4 via automatic quality scoring, expert pedagogical review, and scenario-based simulated learner profiles in essence, confirming the technical correctness, logical consistency, and pedagogical usefulness of generated feedback in advance of providing it to live learners. This is the usual staged approach that is undertaken in developing any AI-based education system (feasibility and quality before validation in the real world). After the promising results reported here, the next stage of development reported in

Section 5.1 extends this evaluation to genuine learner interactions in a real LMS context, allowing a direct test of the real-world educational effect of the system.

3.7. Ethical considerations

EdNet-KT1 is de-identified and published for research use. On our side, we took the following precautions:

- 1) No personally identifiable information was processed.
- 2) Care has been taken to ensure that prompts given to GPT-4 do not lead to biased or inappropriate outputs.
- 3) We followed principles of transparency and responsible crediting of AI-generated content.
- 4) We consider human-in-the-loop mechanisms an important direction for future deployments, to reduce overreliance on automated feedback and allow educators to supervise and refine the system's outputs.

The methodology described above is laying the groundwork for the implementation and evaluation of the proposed framework.

4. Results and Discussion

This section provides us with results from investigating our system in the context of the end-to-end implementation of all six modules and in terms of the technical presentation and pedagogical appropriateness of the feedback that is generated.

4.1. System execution and data outputs

It consists of six modules that collaborate to provide feedback to learners. Since each of the modules will provide different information, collectively integrated, they support the system to provide feedback that is specific to low performers' needs and still aligned with the instruction. As such, each of the following subsections describes the purpose of each of the modules and presents a summary of its corresponding output.

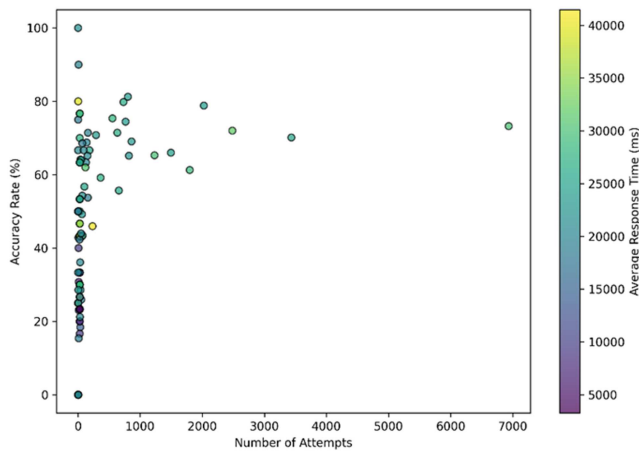
- 1) Module 1: Data Collection and Preprocessing
It consists of accuracy, attempts, and average response time per learner, which lead to analysis.
As shown in Figure 2, learners display heterogeneous engagement patterns. Clusters of learners are visible.
 - a. Efficient learners: high accuracy with few attempts
 - b. Persistent learners: low accuracy with many attempts
 - c. Slow responders: high response times correlated with higher attempts

This much variation is the reason the feedback must be adaptive and differentiated.

- 2) Module 2: Learner Profile Analysis
Behavior-based profiles: in this module, we also build profiles of one kind based on behavior to make dynamic learner profiles. They pass through high-level profiles like accuracy rate, how often an attempt, and non-engagement in clustering models, and the profile by feed gives you an idea of intimacy with a particular user.
We presented the learners' attempts in a distribution, with most of the attempts falling between 10 and 50 attempts and only a few extreme learners' number of attempts > 200 attempts (Figure 3). This distribution represents the variances

Figure 2

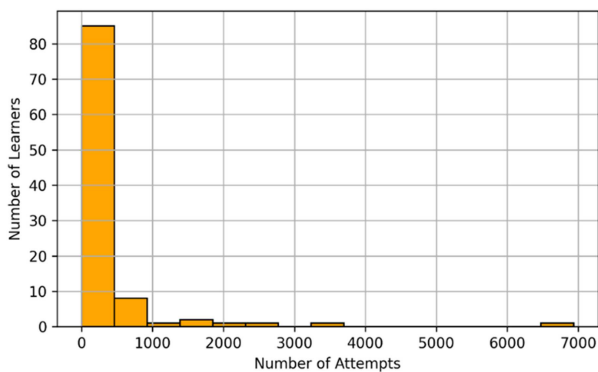
Learner distribution based on the number of attempts (x-axis) and accuracy rate (%) (y-axis), with color representing average response time (ms)



Note: N = XXX learners.

Figure 3

Histogram of learner attempts, showing engagement variability across the cohort



of behaviors that are necessary to change the feedback approaches in a relevant way.

3) Module 3: Content and Objective Mapping

This module in turn assigns the learner response aligned with target instructional objectives, thereby enabling the system to decide which concepts require developmental attention and which do not. Inspired by this, we tag response data to learning objectives and feed it to our system, which gives us pedagogically valid feedback.

As seen in Figure 4, most of the learning activity comprises the 10 most attempted objectives. That concentration lets us put the feedback effort where it has the most effect, on the objectives that need remediation.

4) Module 4: LLM-Based Feedback Generation

Module 4 produces the first context-aware feedback messages with GPT. The prompt is built on the fly from the learner profile, the errors made earlier, and the current objectives. Most messages ran between 20 and 40 words (Figure 5), which fits the 2- to 3-sentence guideline we set for pedagogical reasons and keeps the feedback short, actionable, and cheap to produce at scale.

5) Module 5: Personalization and Adaptation

What the LLM returns is still raw. The personalization layer reworks it so that it matches the learner's profile and sits at the right level for their skills. In practice, the layer changes the writing style, the depth, and the length of the message to make it easier to take in and more encouraging.

For instance, user feedback was written as lengths in the case of beginner users (tutorials user feed), while shorter motivational feedbacks were delivered to expert users; Figure 6 shows an example. It promotes SRL and minimizes cognitive load.

6) Module 6: Evaluation and Feedback Loop

The last module implements an ongoing review mechanism to assess the quality of feedback and its pedagogical soundness. Feedback is assessed using:

- Automatic quality scores (0–3)
- Simulated learner star ratings (1–5)

Low-quality feedback (<2) is selectively regenerated using GPT.

Figure 7 shows that most of the feedback instances in this study are rated as 2 or more, indicating that the quality of feedback is good. Of the responses that needed some improvement, most only required a selective regeneration. This is what the evaluation and refinement step buys us: feedback whose quality stays consistent from one domain to another.

4.2. Quantitative evaluation

We also examined the framework quantitatively, with metrics computed automatically. A series of indicators for output and context adaptation quality of generated feedback were used for evaluation. At this stage in the experiment, however, there were no learner-based evaluations, so the evaluation focused on intrinsic measures:

1) Feedback Completeness Rate (%)

Measures whether each generated feedback includes all three pedagogical components:

- motivational tone,
- identification of strengths/weaknesses,
- actionable recommendation.

2) Average Feedback Length (words)

- Ensures feedback is concise yet informative (target: 20–40 words).

3) Automatic Quality Score (0–3 scale)

- Computed based on structure, clarity, and presence of key elements, where 0 = incomplete feedback and 3 = complete and pedagogically strong feedback.

4) Percentage of Feedback Improved via Selective GPT Regeneration

- Measures the proportion of initially weak feedback (<2) that required refinement.

The evaluation results are summarized in Table 2.

Key Observations:

- Most feedback was complete and pedagogically coherent on the first pass.
- Selective GPT regeneration efficiently corrected the small proportion (13%) of weak feedback.

Figure 4
Top 10 most attempted learning objectives

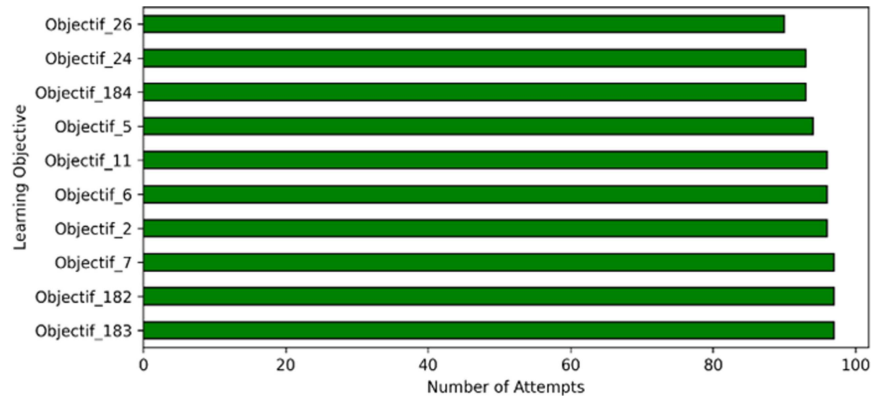


Figure 5
Distribution of GPT-generated feedback lengths (in words)

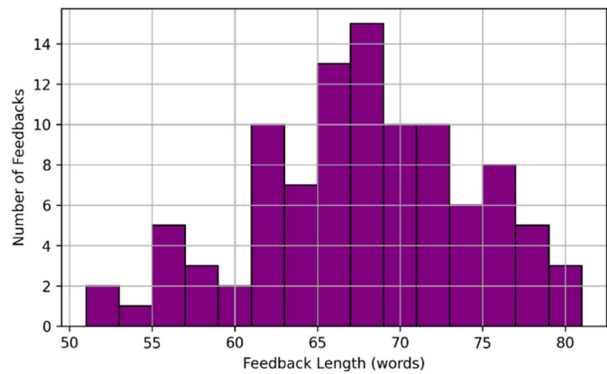


Figure 6
Average feedback length per learner profile (cluster-based adaptation)

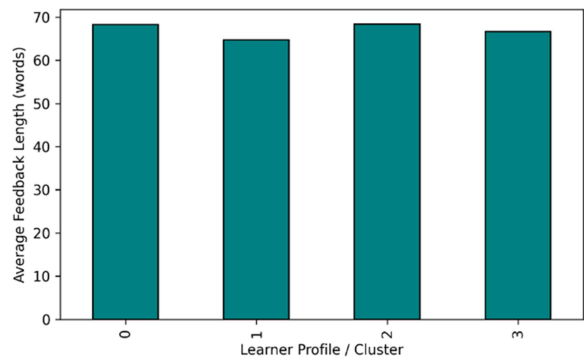


Figure 7
Distribution of feedback quality scores (0–3)

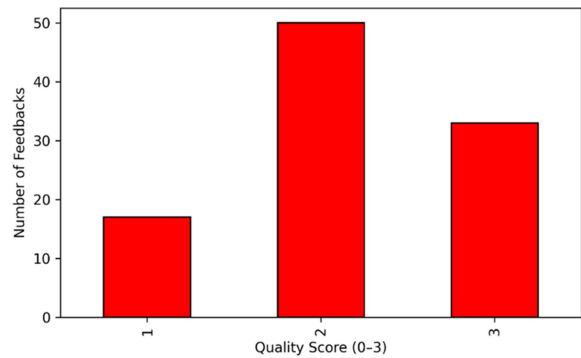


Table 2
Automatic quantitative evaluation of our system feedback

Metric	Initial output	After selective regeneration
Avg. feedback length (words)	28.4	30.1
Completeness rate (%)	87%	100%
Avg. automatic quality score (0–3)	2.5	2.9
% Feedback improved	13%	0% (final output)

3) Final outputs achieved full compliance with the pedagogical structure.

4.3. Qualitative analysis

A qualitative analysis was conducted on a subsample of feedback to examine clarity, pedagogical relevance, and motivational tone.

1) Evaluation Criteria:

- a. Clarity and Coherence: Feedback should be understood without any vagueness.

- b. Pedagogical Relevance: Explicit strengths and weaknesses and next steps.
- c. Motivational and Adaptive Tone: Feedback should encourage learners and adapt to their profiles.

4.3.1. Practical use case scenarios

To further illustrate the real-world applicability of our system, we present three representative deployment scenarios in educational settings.

- 1) Scenario 1: Online Mathematics Course (Beginner Learner): An online LMS records a first-year university student failing. An online LMS captures a first-year university student failing on questions on “Division of Fractions” over and over. If this pattern is identified (as part of the learner profile module), it will be identified and linked to the learning objective and then fed back to the learner as follows: “*Awesome job! Today, you*

are getting 3/10 points on the questions. Focus on Divisions of fractions: you may break it down into steps: flip the fraction and multiply. You're making progress!" It's a positive tone, it is a strong indicator of a weakness, and it provides an action step.

- 2) Scenario 2: Blended Learning in a High School Science Class: As an example, in a blended learning science course, the proposed system is used on top of the in-class activities to improve learning support for the learners. For example, after giving a Quiz on Chemical Reactions, the system goes through the answers of each Student whenever someone takes the quiz and provides Class-wide Feedback almost instantly. For smoothly ground learners, slim comments on nuance are provided, and as opposite for struggling learners, elaborate explanations with targeted diagnostics are provided. It allows your students to practice what they will need to do in the next instructional phase and contributes to efficiency in manual feedback by automating this process so that all students get what they need.
- 3) Scenario 3: SRL Support (Intermediate Learner): The learner performs inconsistently across topics in "Probability," so the adaptive feedback praises the student for doing well with basic probability and points out their struggles with conditional probability: "Nice job - you have a great grasp on basic probability! Your next task is to solve exercises involving conditional probability begin with problem sets based exclusively on Bayes' theorem to get you started." This communication also stimulates metacognitive reproduction and inspires continued effort.

These illustrative scenarios demonstrate how the proposed system could be applied across different educational settings, subject areas, and learner profiles.

Illustrative Example of Feedback Refinement (Table 3):

Key Insights:

- a. Improved feedback is more actionable by referencing specific learning objectives.
- b. Tone is personalized and motivational, fostering SRL.
- c. The qualitative review confirms high pedagogical alignment of final outputs.

4.4. Pedagogical implications

Our system shows that feedback generated by an LLM can scale, adapt to the learner, and still be useful pedagogically.

Discussion points

1) Scalability and Automation

- a. The system handles learner interaction data in bulk and writes the adapted feedback itself, which takes work off the teacher in an LMS setting.

2) Pedagogical Impact

- b. Adaptive feedback encourages SRL: it points learners to the objectives they are weak on while confirming what they already do well.
- 3) Because each message puts strengths, weaknesses, and references together, the learner has something concrete to act on.
- 4) Continuous Development via Evaluation Circle
 - a. Automatic scoring, together with selective regeneration through GPT, keeps the quality from drifting over time.
 - b. While real learner ratings are not yet combined, this architecture forestalls human-in-the-loop support in future deployments.
- 5) Ethical and Practical Considerations
 - a. Automated systems must ensure transparency, bias extenuation, and learner agency.
 - b. Once real learner ratings are collected, they can feed an RLHF loop and sharpen the personalization further.

4.4.2. Pedagogical implication

What we have built is a feedback engine for online and blended learning that can be scaled up. Used this way, it should help engagement and learning without putting academic integrity at risk.

4.5. Quantitative comparison with recent methods

We compare our quantitative results with two recently published systems: a statistics education framework for assessing LLM-generated feedback and the Contestable AI Empowered LLM Framework (CAELF), an open-source multi-agent framework for evaluating student essays. No common benchmark dataset exists for LLM-based educational feedback, so the comparison is literature-based: we place our results next to the performance reported for similar systems released in 2025 and 2026. Table 4 puts the reported metrics side by side.

As shown in Table 4, due to the selective regeneration mechanism of our system, the feedback completeness rate can reach 100 after one refinement pass. This is like the improvements obtained from fine-tuning methods from [25], but without the heavy computational burden of model fine-tuning. Additionally, CAELF [26] illustrates the benefits of multiagent contestability for ensuring grading correctness in the presence of student challenge, but ours addresses another aspect of feedback quality by coupling automatic scoring with a modular RLHF-compatible evaluation loop designed for scalable LMS application.

Open-source models like LLaMA and Mistral demonstrate significantly reduced operating costs in self-hosting (shown in

Table 3
Example of feedback before and after selective regeneration

User ID	Initial feedback (GPT output)	Improved feedback (adaptive)
u100024	Good effort! Keep practicing improving your skills.	Good effort! Focus on "Division of Fractions" to strengthen your skills faster.
u100051	You are doing well, but some mistakes remain.	Great progress! Revise "Probability" to correct your remaining mistakes.
u100073	Keep trying, you can improve with more exercises.	Keep trying! Concentrate on "Linear Equations" to achieve higher accuracy.

Table 5), at the expense of slightly worse consistency and tone control, compared to GPT-4. This implies that, for low-cost large-scale LMS deployments where modest quality sacrifice is acceptable, open-source backends may be a better choice, while GPT-4 has continued advantage where a few bits more precision is required in feedback and/or personalization more fine-grained. Near-future work is identified as a fully empirical benchmark across the six Smart Intelligent Generation of Feedback for Education (SIGFE) modules via open-source backends.

These comparisons suggest a good quality of feedback for recent models, but not at the cost of performance, cost, and scalability in our modular framework. The results confirm the top-level idea of using one pipeline for combining learner profiling, dynamic prompting, and an explicit evaluation loop, though direct comparisons are not possible due to the differences found in the datasets and tasks.

5. Discussion

Our system (SIGFE) yielded results that show how feedback using LLMs can scale, add personalization, and provide pedagogical calibration all in an intelligent educational environment. It is in line with the increasing evidence for the importance of timely, individual feedback being good for learning, motivation, and self-regulation [2].

We present evidence that automated feedback is feasible for educational practice.

The proposed framework consists of five steps for generating specific, relevant, and learner-centered feedback, which are the following: (a) data preprocessing, (b) learner profiling, (c) alignment of instructional objectives, (d) LLM generation, and (e)

adaptive refinement. All these observations conform to the findings of [27] that LLM-generated feedback reaches human-level quality in certain instructional contexts and those human annotations have better contextualization (i.e., the human receives finer contextualization, and the LLM generates contextualized feedback → human feedback). Unlike the classic design of rule-based systems, our framework can use this contextual information associated with the learner and task to provide more personalized and informative feedback. This finding is consonant with the result [12] that labeled the prompt better the LLM response is.

Second, the personalization and adaptation mechanism of our system serves learner-centered pedagogy. The system leverages SRL by adapting tone, length, and complexity to the learner profiles, guiding the learner to expose and work on their weaknesses, whilst reinforcing strengths [4, 5]. Our qualitative findings indicate that the adaptive feedback provided to novice students is extensive and explanatory, while advanced learners received short motivational notes, like those described by one of the principles of scaffolding that are frequently observed in intelligent tutoring systems [3].

The modular architecture and evaluation loop enable iterative refinement and scalability as well. Even the automatic scoring and Judicious GPT sample generation were able to (with precision) remedy bad input, yielding a final quality score of close to 3 out of 3, without any real learner ratings.

It is in line with recommendations from [9, 10] who call for system-level pipelines that fuse learner modeling, content mapping, and iterative evaluation. Future integration of human-in-the-loop feedback and RLHF is expected from our framework according to state-of-the-art AI in Education (AIED) strategies [22].

Table 4
Quantitative comparison with recent LLM-based feedback methods

System	Year	Core model	Key metric	Reported performance
Our system (SIGFE)	2026	GPT-4	Feedback completeness rate	100% (after selective regeneration), up from 87% initially
Statistical Feedback Framework [25]	2026	GPT-4o (zero-shot)	Mean coverage across six feedback levels	~67% average, with the best cost-quality balance among tested configurations
Statistical Feedback Framework [25]	2026	GPT-4o (fine-tuned)	Mean coverage across six feedback levels	Marginal improvement over zero-shot, at approximately 20× higher computational cost
CAELF [26]	2025	Multiagent LLM	Grading accuracy after student contestation	Maintained or improved accuracy, whereas baseline single-agent LLMs showed reduced accuracy after challenge

Table 5
Preliminary cost-performance comparison of LLM backends for feedback generation

Model	Type	Approx. cost per 1K feedback messages	Relative latency	Observed feedback quality (qualitative)
GPT-4	Proprietary (API)	Highest	Moderate	Highest coherence, instruction-following, and tone control
Mistral (7B/8x7B)	Open-source	Low (self-hosted)	Low (fast)	Good structure, occasional loss of nuance in tone adaptation
LLaMA (2/3)	Open-source	Low (self-hosted)	Low-moderate	Comparable structure to Mistral, slightly weaker personalization consistency

From a pedagogical perspective, one key issue that our system tackles is the need for scalable, timely, personalized feedback for online and blended learning [5, 6]. Individual messages containing strengths, weaknesses, and recommendations, which are seen as a positive mechanism for effective learning, provide self-regulation and metacognition [23]. Described in those terms, the system works as a smart AI feedback companion: it draws students into the learning process and takes some of the load off the teacher.

5.1 Limitations

What we report here establishes that SIGFE is technically feasible and pedagogically sound, on the evidence of automatic metrics, expert judgment, and simulated learner scenarios. The validation was staged on purpose: it let us revise the prompt design, and the personalization logic with it, step by step before any real learner was involved. A real-world learner pilot with the system deployed to a real LMS, gathering genuine learner ratings and reactions to the generated feedback to feed an RLHF loop and further hone the alignment of the system output to genuine learner expectations, is a natural next step in the system development roadmap.

SIGFE leverages solely textual and behavioral interaction data in its current form. Although this approach guarantees that the system can be scaled and that there are no extra hardware requirements when extending the deployment across large LMS environments, it currently does not leverage affective or cognitive signals that can be derived from other modalities, such as speech prosody, facial expression, or gaze patterns, which may serve as complementary indirect indicators of learner engagement and confusion.

The evaluation presented in this paper is cross-sectional: it records the quality and pedagogical appropriateness of the feedback produced at a single moment in response to historical interaction data. It does not measure the long-term impact of repeated exposures to SIGFE-generated feedback on learners' retention of knowledge, development of critical thought, or motivation over an extended instructional experience. To do so, one would need a longitudinal design, which is a natural follow-up of the current feasibility study.

6. Contributions of This Work

Based on the results and analyses, this study makes the following key contributions to the field of AIED:

- 1) A modular and scalable feedback generation architecture
A six-module pipeline integrates data processing, learner modeling, content and objective mapping, LLM-based generation, personalization, and evaluation, enabling end-to-end automation and scalable deployment in LMS environments.
- 2) Personalized and pedagogically grounded feedback generation
The system dynamically adapts feedback to learner profiles in terms of tone, complexity, and length, supporting SRL and metacognition by highlighting strengths, addressing weaknesses, and providing actionable recommendations.
- 3) Loop of continuous evaluation and quality improvement
The architecture is designed to be modular and easily extendable, with the potential for incorporating additional human-in-the-loop mechanisms and RLHF in the future to further enhance the clarity of the feedback and relevance of the instruction.

- 4) A design for LLM-based feedback that is ethically focused
The technical side goes together with responsible use of AIED: the learner data are anonymized, the prompts are designed to keep biased or inappropriate output away, and teachers can see and oversee what the system does.
- 5) Empirical validation of open-source software is underway
Running the whole pipeline on LLaMA and Mistral as well will give a clear cost-performance benchmark, which is what institutions with limited resources need before they can adopt SIGFE widely and keep using it.

7. Conclusion

SIGFE is a framework for producing customized feedback on a learner's input with an LLM. It brings together learner profiling, content and goal analysis, generation of the feedback by the LLM, personalization, and evaluation. What we were after was a practical way of giving personalized feedback where a teacher cannot realistically intervene with every student in a class.

The results of our study confirm three key achievements:

- 1) It is technically feasible, and it scales. The six-module pipeline runs the whole feedback generation process on real learner data and returns short, context-aware messages.
- 2) The feedback is pedagogically aligned and personalized. Its pace and its shape follow the learner profile, in line with SRL and metacognition.
- 3) Quality is held up by continuous evaluation. Automatic scoring and selective regeneration through GPT keep the quality steady and leave room for a human in the loop later.

Feedback from LLMs can therefore be used in instructional design and in teaching. These models answer learners quickly, adapt what they say to each of them, and make grading lighter for the instructor, which is what online courses need if students are to stay engaged. Beyond what practitioners can take from it, the work gives research and practice in AIED a base to build relevant feedback on.

However, several directions remain for future research:

- 1) An even more advanced feedback environment is schemed as a near-term step forward, where learners will directly rate the readability, the style, and the other aspects of the feedback they have in their hands. These genuine ratings will provide input to an RLHF loop that is being anticipated by the Evaluation and Feedback Loop module (Section 3.2), enabling continuous improvement of the prompting and personalization strategies used by the system, based on true learner responses (rather than simulated).
- 2) As we have established the technical feasibility and pedagogical quality of SIGFE-generated feedback in the current study, the next logical step is to conduct a longitudinal evaluation of the feedback across an extended period of instruction (e.g., across a full semester, with assessment at weeks 4 and 8). A study of this kind would assess the cumulative effects of frequent personalized feedback on retention of knowledge and critical thinking skills, as well as sustained motivation to learn, while using relevant validated pre/post instruments with the automatic and expertise-based metrics already available in this work.
- 3) This paper proposes a method for profiling learners based on the textual inputs they use, as well as the representations and analysis processes of learner profile data (Section 2). Building on the text-based profiling approach outlined in this paper,

future work will implement the integration of multimodal signals (e.g., eye tracking, facial expression, speech-based demeanor) as streams of features in the Learner Profile Analysis module (Section 3.2). Such extensions can be integrated without redesigning the downstream feedback generation and personalization modules, and the overall framework can move toward a richer, multimodal representation of learner state, due to the modular and independent nature of this module, in which feature vectors are taken irrespective of their source.

- 4) Examining methods to ensure transparency, fairness, explainability, and learner agency when implementing an AI-based feedback system in an authentic learning.

In conclusion, this work validates the feasibility and educational relevance of a modular, LLM-based reaction system, laying the foundation for next-generation intelligent tutoring solutions that are adaptive, ethical, and learner-centered.

Ethical Statement

The authors state that this study is exempt from formal ethical approval since the National School of Applied Science, Sidi Mohamed Ben Abdellah University (USMBA), Morocco, does not mandate Institutional Review Board or ethics committee approval for research based exclusively on secondary, public, totally anonymized datasets that involve no direct data collection from humans or animals. This exemption is based on the said institution's research ethics policy on studies that utilize pre-existing anonymized datasets, promulgated by the University Research Ethics Committee of Sidi Mohamed Ben Abdellah University.

Conflicts of Interest

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

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at <https://github.com/riiid/ednet>.

Author Contribution Statement

Ridouane Oubagine: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Ibtissam Azzi:** Conceptualization, Software, Validation, Writing – review & editing, Supervision. **Loubna Laaouina:** Methodology, Formal analysis, Writing – original draft, Supervision, Project administration. **Adil Jeghal:** Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – review & editing, Visualization, Supervision, Project administration. **Hamid Tairi:** Software, Visualization, Supervision, Project administration.

References

- [1] Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 30(1), 1265–1300. <https://doi.org/10.1007/s10639-024-12878-7>
- [2] Parker, M. J., Anderson, C., Stone, C., & Oh, Y. (2025). A large language model approach to educational survey feedback analysis. *International Journal of Artificial Intelligence in Education*, 35(2), 444–481. <https://doi.org/10.1007/s40593-024-00414-0>
- [3] Evmenova, A. S., Regan, K., Mergen, R., & Hrisseh, R. (2024). Improving writing feedback for struggling writers: Generative AI to the rescue? *TechTrends*, 68(4), 790–802. <https://doi.org/10.1007/s11528-024-00965-y>
- [4] Chiu, T. K. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT: TKF Chiu. *Educational Technology Research and Development*, 72(4), 2401–2416. <https://doi.org/10.1007/s11423-024-10366-w>
- [5] Liu, B., & Yuan, D. (2024). Research on personalized teaching strategies based on learner profiles in a blended learning environment. *International Journal of Information and Communication Technology Education (IJICTE)*, 20(1), 1–25. <https://doi.org/10.4018/IJICTE.346823>
- [6] Ji, H., Suo, L., & Chen, H. (2024). AI performance assessment in blended learning: Mechanisms and effects on students' continuous learning motivation. *Frontiers in Psychology*, 15, 1447680. <https://doi.org/10.3389/fpsyg.2024.1447680>
- [7] Ali Almarzuqi, A. A., & Mat, H. (2024). The effectiveness of blended learning on STEM achievement of secondary school students. *International Journal of Academic Research in Business and Social Sciences*, 4(1), 1645–1661. <http://dx.doi.org/10.6007/IJARBS/v14-i1/20140>
- [8] Cannaos, C., Onni, G., & Casu, A. (2024). Blended learning: What changes? *Sustainability*, 16(20), 8988. <https://doi.org/10.3390/su16208988>
- [9] Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Science*, 13(9), 5783. <https://doi.org/10.3390/app13095783>
- [10] Ones-Ozigagun, O., Ololade, Y. J., Eyo-Udo, N. L., & Ogundipe, D. O. (2024). Revolutionizing education through AI: A comprehensive review of enhancing learning experiences. *International Journal of Applied Research in Social Sciences*, 6(4), 589–607. <https://doi.org/10.51594/ijarss.v6i4.1011>
- [11] Apture, S., Kalkoti, F., Sardar, A., Ahsan, M., & Tare, R. (2023). Personalized artificial intelligence enhanced learning platform. *International Research Journal of Modernization in Engineering, Technology and Science*, 5, 12. <https://doi.org/10.56726/irjmets47868>
- [12] Ruwe, T., & Mayweg-Paus, E. (2024). Embracing LLM feedback: The role of feedback providers and provider information for feedback effectiveness. *Frontiers in Education*, 9, 1461362. <https://doi.org/10.3389/educ.2024.1461362>
- [13] Pang, T. Y., Kootsookos, A., & Cheng, C. T. (2024). Artificial intelligence use in feedback: A qualitative analysis. *Journal of University Teaching and Learning Practice*, 21(6), 108–125. <https://doi.org/10.53761/40wmcj98>
- [14] Luo, T., Muljana, P. S., Ren, X., & Young, D. (2025). Exploring instructional designers' utilization and perspectives on generative AI tools: A mixed methods study. *Educational Technology Research and Development*, 73(2), 741–766. <https://doi.org/10.1007/s11423-024-10437-y>
- [15] Weng, X., Ye, H., Dai, Y., & Ng, O. L. (2024). Integrating artificial intelligence and computational thinking in educational contexts: A systematic review of instructional design

- and student learning outcomes. *Journal of Educational Computing Research*, 62(6), 1420–1450. <https://doi.org/10.1177/07356331241248686>
- [16] Liu, F., Wang, Y., Feng, Q., Zhu, L., & Li, G. (2024). Optimizing e-learning environments: Leveraging large language models for personalized education pathways. In *2024 5th International Conference on Education, Knowledge and Information Management (ICEKIM 2024)*, 811–817. https://doi.org/10.2991/978-94-6463-502-7_86
- [17] Zhang, X., Zhang, X., & Liu, H. (2024). Reflections on enhancing higher education classroom effectiveness through the introduction of large language models. *Journal of Modern Educational Research*, 3, 19. <https://doi.org/10.53964/jmer.2024019>
- [18] Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. *British Journal of Educational Technology*, 55(5), 1982–2002. <https://doi.org/10.1111/bjet.13505>
- [19] Talgatov, Y., Kassymova, G., & Nurtanto, M. (2024). AI in the classroom: A boon or a threat to pedagogical practices. *Materials of International Scientific-Practical Internet Conference “Challenges of Science*, 61, 128–134. <https://doi.org/10.31643/2024.19>
- [20] Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ..., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- [21] Harry, A., & Sayudin, S. (2023). Role of AI in education. *Interdisciplinary Journal and Humanity (INJURITY)*, 2(3), 260–268. <https://doi.org/10.58631/injury.v2i3.52>
- [22] Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024). AI or human? Evaluating student feedback perceptions in higher education. In R. F. Mello, N. Rummel, I. J. G. Pishtari, & J. A. R. Valiente (Eds.), *Technology enhanced learning for inclusive and equitable quality education* (pp. 284–298). Springer. https://doi.org/10.1007/978-3-031-72315-5_20
- [23] Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, 55(4), 1328–1353. <https://doi.org/10.1111/bjet.13454>
- [24] Liu, J., Jiang, B., & Wei, Y. A. (2025). LLMs as promising personalized teaching assistants: How do they ease teaching work? *ECNU Review of Education*, 8(2), 343–348. <https://doi.org/10.1177/20965311241305138>
- [25] Ippisch, N., Herklotz, M., Haensch, A. C., & Schwemmer, C. (2025). Beyond correctness: Evaluating and improving LLM feedback in statistical education. arXiv. <https://doi.org/10.48550/arXiv.2511.07628>
- [26] Hong, S., Cai, C., Du, S., Feng, H., Liu, S., & Fan, X. (2025). “My Grade is Wrong!”: A contestable AI framework for interactive feedback in evaluating student essays. In *International Conference on Artificial Intelligence in Education*, 27–35. https://doi.org/10.1007/978-3-031-98462-4_4
- [27] Jansen, T., Höft, L., Bahr, L., Fleckenstein, J., Möller, J., Köller, O., & Meyer, J. (2024). Empirische arbeit: Comparing generative AI and expert feedback to students’ writing: Insights from student teachers. *Psychologie in Erziehung und Unterricht*, 71(2), 80–92. <https://doi.org/10.2378/peu2024.art08d>

How to Cite: Oubagine, R., Azzi, I., Laaouina, L., Jeghal, A., & Tairi, H. (2026). Large Language Models (LLMs) for Generating Personalized Educational Feedback: Architecture, Methods, and Perspectives. *Artificial Intelligence and Applications*. <https://doi.org/10.47852/bonviewAIA620210731>