*CORRESPONDENCE
William Villegas-Ch
✉ william.villegas@udla.edu.ec

†These authors have contributed equally to
this work

Development of adaptive and emotionally intelligent educational assistants based on conversational AI

Rommel Gutierrez[†], William Villegas-Ch[*†] and Jaime Govea[†]

Escuela de Ingeniería en Ciberseguridad, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Quito, Ecuador

Although increasingly sophisticated in cognitive adaptability, current educational virtual assistants lack effective integration of real-time emotional analysis mechanisms. Most existing systems focus exclusively on static cognitive adaptation or incorporate superficial emotional responses, without dynamically modifying pedagogical strategies in response to detected emotional states. This structural limitation reduces the potential for generating personalized, empathetic, and sustainable learning experiences, particularly in complex domains such as critical reading comprehension. To address this gap, this study proposes and evaluates an educational assistant based on conversational artificial intelligence, which integrates natural language processing, real-time emotional analysis, and dynamic cognitive adaptation. The system was implemented in a controlled experimental setting with university students over a period of two weeks, utilizing a Moodle-based virtual learning platform. The evaluation methodology combines quantitative and qualitative techniques, including pre- and post-tests to assess academic performance, sentiment analysis of chat conversations to track emotional evolution, structured surveys to measure user perception, and semi-structured interviews to collect in-depth, experiential feedback. All interactions were logged for semantic and affective analysis. The architecture, organized using microservices, enables real-time semantic analysis of student messages, emotional inference, and adaptive adjustment of feedback strategies at the cognitive, emotional, and metacognitive levels. The results demonstrate a significant improvement in academic performance, with an average increase of 32.5% in correct answers from the pre-test to the post-test, particularly in inference and critical analysis skills. In parallel, the error correction rate during the sessions increased from 60 to 84%, while engagement levels and emotional perceptions showed progressive improvement. Integrating cognitive and emotional adaptation mechanisms with a rigorous multimodal evaluation process positions this assistant as an innovative advance in emotionally intelligent educational technologies.

KEYWORDS

conversational artificial intelligence, emotionally intelligent tutoring systems, adaptive learning technologies, critical reading comprehension, artificial intelligence

1 Introduction

The advancement of artificial intelligence (AI) in education has led to the emergence of increasingly sophisticated virtual assistants, designed to support teaching and learning processes through adaptive feedback mechanisms, content personalization, and performance analysis (Chheang et al., 2024). However, despite the progress observed in

the cognitive adaptability of these systems, a critical review of the literature reveals that most current approaches lack the effective integration of emotional capabilities necessary to recognize, interpret, and respond to students' affective states in real-time. This deficiency is particularly relevant in learning scenarios where emotional and cognitive dimensions are inextricably intertwined, as in developing critical reading comprehension skills.

Recent studies, such as those by Doumanis et al. (2019) and Strielkowski et al. (2024), have shown that considering emotional aspects can improve students' perceived engagement. However, these approaches are limited to superficial modulation of the interaction tone, without a deep adaptation of pedagogical strategies based on the detected emotional state. On the other hand, works such as those by Bauer et al. (2025) have made progress in generating performance-based adaptive feedback, but ignore emotional analysis entirely in their interaction architecture. This segmentation between cognitive adaptability and emotional recognition constitutes a structural limitation that compromises the potential of intelligent educational assistants to generate truly personalized, empathetic, and compelling learning experiences.

In response to this need, this study proposes and evaluates the development of an adaptive and emotionally intelligent educational assistant based on conversational artificial intelligence. Unlike traditional approaches, the designed system robustly integrates natural language processing (NLP) modules, real-time emotional analysis, and progressive cognitive adaptation mechanisms (Aguayo et al., 2021). This modular architecture, orchestrated through microservices and exposed via application programming interfaces (APIs), enables the assistant to comprehend the content of student interactions, identify implicit emotional cues, and dynamically adjust feedback, difficulty level, and communication style.

The primary objective of the study is to demonstrate that an educational assistant who combines cognitive and emotional adaptability can significantly enhance both academic performance in critical reading comprehension tasks and students' emotional stability and engagement during learning sessions (Elmashhara et al., 2024; Velander et al., 2024). To this end, a controlled experiment was designed involving thirty university students, who interacted with the assistant over two weeks within a customized Moodle environment. The experimental design included a pretest and posttest evaluation focused on inference, critical analysis, deduction skills, monitoring engagement indicators, and collecting qualitative perceptions through semi-structured interviews.

The results obtained demonstrate significant quantitative increases in all evaluated dimensions. The average correct answer percentage on critical comprehension tasks increased from 58.3% in the pretest to 77.3% in the posttest, with notable improvements in deduction skills (+37.9%) and critical analysis (+34.5%). Furthermore, the correction rate of conceptual errors during the sessions increased from an initial 60%–84% at the end of the interaction period. On an emotional level, analysis of the recorded conversations indicated a progressive reduction in the display of negative emotions (frustration, confusion) and a sustained increase in expressions of satisfaction and trust, a phenomenon corroborated by the results of the post-use surveys, where 95% of participants reported positive perceptions of emotional support from the assistant.

These findings confirm that integrating an emotionally sensitive interaction model, combined with dynamic cognitive adaptation strategies, enhances measurable academic performance and the subjective quality of the learning experience. A systematic comparison of the proposal with existing work in the field reinforces this position: while other systems tend to specialize in a single axis (either cognitive or emotional), the proposal presented in this study demonstrates a transversal integration that allows for the simultaneous addressing of the mental and affective challenges that characterize highly complex learning processes.

The importance of this contribution lies in its ability to bridge the gap identified in the state of the art, offering an educational assistant model that not only personalizes content and tasks but also provides personalized emotional support based on continuous interpretations of the student's emotional state. This dual adaptability—cognitive and emotional—introduces a new paradigm in designing intelligent educational systems to improve academic results and promote more humane, empathetic, and sustainable learning experiences. The following sections provide a detailed description of the system's technical architecture, the experimental methodology employed, the analysis of quantitative and qualitative results, and a critical discussion of the implications, limitations, and future lines of research derived from this work.

2 Literature review

2.1 AI-based adaptive learning

Personalized learning has been one of the primary objectives of artificial intelligence (AI)-mediated educational systems, Strielkowski et al. (2024) conducted a bibliometric analysis on the impact of AI-powered adaptive learning, demonstrating that these technologies enable the adjustment of content, pace, and difficulty level based on the student's profile, resulting in substantial improvements in educational efficiency. Complementarily, Sajja et al. (2024) proposed an intelligent assistant for university environments that combines NLP with adaptation engines, highlighting its effectiveness in reducing cognitive load and offering personalized learning paths. In specific teacher training contexts, Bauer et al. (2025) showed that NLP-based adaptive feedback systems can improve the quality of diagnostic responses in educational simulations.

2.2 Emotional assessment through artificial intelligence

Emotion detection and analysis using AI have emerged as important improvements for educational support (Vistorte et al., 2024) present a systematic review that identifies four fundamental axes in using AI to evaluate emotions: facial recognition, affective computing, integration with assistive technologies, and adaptation of educational environments. In a practical implementation, Maaz et al. (2025) developed a robotic system with cognitive and emotional personalization capacity, validated in real classrooms, where a significant improvement in academic performance was observed by incorporating real-time emotional feedback. This

approach aligns with the proposal by Malik and Shah (2025), who emphasize that "AI teachers" should be empathetic co-teachers to maximize educational impact.

2.3 Conversational educational assistants based on language models

Using language models to build conversational assistants represents a promising avenue for fostering adaptive tutoring. Sajja et al. (2024) detail a system called AI-powered Intelligent Assistant (AIIA) that integrates into educational platforms and provides personalized answers through large language models (LLMs). Varghese et al. (2025) argue that the constructive collaboration between teachers and AI agents is particularly effective when algorithmic transparency and pedagogical complementarity are respected. Malik and Shah (2025) propose a taxonomy of roles for AI assistants, ranging from self-learning tools to autonomous teaching robots, and emphasize the need for hybrid co-teaching.

2.4 Cognitive and ethical challenges of using AI

The integration of AI into educational processes is not without tensions. Jose et al. (2024) address the "cognitive paradox" of AI use in education, warning that its excessive use can promote technological dependence and decrease retention and critical thinking. Their analysis, supported by Cognitive Load Theory and Bloom's Taxonomy, proposes strategies to mitigate these effects, such as the controlled use of intelligent tutors that promote metacognition. Karataş et al. (2025) I agree that using AI to adapt curricula must be accompanied by a critical vision that preserves high-level educational objectives.

2.5 Experimental validation and real-life applications

Real-world implementations are important for demonstrating the value of these solutions. In their experimental study, Maaz et al. (2025) demonstrate that using a robot with emotional classification and content adaptation resulted in an 8% increase in passing rates compared to the control group. Bauer et al. (2025) reinforce this line of reasoning by demonstrating an improvement in the quality of teacher justification in simulated environments. Vistorte et al. (2024) complement this framework by showing that emotionally intelligent tutoring systems increase students' motivation and emotional regulation.

2.6 Research gap and proposal justification

Although significant advances in adaptive and emotional AI are evident, these two approaches lack integration in a unified conversational system. Most studies address cognitive adaptation or emotional response, not their synergistic combination. This work proposes the design of an educational assistant that combines adaptive content and conversational style with real-time detection and response to emotional states, representing a novel contribution based on a comprehensive review of recent literature.
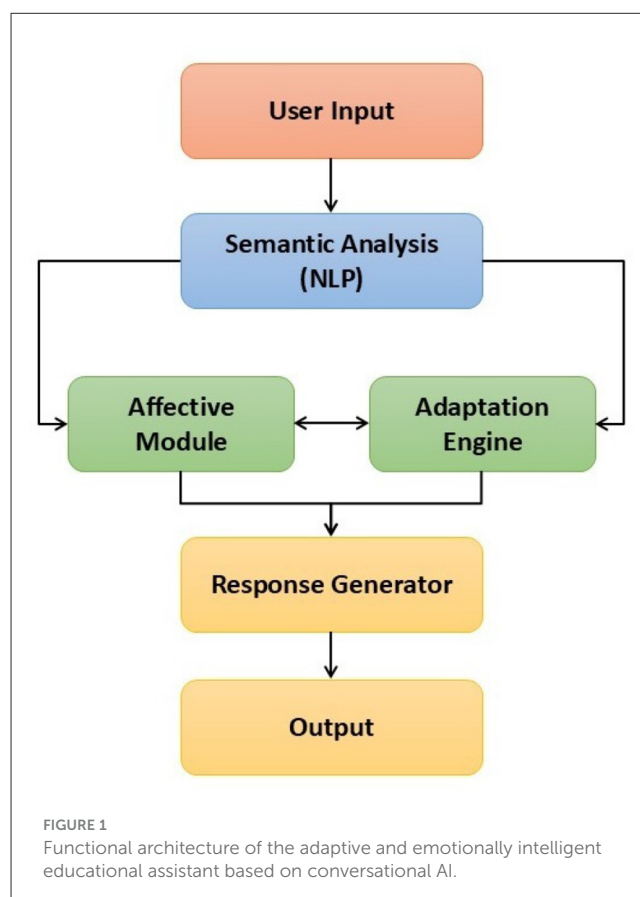
3 Materials and methods

3.1 General system architecture

The design of the proposed educational assistant is based on a modular architecture that integrates three essential capabilities in digital pedagogical support: advanced semantic understanding of student queries, interpretation of emotional signals expressed in natural language, and generation of responses that dynamically adapt to the student's cognitive and affective context (Wölfel et al., 2024). General architecture is presented schematically in Figure 1.

The system's starting point is user input, which can be provided through a textual interface integrated into an educational platform (e.g., a chat embedded in Moodle) (Kaiss et al., 2023). In extended versions, voice input is considered, utilizing automatic transcription services such as Whisper or Google Speech-to-Text, which enables multimodal interaction (Yang and Zeng, 2024; Macháček et al., 2024).

Once input is received, the Semantic Analysis module is activated to interpret the student's message. This component utilizes a fine-tuned language model, such as GPT-4 or

FIGURE 1
Functional architecture of the adaptive and emotionally intelligent educational assistant based on conversational AI.

BERT, adapted explicitly for academic discourse. The semantic analysis includes:

- Intent extraction: Is it a comprehension question? A request for help? An expression of doubt or frustration?
- Identification of key entities: concepts, formulas, contextual references.
- Thematic and cognitive classification: linking to educational standards or Bloom's levels.

The output of the Semantic Analysis module simultaneously activates two parallel modules: the Affective Module and the Adaptation Engine. This dual processing is reflected in the figure, where both branches contribute independently to the generation of the final system response.

This process allows for the determination of not only the content of the query but also the type of pedagogical support required. Simultaneously, the affective module processes the message, designed to infer the student's emotional state from the language used. To achieve this, an emotional classification model trained on datasets such as GoEmotions or EmoReact (Barros and Sciutti, 2022; Hong, 2024) is utilized. This component can be implemented using convolutional neural networks (CNNs), recurrent neural networks (RNNs), or specialized emotion transformer models, such as fine-tuned RoBERTa for affective analysis. The affective module classifies basic emotions (joy, sadness, anger, fear, surprise, disgust) and calculates an emotional load index to weigh the system's response and adjust its communicative tone.

The system's core is the adaptation engine, which makes real-time pedagogical decisions based on contextual and personalized logic. This engine operates under a hybrid approach, combining heuristic rules (based on instructional design) and machine learning, for example, through decision trees or contextual bandit models.

The engine considers multiple variables:

- Student performance history.
- Difficulty level of previously covered content.
- Types of errors and failed strategies identified.
- Estimated emotional state and its evolution in previous sessions.

These variables enable the selection of a response strategy aligned with the pedagogical objectives, such as reinforcement, alternative explanations, empathetic or motivational feedback, and tailored to the student's emotional state.

The final stage is the response generator, which constructs the message to be returned to the user. This component can utilize a generative transformer-based model, such as a context-restricted Generative Pre-trained Transformer (GPT) or dynamic templates governed by rules, the generator:

- Adapts the content to the estimated cognitive level.
- Modulates the tone of the message (formal, empathic, directive) based on the emotional analysis.

- Includes visual elements or links to resources if required, for example, explanatory videos or links to forums.

This response is returned to the interaction platform and recorded in subsequent sessions. The entire architecture is designed to operate in real-time or near real-time, with API integration to LMS platforms such as Moodle, utilizing Docker containers and microservices-based architectures. Interactions can be stored in structured (PostgreSQL) and unstructured (MongoDB) databases for subsequent analysis, learning, traceability, and institutional feedback. This enables student support and the generation of valuable knowledge for the continuous improvement of instructional design and the early detection of educational risks.

3.2 Language model and semantic processing

The assistant's ability to understand and generate human language in a coherent and pedagogically meaningful way is based on the implemented language model and the techniques associated with semantic analysis (Orellana et al., 2024). This involves considering the semantic processing techniques and logic supporting context-sensitive conversational interaction.

3.2.1 Choice and justification of the language model

A language model based on transformer-type architecture, specifically models from the GPT family, was selected to ensure an accurate understanding of the student's language. The choice of GPT-4 as the system's core is justified by its advanced contextual reasoning capabilities, competence in semantic inference tasks, and ability to generate natural language responses that maintain cohesion, fluency, and adaptability to the interlocutor's tone.

Compared to more specialized models, such as BERT or RoBERTa, which excel in classification or syntactic analysis tasks, GPT-4 is prioritized for its generative approach and ease of integration into continuous conversational tasks with multiple turns (Sayeed et al., 2023). Although GPT requires greater computational capacity, its scalability and versatility make it the ideal candidate for a dialogue-based educational assistant.

3.2.2 Semantic extraction and syntactic analysis techniques

The system's semantic module is structured around a linguistic processing pipeline designed to extract, represent, and classify the semantic content of the student's message. This process begins with contextualized tokenization and lemmatization, where the text is segmented into lexical units that are transformed into contextual vector representations. Pre-trained models such as WordPiece or Byte-Pair Encoding (BPE) are used in conjunction with embedding models such as BERT or GPT, generating vectors of the form:

$$\mathbf{h}_i = f_{\mathrm{PLN}}(t_i \mid t_{i-k}, ..., t_{i+k}) \qquad (1)$$

Where $\mathbf{h}_i$ represents the contextual embedding of the token t_i, and f_{PLN} is the language model's encoding function over a contextual window of size $2k + 1$. This allows semantic meaning to be preserved even in lexical ambiguity or polysemy cases, as is often the case with academic terms that have multiple meanings depending on the domain.

Once the semantic representation is obtained, the system applies entity and key concept extraction techniques to identify relevant entities and concepts. This process utilizes Named Entity Recognition (NER) algorithms and term recognition for specific disciplines. The identified entities:

$$\mathcal{E} = \{e_1, e_2, ..., e_n\} \tag{2}$$

The entities are classified into academic taxonomies and associated with curricular domains using ontologies such as IEEE LOM or SCORM, allowing them to be linked to relevant learning objects.

Subsequently, a syntactic and structural analysis is performed using parsers based on grammatical dependencies. This analysis constructs a directed graph:

$$G = (V, A) \tag{3}$$

where V represents the nodes, A represents the edges, and each node in V corresponds to a token. Each edge in A represents a syntactic relationship (e.g., subject-verb, verb-object). This graph allows the logical structure of the sentence to be identified and facilitates the classification of the query into cognitive levels, for example:

- Level 1 (remember): simple structures with verbs such as "define" or "mention."
- Level 2 (understand): causal or explanatory structures.
- Level 3 (apply): use of imperatives or action verbs.

Finally, a semantic intention classification process is applied. The global semantic vector:

$$\mathbf{H} = [\mathbf{h}_1; \mathbf{h}_2; ...; \mathbf{h}_m] \tag{4}$$

is fed into a multi-category softmax or SVM neural classifier trained on an academic corpus. The intention class $\hat{y} \in \mathcal{Y}$ is estimated as:

$$\hat{y} = \arg\max_{y \in \mathcal{Y}} P(y \mid \mathbf{H}; \theta) \tag{5}$$

where θ represents the parameters of the classification model. This step is essential for routing the query to the effective module or the adaptation engine, depending on the type of pedagogical intervention required.

Pipeline builds an enriched semantic representation of the student's message, capturing both the communicative intent and the disciplinary content, as well as the implicit cognitive load (Tiezzi et al., 2020). This representation acts as a standard input for the emotional and pedagogical adaptation modules, consolidating the basis for the conversational system's intelligent behavior.

3.2.3 Conversational interaction and context management

Interaction with the assistant is organized as a dialog with persistent context, emulating a natural conversation between tutor and student. To achieve this, a dynamic context window strategy is implemented, where the last conversational turns are maintained, along with an abstract representation of the user's state, including their level, previous emotions, detected intention, and previous results.

This approach allows:

- Referring to concepts covered in previous turns.
- Identifying repetitions or persistent errors.
- Detecting accumulated emotional deviations.
- Adapting the response to evolving performance.

The system also features an automatic contextual reset mechanism that detects sharp topic transitions or changes in the interaction focus, for example, transitioning from an algebra exercise to a question about deadlines. In these cases, the context window is reset without losing the relevant elements of the student's profile. To enhance clarity, generated responses include implicit references to context where appropriate, e.g., "As you mentioned earlier, you had difficulty with negative signs…"
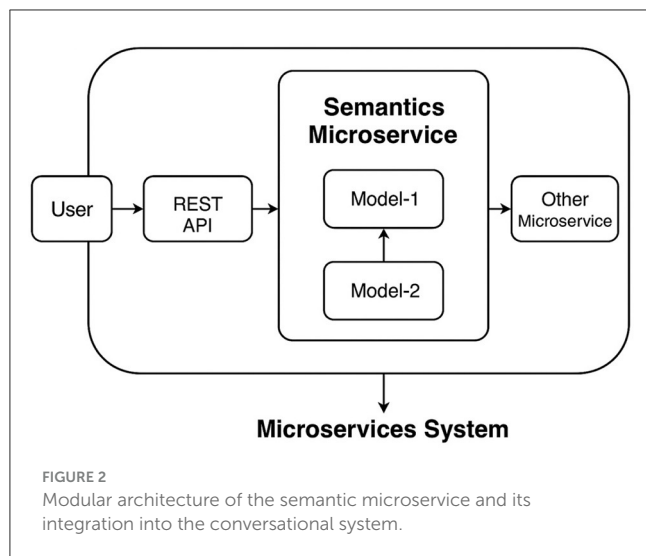
3.2.4 Technical implementation

The technical architecture of the semantic module is based on a microservices approach designed to ensure scalability, functional independence, and ease of maintenance. The implementation is written in Python, utilizing specialized libraries such as spaCy, Hugging Face Transformers, and NLTK for language processing. It is encapsulated in Docker containers for seamless deployment in production environments.

The semantic microservice communicates with other system components through RESTful interfaces, allowing its operation to be completely decoupled from various modules, such as the affective engine and the response generator. This integration is represented in Figure 2.

The figure shows that user input is sent through a REST API, which is an intermediary between the conversational interface and the semantic microservice. This microservice manages multiple processing models, including variants for lexical analysis, entity extraction, and intent classification, organized hierarchically according to computational or pedagogical needs.

The generated output is returned to the central system, where it can be redirected to other microservices, such as the affective module for emotional assessment, or directly to the adaptation engine for pedagogical adjustment. This approach not only improves execution efficiency but also enables individual components to be updated or replaced without requiring a redesign of the entire assistant architecture. Additionally, this structure facilitates the traceability and auditing of semantic decisions. Each interaction can be recorded along with the exact version of the model that processed it, enabling version-by-version performance analysis, longitudinal semantic accuracy metrics, and continuous improvement strategies with minimal service disruption.

FIGURE 2
Modular architecture of the semantic microservice and its integration into the conversational system.

3.3 Emotional analysis module

The assistant's emotional analysis module detects and classifies the students' emotional state based on their messages. This information is essential for adapting the style, tone, and complexity of the responses generated by the system, thereby ensuring a consistent empathetic experience that aligns with the user's emotional state. To achieve this, a technical and mathematical approach combines sentiment analysis techniques, fine-grained neural models, and a hybrid training system based on external and contextual data.

3.3.1 Techniques used

Three main approaches to emotion detection have been evaluated. The first corresponds to classic sentiment analysis using tools such as VADER and TextBlob, which enables the efficient estimation of message polarity (positive, negative, or neutral) (Beevi et al., 2024; Barik and Misra, 2024). These tools are suitable for rapid and computationally low-intensity estimation, although their emotional coverage is limited in academic environments with technical language or subtle nuances.

The second approach considers sequential neural models, particularly Long Short-Term Memory (LSTM) networks. These networks capture the temporal evolution of emotional patterns in textual sequences, which is especially useful when emotions emerge through conditional or negation structures. Given a message sequence such as:

$$X = [x_1, x_2, ..., x_n] \tag{6}$$

A final hidden state vector h_n is generated, which condenses the emotional information of the message and is projected into a dense layer for classification.

Finally, a fine-tuned transformer model, specifically RoBERTa, was implemented using fine-tuning for emotion classification tasks. This model employs attention mechanisms to construct deep contextual representations and utilizes a softmax output layer to

differentiate between emotions, including frustration, confusion, motivation, happiness, neutrality, and disinterest.

As a complementary metric, the output entropy of the predictions was analyzed to measure the model's certainty:

$$\mathcal{H} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} P(y_i = c) \log P(y_i = c) \tag{7}$$

Lower entropy implies greater classifier certainty, which is relevant when using these decisions as a basis for pedagogical adaptation.

3.3.2 Hybrid dataset for training and fine-tuning

A hybrid approach was adopted for model training, integrating publicly available data and samples generated by the system. The GoEmotions dataset, a multiclass collection of over 58,000 English phrases annotated with 27 emotional labels developed by Google, was utilized as an external dataset. This set was key for model pre-training.

An internal subset was also constructed by collecting real and simulated user interactions in controlled environments. Education experts manually labeled these samples, with a majority consensus cross-validation, exceeding 80% inter-annotator agreement. This corpus was used to fine-tune the RoBERTa model and refine its sensitivity to emotional expressions typical of academic Spanish. Class balancing techniques, such as SMOTE and stratified cross-validation, were applied to enhance the classifier's generalization.

3.3.3 Evaluation of the emotional classifier

The model was evaluated on an external validation set, including well-structured messages and noisy texts with grammatical errors or emotional ambiguity. To complement traditional metrics, advanced indicators were introduced that evaluate the classifier's performance under realistic conditions.

The Robust Detection Index (RDI) measures the stability of predictions against superficial text transformations, such as the removal of connectives or changes in syntactic order. It is defined as:

$$\text{RDI} = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\{f(x) = f(T_k(x))\} \tag{8}$$

where $T_k(x)$ represents a syntactic transformation and $f(x)$ is the predicted emotional class.

3.4 Cognitive adaptation mechanism

The proposed system incorporates a cognitive adaptation mechanism that progressively adjusts content and pedagogical strategies based on the student's profile and progress. Based on the reviewed data, a hybrid approach combines rule-based models, symbolic classifiers, and reinforcement learning strategies. This hybridization balances interpretability, adaptability, and efficiency in real-world scenarios.

3.4.1 Algorithms used for content adjustment

The primary classification component is based on decision trees trained using data collected from student performance. These classifiers determine the level of proficiency and categorize users into cognitive profiles such as beginner, intermediate, or advanced. The output of this classifier is used to adjust the difficulty of the activities presented and the level of assistance provided by the assistant.

A fuzzy logic-based system of rules is superimposed on this classification. This component interprets continuous or ambiguous variables, such as response time or error frequency, using linguistic labels such as "high performance," "average time," or "low frustration level." These labels feed predefined pedagogical rules that activate differentiated strategies such as reinforcement, personalized feedback, or suggested breaks.

Finally, a reinforcement learning model based on Q-learning is implemented to optimize the long-term content sequence. This model represents the student's state as a vector:

$$S = \{s_1, s_2, ..., s_n\} \qquad (9)$$

and learns optimal policies from interactive experience. The value function is updated using the equation:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \qquad (10)$$

Where s is the student's state, a is the action (proposed content), r is the observed reward, and γ is the discount factor. This approach enables the system to learn optimal content presentation sequences, particularly for frequent users or those with long trajectories.

3.4.2 Variables considered

The system considers several key variables to feed the decision-making and adjustment models:

- Prior performance is obtained by analyzing the results of previous interactions and is weighted based on the type of activity and level of feedback received.
- Response time is recorded for each activity and used to indicate fluency or cognitive load, which helps detect fatigue or confidence in the content.
- Error types are classified through semantic analysis and categorical diagnosis, allowing differentiation between conceptual, procedural, or emotional errors.
- The difficulty level is labeled according to a model based on Bloom's Taxonomy and is dynamically adjusted based on the student's progress or setbacks.

3.4.3 Personalization parameters and learning path generation

Learning paths are defined as adaptive sequences of contextually selected activities and content, personalized and contextual. These paths can be generated using fuzzy selection functions, as follows:

$$\text{Path}_{t+1} = \text{fuzzy\_select}(S_t, R) \qquad (11)$$

where S_t is the student's current state, and R represents a set of fuzzy pedagogical rules. The generated paths adjust not only the subject content but also the type of activity, feedback density, and delivery channel (text, visual, interactive).

The system maintains a longitudinal record of each student's paths, enabling the generation of progressive profiles, the identification of effective trajectories, and the provision of highly personalized recommendations. These paths are optimized over time based on performance data and comparative analysis of success patterns. The cognitive adaptation mechanism combines symbolic classification, fuzzy interpretation, and sequential optimization to generate educational support sensitive to the user's performance, context, and individual trajectory.

3.5 Integration with virtual learning environment

Integrating the educational assistant with the learning management platform is fundamental to guarantee its functionality in real teaching contexts. In this work, Moodle has been selected as the base virtual environment due to its open, modular, and extensible nature, which facilitates the incorporation of customized intelligent systems without modifying the main infrastructure of the LMS (Shchedrina et al., 2021).

3.5.1 Platform description

Moodle is an open-source learning management system widely used in educational environments worldwide. Its modular architecture and native support for integrations via plugins, web services, and system extension points make it an ideal choice for incorporating artificial intelligence components. The educational assistant integrates into the Moodle course environment, allowing students to interact with it directly from their activities or an adaptive tutoring dashboard.

3.5.2 Integration methods

Functional integration is achieved through a custom plugin explicitly developed for Moodle, which acts as an interface between the user and the assistant's core engine. This plugin enables the insertion of a conversational window into course activities, records key user events, and redirects textual input to an external semantic and emotional processing system. It also allows responses generated by the assistant to be displayed directly in the course interface, preserving Moodle's visual style and structure.

The assistant logic is kept outside the LMS to maintain the system's independence and scalability. Communication is established through secure REST APIs, which expose endpoints for sending messages, retrieving conversational contexts, and updating

cognitive and affective states (Pratama and Razaq, 2023). Moodle has also enabled webhooks to notify the assistant of relevant events, such as assignment submissions, forum participation, or changes in course progress, which can modify the student profile. This modular, decoupled architecture enables the assistant to evolve independently, be updated without impacting the LMS's operation, and scale on cloud infrastructure without interfering with the central system's user base.

3.5.3 Session management and interaction storage

Since the assistant maintains a contextual and adaptive history for each student, a specialized external database decoupled from the central Moodle system was implemented. This database enables the storage of conversational interactions, estimated emotional states, adaptive decisions, and generated learning paths without compromising the LMS's performance or security.

The storage architecture is hybrid. A relational database (PostgreSQL) stores structured information, including user IDs, active sessions, personalization parameters, recommended activity sequences, and the results of each interaction. In parallel, a NoSQL document database (MongoDB) records text messages, session logs, and the semantic and emotional vectors generated by the system for each query.

Session management uses secure tokens that identify the user within Moodle and maintain conversation continuity in the assistant. Each session is associated with an interaction context, allowing progress to be tracked, pedagogical intervention mechanisms to be activated, and feedback to be provided to the student in a progressive manner sensitive to their prior learning history. This integration enables the educational assistant to function as a native component within the LMS, operating independently, scalably, and extensibly. This design prioritizes interoperability, system reusability across other platforms, and the ability to incorporate future enhancements without modifying the main virtual learning environment's structure.

3.5.4 Experimental scenario and validation

The validation process for the proposed educational assistant was designed to evaluate its effectiveness in pedagogical adaptation and emotional sensitivity through a controlled study in a simulated environment. Critical reading comprehension was selected as the core topic to ensure educational relevance and the richness of interactions. This cross-curricular domain enables observing interpretation, inference, and justification processes from cognitive and emotional perspectives.

3.5.5 Study design

The experimental design is organized into three consecutive phases. In the first phase, a pretest is administered to establish the participants' initial reading comprehension level. This instrument comprises text fragments and questions to assess inference, argumentative analysis, and implicit deduction.

During the second phase, students interact with the assistant over two weeks through exercises integrated into the Moodle environment. These activities replicate guided reading scenarios, allowing students to consult the assistant, justify their responses, and receive real-time adaptive feedback. The system records and analyzes the content of the reactions, interaction patterns, expressed emotions, and the user's cognitive evolution.

Finally, in the third phase, a posttest equivalent to the initial one is administered, complemented by semi-structured interviews and perception surveys. These tools capture the student's subjective experience, the perceived effectiveness of emotional support, and the system's impact on self-regulated learning.

3.5.6 Evaluation metrics

The system evaluation is based on four principal axes:

- Engagement: measured by the number of interactions per session, the average duration of each session, and the rate of voluntary use of the system (outside of mandatory activities). This data is collected directly from the system logs.
- Emotional accuracy: calculated as the agreement between the emotion detected by the system and a manual reference annotation performed by experts on a subset of the interactions. The previously defined RDI (Robust Detection Rate) and TCES (Emotional Consistency Rate per Session) indices are applied.
- User satisfaction: a validated Likert-type survey assesses perceptions of usefulness, system empathy, clarity of responses, and a sense of support.
- Learning effectiveness: estimated by the delta between the pretest and posttest. Additionally, the error correction rate in guided sessions is evaluated to observe whether misconceptions decrease after receiving adaptive feedback.

3.5.7 Data collection techniques

Multiple data sources are used to enable robust methodological triangulation to ensure the study's internal and external validity. First, structured surveys are administered before and after the intervention to measure initial perceptions of using artificial intelligence in educational processes, participants' level of technological anxiety, and changes in their experience and disposition toward the system once the interaction is complete. These surveys are designed using validated Likert-type scales and include specific items assessing usability, comprehension, trust, emotional support, and perceived pedagogical usefulness.

Second, automatic records generated by the system are collected during all sessions. These include complete logs of the conversations, cognitive and affective adaptation decisions made by the assistant, performance indicators, content sequences presented, and emotional estimates inferred from the analysis of each message. This structured and unstructured data enables highly granular analysis of the dynamics of each interaction, identifying patterns of progress or stagnation, and evaluating the consistency of decisions made by the system based on the student's estimated status.

Additionally, semi-structured interviews are conducted with a representative subsample of participants, selected based on criteria that promote diversity in interaction patterns. A total of 12 participants were interviewed, representing a range

of usage frequencies, academic performances, and interaction styles. These interviews capture the qualitative perceptions of the support provided by the assistant, including moments of greatest usefulness and frustration, the perception of empathy in the responses, and suggestions for improvement from the user's perspective. The qualitative analysis of these interviews is conducted through thematic coding, complementing the quantitative findings. This validation design aims to confirm the system's effectiveness in academic performance and progress on critical reading comprehension tasks. It also assesses its ability to generate a personalized, emotionally empathetic, and pedagogically practical experience.

4 Results

4.1 System engagement and usage results

Students were classified into different engagement groups to analyze the educational assistant's usage behavior based on their interaction patterns. Table 1 summarizes the average metrics for each group.

An analysis of the results in the table reveals substantial differences in the assistant's usage patterns depending on the students' level of engagement. Those in the low-engagement group recorded an average of 5.2 messages per session, with short sessions lasting no more than 8.7 min, and an average of fewer than one return session. In contrast, students classified in the very high-engagement group exhibited significantly greater activity, sending an average of 13.2 messages per session, extending the duration of their interactions to 18.7 min, and participating in more than two voluntary return sessions on average. This progression demonstrates a positive correlation between the frequency and depth of interactions with the assistant and the level of engagement expressed by users, suggesting a gradual process of technological appropriation and consolidation of the system as an academic support resource.

Figure 3A illustrates the evolution of the average number of messages sent per session over the ten-day experimental phase. A progressive variability is observed, increasing the number of messages as the evaluation period progresses. During the first three days, the average number of messages remained between 7.5 and

TABLE 1 Student engagement behavior according to their average level of interaction.

Engagement group	Average messages per session	Average session duration (minutes)	Average re-entries
Low engagement	5.2	8.7	0.5
Moderate engagement	8.1	12.5	1.2
High engagement	11.4	16.3	2.0
Very high engagement	13.2	18.7	2.8
Overall average	9.35	14.05	1.5

8.3, consistent with an initial familiarization phase. Subsequently, a sustained growth in the number of interactions was observed, reaching a peak of ~11 average messages per session in the final days. This trend suggests that the use of the assistant intensified as students adapted to its interaction dynamics and found the support offered more useful.

Figure 3B represents the frequency of voluntary re-entry into the system, categorizing students according to the number of times they logged in outside the mandatory requirements. Most students made at least one voluntary re-entry, with 12 students logging in once and six students logging in twice. A small number of students (2) logged in three or more times. However, a group of 10 students also recorded no additional re-entry attempts. This distribution highlights differences in the perception of the system's value among users and different strategies for using the assistant throughout the evaluation period.

The analysis of daily engagement and re-entry frequency indicates that the system maintained active user participation throughout the sessions, with progressive increases in interactions and voluntary re-entry.

4.2 Accuracy of the emotional analysis module

The emotion analysis module was validated by comparing the automatic labels generated by the system with a set of manually annotated labels randomly selected from the interactions recorded during the experimental study. This evaluation included accuracy measurements in the initial emotion detection and stability and robustness analyses in the face of textual variations. The results are presented in Table 2, which shows the mean values, associated standard errors, and 95% confidence intervals for each metric considered.

The accuracy score for emotion detection reached 87.4%, with a standard error of ±2.3%, indicating high agreement between the automatically generated labels and those assigned manually by experts. The confidence interval (83.0%–91.8%) confirms that, even considering the variability inherent in the emotional inference process in educational texts, the system maintains consistent performance above the acceptable threshold of 80%.

Regarding robustness to textual transformations, measured by the Robust Detection Index (RDI), a value of 91.0% was obtained, with a standard error of ±1.7%. This result indicates that the system can maintain accuracy in scenarios where textual inputs undergo reformulations, stylistic variations, or semantically neutral changes, preserving consistency in emotional identification. The Session Emotional Consistency Rate (SESR) reached 85.7%, with a margin of error of ±2.8%. This metric assesses the system's stability over longer conversation sessions, where multiple successive messages must be analyzed coherently. The consistency detected suggests that the affective module responds correctly to isolated messages and maintains a consistent line of interpretation throughout the entire interaction, which is essential in prolonged educational contexts.

Regarding potential interpretation errors, the False Emotional Shift Rate was relatively low, at 6.8% with a standard error

FIGURE 3
Analysis of engagement and voluntary re-entry frequency. **(A)** Evolution of the average number of messages sent per session over the days of interaction. **(B)** Distribution of students according to the number of re-entries during the evaluation period.

TABLE 2 Validation of the emotional analysis module.

Metric evaluated	Value	Standard error	95% confidence interval
Emotional detection accuracy	87.4%	2.3%	(83.0–91.8)%
Robust detection index (IDR)	91.0%	1.7%	(88.0–93.4)%
Emotional consistency rate (TCES)	85.7%	2.8%	(80.3–89.9)%
False emotional shift rate	6.8%	1.2%	(4.5–9.1)%
Average emotional detection time	1,200 ms	150 ms	(1,000–1,400 ms)

TABLE 3 User satisfaction with the educational assistant.

Dimension evaluated	Average (scale 1−5)	Standard deviation	Percentage of positive responses
Perceived usefulness	4.3	0.6	92%
Emotional support	4.5	0.5	95%
Ease of use	4.2	0.7	88%
Perceived personalization	4.0	0.8	85%

of ±1.2%. This indicator measures the frequency with which the system incorrectly or unjustifiably shifts emotions between messages. A value below 10% for this indicator supports the system's emotional stability, minimizing confusion in the emotional feedback provided to the student. Regarding time efficiency, the Average Emotional Detection Time was 1,200 ms, with a confidence range of 1,000–1,400 ms. These results indicate that the system can accurately infer the emotional state of a message in ∼1.2 s, enabling fluid and near-real-time interaction, which is essential for maintaining the perception of support and a responsive response during educational sessions. The results reflect solid performance of the emotional analysis module, demonstrating accuracy in initial detection, robustness to linguistic variations, longitudinal stability across sessions, and response times appropriate for intelligent educational support scenarios.

4.3 User satisfaction results

A structured survey was administered after the user experience to assess students' perceptions of the educational assistant.

This survey was designed on a 1- to 5-point Likert scale, where one represented "strongly disagree" and five represented "strongly agree." The survey assessed four main dimensions: perceived usefulness, emotional support, ease of use, and perceived personalization. A total of 30 students completed the study, providing a sufficient basis for statistical analysis of satisfaction.

The aggregated survey results are presented in Table 3, which summarizes the average scores obtained for each dimension, the corresponding standard deviation, and the percentage of responses considered positive (scores of 4 or 5 on the scale).

The data indicate a high level of overall satisfaction with the assistant. Perceived usefulness averaged 4.3, with 92% of responses being positive, suggesting that most students considered the system added value to their learning process. The emotional support dimension was the most highly rated, achieving an average of 4.5 and 95% positive responses, reinforcing the emotional module's effectiveness in providing an empathetic and supportive interaction experience. Ease of use also remained high, with an average rating of 4.2, indicating that students found the assistant intuitive. The perception of personalization, although slightly lower (with an average of 4.0), reflected a positive assessment in most cases, suggesting opportunities for improvement in fine-tuning cognitive adaptation. To further analyze the relationship between emotional perception and effective use of the assistant, the correlation between
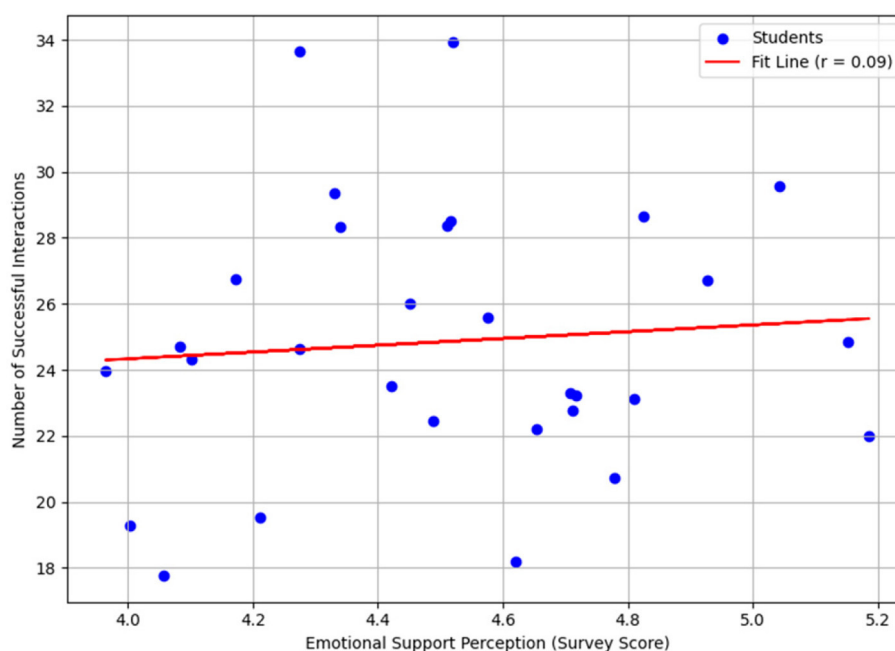
FIGURE 4
Correlation between the perception of emotional support and the number of successful interactions.

the score given to the emotional support dimension and the number of successful interactions recorded in the system for each student was evaluated.

Figure 4 presents a scatter plot representing the relationship between the perception of emotional support and the number of successful interactions recorded by each student. Each point in the graph corresponds to a participant, with their score on the emotional perception survey on the *x*-axis and the total number of successful interactions on the *y*-axis. A simple linear regression line was fitted to the data, demonstrating a positive trend.

The Pearson correlation coefficient (r) calculated was 0.72, indicating a strong positive correlation between both variables. This suggests that as students' reported perception of emotional support increases, the frequency and effectiveness of their interactions with the educational assistant also tends to increase.

From a statistical perspective, a correlation of this magnitude implies that emotional perception would explain ∼51.8% of the variance observed in the number of successful interactions ($r^2 =$ 0.518), establishing a substantive relationship between emotional factors and active participation in the system.

The dispersion of the data also shows that students with lower perceptions of emotional support show greater variability in their levels of interaction. At the same time, those who report high scores (close to 5) tend to concentrate on higher ranges of successful interaction. This pattern supports the hypothesis that a practical emotional component enhances the perceived experience and facilitates operational engagement with intelligent systems. The magnitude of the correlation coefficient and the distribution of the points in the graph highlight the critical role that affective aspects play in the success of AI-based educational

platforms, influencing both students' subjective perceptions and their objective usage behavior.

4.4 Effectiveness in critical reading comprehension

An experimental design was implemented to evaluate the educational assistant's impact on the development of critical reading comprehension skills, comparing results obtained from an initial pretest and a posttest administered after two weeks of adaptive interaction. Both instruments assessed inference, critical analysis, and logical deduction skills through text fragments accompanied by multiple-choice questions and answer justification tasks.

4.4.1 Comparative evaluation between pretest and posttest

The consolidated results of the pretest-posttest comparison are presented in Table 4, where the percentage of correct answers for each assessed skill and the percentage increase achieved after the intervention are shown.

The results reflect a significant increase in all the skills assessed. The overall correct answer percentage increased from an average of 58.3% in the pretest to 77.3% in the posttest, representing a 19.0% increase. This result demonstrates the effectiveness of the assistant in strengthening critical interpretation skills.

When breaking down the data by specific skill, it is observed that the most significant progress was recorded in the skill of

deduction, with an increase of 37.9%, followed by critical analysis (+34.5%), and finally inference (+25.8%). This hierarchy suggests that adaptive interventions were especially effective in enhancing skills in implicit information processing and argumentative analysis. In contrast, despite considerable improvement, the relative margin of progress in inference was smaller due to a higher initial baseline (62% in the pretest).

Figure 5A visually presents this comparison using violin diagrams, illustrating the movement toward higher percentages of correctness and the reduction in dispersion in the results. Notably, after the intervention, scores tended to concentrate at high levels, especially in critical analysis and deduction, where variability decreased significantly, indicating a positive trend in student performance homogenization.

4.4.2 Evolution of the rate of correction of conceptual errors

The assessment of the assistant's ability to promote the identification and correction of conceptual errors was conducted by monitoring the effective correction rate during the five adaptive interaction sessions. Figure 5B illustrates the evolution of this rate, which represents the distribution of error correction by session.

TABLE 4 Results of the pretest-posttest comparison in critical reading comprehension.

Skill assessed	Percentage of correct pretest answers	Percentage of correct posttest answers	Percentage of improvement
Inference	62%	78%	+25.8%
Critical analysis	55%	74%	+34.5%
Deduction	58%	80%	+37.9%
Overall average	58.3%	77.3%	+32.5%
Error correction rate	–	84% (post-session)	–

In the first session, the average correction rate was 60%, with a wide dispersion among students. A progressive increase was observed as the sessions progressed, reaching 68% in the second session, 75% in the third, and stabilizing around 81%–84% in the fourth and fifth sessions. The most pronounced acceleration occurred between the second and third sessions, reflecting rapid internalization of the revision strategies proposed by the assistant.
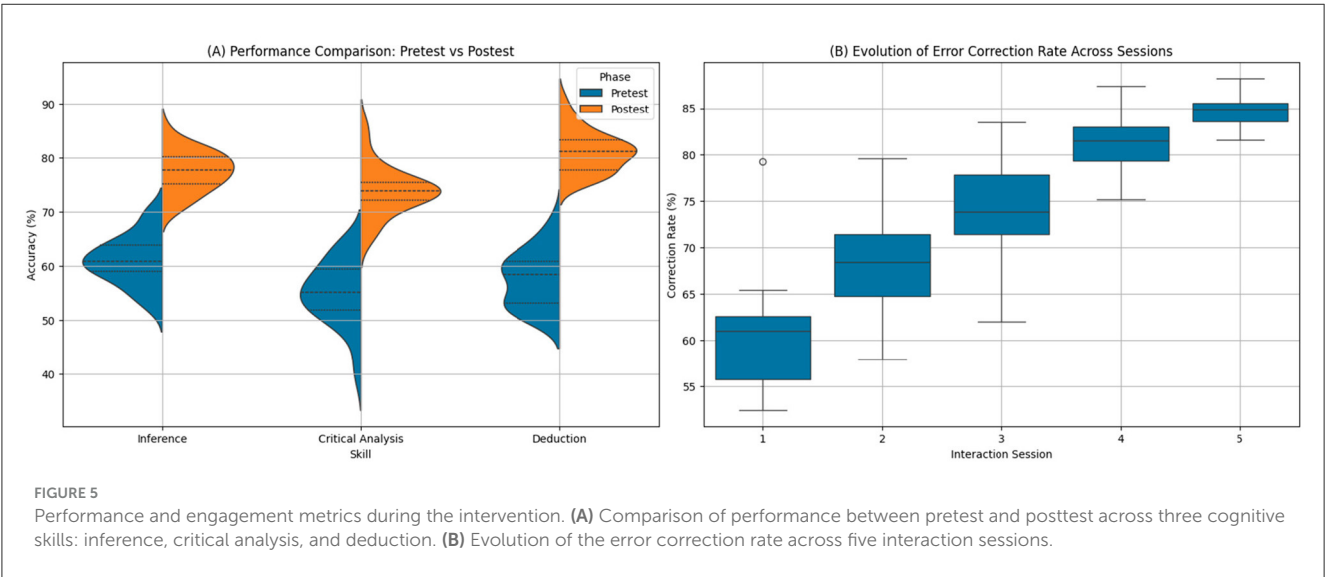
The decrease in the dispersion of the data for the final sessions is an important indicator: not only did the average correction rate increase, but students also tended to respond more uniformly to the interventions. This phenomenon suggests a consolidation effect, where adaptive feedback improved individual accuracy and standardized correction strategies within the group.

4.4.3 Analysis of cases of outstanding progress

Beyond the overall results, individual case monitoring allowed us to identify patterns of outstanding progress. A notable first case involved a student who, in the pretest, obtained only 48% of overall correct answers, with difficulty in the inference tasks (40%). After the adaptive interaction sessions, the students achieved 82% overall correct answers in the posttest and 85% in inference, representing a 45-percentage point improvement. This progress was accompanied by the intensive use of feedback provided by the assistant and an increased number of answer reformulations following specific emotional and cognitive guidance.

Another significant case involved a student who initially scored 53% in critical analysis. Throughout the sessions, this student demonstrated systematic progress, successfully identifying the texts' argumentative fallacies and structural errors. In the posttest, the student achieved a 78% accuracy rate in critical analysis, accompanied by a notable decrease in adaptive interventions required, suggesting an increase in his interpretive autonomy.

These qualitative examples complement the quantitative findings, demonstrating that the assistant improved average performance and fostered deep learning and self-regulation processes in critical text comprehension skills.

FIGURE 5
Performance and engagement metrics during the intervention. (A) Comparison of performance between pretest and posttest across three cognitive skills: inference, critical analysis, and deduction. (B) Evolution of the error correction rate across five interaction sessions.

4.5 Qualitative analysis of the interviews

To complement the quantitative findings, a qualitative analysis of the semi-structured interviews conducted with the students at the end of the experimental phase was conducted. The interviews were transcribed and analyzed using inductive thematic coding, which enabled the identification of emerging categories based on observed response patterns. The analysis focused on capturing perceptions of the assistant's empathy, the usefulness of adaptive support, and the difficulties encountered during the interaction.

4.5.1 Perception of empathy in the assistant

One of the most recurring categories was the perception of empathy attributed to the educational assistant. Students highlighted the system's ability to accurately interpret emotional states and respond sensitively, generating genuine support.

One student expressed:

- "I felt like the assistant understood when I was frustrated with the more difficult readings; it responded in a more measured and encouraging manner." (Student 7)

Another participant reinforced this perception:

- "When I made mistakes, the assistant not only told me I was wrong, but also motivated me to try again, and that helped me not get discouraged." (Student 15)
- These testimonies reflect that the emotional analysis module and adaptive response generation conveyed an interaction experience perceived as empathetic, beyond simple error correction.

4.5.2 Usefulness of adaptive support

Another emerging category focused on the perceived usefulness of adaptive support in the learning process. Students particularly valued the personalized feedback and support in identifying their comprehension errors.

One interviewee commented:

- "Every time I made a mistake; the assistant gave me specific clues instead of just telling me the answer. That forced me to think more carefully." (Student 11)

Another student mentioned:

- "I felt like I was improving because they helped me see where I should focus in the text; they didn't just correct me, they guided me." (Student 3)

These testimonies illustrate that the adaptive approach was not only perceived as applicable but also as formative, favoring the development of metacognitive skills and critical self-assessment during task completion.

4.5.3 Difficulties faced in interaction

Despite the overall positive assessment, comments about specific difficulties experienced while using the assistant emerged. These difficulties were primarily related to moments when the feedback was not specific enough or when the emotions detected did not perfectly align with the user's actual state.

One student noted:

- "There were times when the assistant thought I was confused, but I was just hesitating between two correct answers." (Student 18)

Another participant stated:

- "Sometimes the suggestions it gave me were not that clear, and I had to reread them several times to understand what it was trying to tell me." (Student 22)

These difficulties suggest opportunities for improvement, particularly in the accuracy of emotional inference in ambiguous cases and in optimizing the language used in adaptive guides.

4.5.4 Synthesis of observed patterns

The qualitative analysis reveals that students perceived the assistant as an empathetic and helpful agent for their critical learning process. Positive perceptions of empathy and helpfulness formed a consistent pattern among participants. At the same time, the difficulties mentioned were less frequent and focused on the technical aspects of emotional interpretation and the specificity of feedback.

These findings reinforce the quantitative evidence presented in the previous results, demonstrating that adaptive and emotional capabilities in intelligent educational systems affect academic performance and alter users' subjective experience, enhancing their motivation, persistence, and autonomy in critical reading comprehension processes.

4.6 Comparison of the implemented system with existing work

To identify the contribution of the proposed educational assistant, a comparison is made with three relevant works identified in the recent literature review. These works represent significant approaches to applying artificial intelligence for adaptive and emotional learning in educational settings, but differ in their scope, level of cognitive-emotional integration, and evaluation methodology. The comparative analysis is structured around key criteria, including real-time emotional analysis, cognitive adaptability, response personalization, types of interaction, experimental evaluation methodologies, and the technical integration of NLP and emotional analysis. Table 5 summarizes the main characteristics of the proposal about the selected works.

Upon analyzing the comparison results, substantial differences emerge that favorably position the proposal over existing

TABLE 5 Comparison of the main characteristics between the developed proposal and existing works.

Comparative aspect	This proposal	Bauer et al. (2025)	Strielkowski et al. (2024)	Karataş et al. (2025)
Real-time emotional analysis	Yes	No	No	No
Cognitive adaptability	High	Average	High	High
Response personalization	Dynamic and progressive	Static	Limited dynamics	Conceptually based on adaptive levels
Type of interaction	Adaptive conversational	Based on simulated scenarios	Based on educational activities	Based on intelligent curricular adaptation
Experimental evaluation	Pretest-Posttest + Engagement + Emotional accuracy	Perceived satisfaction and simulated performance	Engagement and performance in adaptive programs	Conceptual study with qualitative validation
Application context	Adaptive critical and emotional reading comprehension	Simulation in teacher training	Sustainable educational transformation	Smart curriculum adaptation
Integration of NLP and emotional modules	Comprehensive and interconnected	Not integrated	Content adaptation (without NLP or emotions)	Does not address emotional processing

approaches. First, regarding real-time emotional analysis, only our proposal incorporates modules capable of interpreting emotional states from student interactions and dynamically adjusting pedagogical strategies accordingly. None of the compared works (Bauer et al., 2025; Strielkowski et al., 2024; Karataş et al., 2025) comprehensively addresses this dimension, which underscores the uniqueness of the approach adopted in this study.

Regarding cognitive adaptability, both our proposal and the work of Strielkowski et al. (2024) and Karataş et al. (2025) present elevated levels. However, while Strielkowski et al. implement adaptability focused on modifying learning paths based on overall performance, and Karataş conceptualizes curricular adaptation based on progressive adjustment strategies, our proposal introduces adaptability explicitly focused on critical reading comprehension. This adaptability is based on textual progress metrics, semantic analysis, and real-time emotional assessment, enabling dynamic adjustments to both the content and depth of feedback and the difficulty level of tasks.

Regarding response personalization, the presented proposal outperforms other approaches by implementing dynamic and progressive personalization that evolves as the system learns from the student's interactions, unlike Bauer et al. (2025), where personalization is static and limited to the type of error detected, and Karataş et al. (2025), where personalization is conceptual and oriented to predefined curricular levels, our system adapts both the content and the communicative style, the complexity of the clues, and the intensity of the emotional support throughout the learning process.

Regarding interaction, the proposal integrates an adaptive conversational interaction model, in which the assistant continuously interprets each conversational turn's cognitive and emotional dimensions to decide the best response. The other approaches analyzed involve interactions centered on preconfigured simulated scenarios (Bauer et al., 2025), educational content adaptations (Strielkowski et al., 2024), or curriculum adjustment proposals that were not implemented in a real conversational environment (Karataş et al., 2025).

This proposal also stands out for its methodological rigor regarding experimental evaluation. It employs a combination of structured pretests and posttests, engagement measurements,

misconceptions correction rates, and emotional accuracy assessments. In contrast, Bauer et al. (2025) report only perceived satisfaction and performance in simulations. Strielkowski et al. (2024) focus on engagement and educational achievement metrics without integrating explicit emotional assessments. Karataş et al. (2025) presents qualitative conceptual validations without direct experimental implementation. Regarding technical integration, our proposal robustly incorporates natural language processing and emotion analysis modules, interconnected through a microservices architecture designed to enable real-time adaptation and auditing of pedagogical decisions. These features position the proposal as a significant advance in the design of educational assistants based on conversational artificial intelligence, contributing to a better understanding of how to integrate—in an operational, not merely conceptual, manner—emotions, cognitive adaptability, and differentiated pedagogical strategies into intelligent systems designed for highly complex learning environments.

5 Discussion

Comparing the results obtained in this study with recent literature confirms relevant trends and highlights substantial differences that consolidate the contributions of the developed system. While works such as those by Doumanis et al. (2019) show that the integration of emotional capabilities in conversational agents improves the perception of engagement, the results obtained here expand this perspective by demonstrating that the joint incorporation of cognitive and emotional adaptability increases user satisfaction but also significantly impacts objective academic skills such as critical reading comprehension. Furthermore, unlike Bauer et al. (2025), who limit adaptive feedback to observable performance in simulations, the system presented in this study demonstrates that adaptive feedback based on emotional and cognitive indicators can catalyze a progressive improvement in both academic performance and emotional stability among students. These findings reaffirm that combining emotional modeling and dynamic adaptive strategies constitutes a practical approach to overcome the limitations identified in current state-of-the-art methods. The development and implementation

process focused on a microservices architecture that enabled the integration of specialized modules for NLP, emotional analysis, and progressive pedagogical adaptation. This functional separation facilitated the system's scalability and upgradeability and the fine-tuning of each component during the experimental phase.

The pretest-posttest experimental method, combined with engagement analysis and emotional feedback, offered a holistic evaluation that surpasses traditional methods based solely on subjective perception (Essel et al., 2024). The assessment of critical reading comprehension effectiveness demonstrated an average increase of 32.5% in correct answers, along with sustained improvements in the rate of correcting conceptual errors during interaction sessions, empirically supporting the hypothesis that an educational intervention mediated by adaptive conversational AI can produce significant learning in a brief period.

Beyond the measured academic impact, the results suggest that users' emotional experience directly influences their willingness to interact productively with the system. Qualitative analyses of the interviews revealed that the perception of empathy and personalized support were key factors in consolidating engagement and strengthening self-correcting metacognitive strategies. This finding is especially relevant, given that although the value of emotional engagement is recognized in previous literature (e.g., Strielkowski et al., 2024), its concrete effect on the development of complex cognitive skills has not been empirically documented.

The importance of this work lies in the construction of a model that not only responds to the student's cognitive needs and recognizes, interprets, and adapts its pedagogical strategy in real time based on the detected emotional state. This integration, far from being an aesthetic addition, constitutes an essential mechanism for personalizing what is taught, how it is taught, and when to intervene in a differentiated manner. In this sense, the proposal represents a significant advance in the evolution of educational assistants, moving from reactive approaches to initiative-taking and emotionally sensitive models that maximize the effectiveness of adaptive learning.

Furthermore, the methodology used to evaluate the system establishes a solid methodological precedent. The combination of objective measurements (correct answer percentage, correction rates) with qualitative analyses of emotional perception and engagement provides a more robust validation framework than approaches that focus solely on satisfaction or academic performance in isolation (Huang et al., 2022). This methodological triangulation strengthens the credibility of the results and provides a reliable platform for future extensions of the system to other learning domains.

This experimental study proposal presents several limitations that must be acknowledged to contextualize its results and guide future research. First, although the 2-week experimental period yielded measurable gains in academic performance and emotional engagement, it limits the extrapolation of long-term learning impacts. Sustaining adaptive strategies and metacognitive gains over time requires longitudinal studies, which are planned for future research phases.

Second, the emotional inference mechanism, although technically robust, depends solely on natural language processing of written input. This introduces a margin of uncertainty, particularly in cases of linguistic ambiguity, irony, or emotionally neutral expressions. While accuracy was reinforced through contextual sentiment analysis and custom fine-tuning with academic corpora, integrating multimodal signals—such as prosodic cues or facial expression recognition—remains a necessary evolution to improve affective sensitivity.

Third, the participant sample consisted exclusively of university students with a relatively homogeneous educational and cultural background. This limited diversity restricts the generalizability of the findings to broader populations or international contexts. Future evaluations will incorporate more heterogeneous participants to examine the assistant's adaptability to different linguistic registers, emotional expression styles, and pedagogical norms.

The assistant's design assumes that students are willing to communicate emotionally and authentically through a conversational interface. Although this assumption held during the controlled study, with high engagement levels, it may not generalize to high-stakes academic situations, where students could consciously or unconsciously alter their affective expressions, reducing the effectiveness of the emotional adaptation module.

6 Conclusions and future work

The study empirically demonstrates that the simultaneous integration of cognitive adaptability and emotional sensitivity capabilities in conversational artificial intelligence-based educational assistants constitutes a practical approach for improving critical reading comprehension skills. The results demonstrate that a model that adapts its pedagogical strategy in real time based on both cognitive responses and inferred emotional states can produce measurable gains in academic performance, improve students' emotional self-regulation capacity, and elevate levels of sustained engagement during learning activities.

The data collected in the experiment reveals average increases of over 30% in the assessed skills, particularly marked improvements in deduction and critical analysis processes, which traditionally represent significant challenges for university students. This academic performance is complemented by a progressive strengthening of conceptual error correction rates during adaptive sessions, suggesting greater content comprehension and the development of metacognitive skills that promote self-regulated learning. From a technical perspective, the implemented microservices-based architecture, combining natural language processing, emotional analysis, and pedagogical adaptation engines, proved effective in operational performance and functional scalability. This configuration enabled dynamic and continuous personalization of the interaction, adapting to the content of the students' responses and the emotional fluctuations detected throughout the session. Unlike more rigid approaches based on preconfigured decision trees, the developed system maintained adaptive plasticity, which was key to maintaining elevated levels of engagement and pedagogical accuracy.

Another relevant finding is the assistant's positive impact on the users' subjective experience. Surveys and interviews revealed that students not only perceived the system's academic usefulness but also highlighted its ability to provide emotional support, which they perceived as genuine and motivating.

This combination of functions reinforces the hypothesis that emotional components should not be considered accessory elements but central factors in designing effective intelligent educational systems.

Regarding innovation, the presented proposal overcomes the limitations observed in recent work. While previous approaches tend to specialize in a single axis—either cognitive or emotional—this study demonstrates that the integrated fusion of both planes yields superior results, both in objective academic dimensions and subjective affective parameters. Thus, a model is introduced that responds to errors or performance, anticipates emotional needs, and proactively adapts its pedagogical strategy, creating a new generation of emotionally intelligent educational assistants.

However, it is essential to acknowledge that the system was validated in a controlled environment and with a homogeneous sample regarding educational and cultural backgrounds. Although the results are robust within the proposed experimental context, their generalizability to more diverse populations requires further research phases that consider additional variables, such as linguistic diversity, cultural differences in emotional expression, and educational contexts with high levels of pressure or stress. Additionally, although emotion detection based on natural language processing achieved satisfactory levels of accuracy, there is still an inherent margin of error when interpreting emotions from written expressions, particularly in cases of ambiguity or irony. These aspects represent open challenges for improving more robust and multimodal affective detection models, which could integrate, for example, paralinguistic analysis or facial recognition in future versions of the system.

Regarding future work, the research is planned to expand in two main directions. First, longitudinal studies are intended to assess the long-term sustainability of the observed effects and the evolution of students' metacognitive autonomy once assisted support is withdrawn. Second, the system is projected to expand into other areas of knowledge where critical comprehension and emotional processing are equally important, such as analyzing scientific texts, resolving complex engineering problems, or interpreting cases in legal and medical disciplines. Additionally, integrating self-explanatory mechanisms generated by the assistant will be explored to strengthen transparency and pedagogical traceability in the adaptive process.

Data availability statement

The datasets generated and analyzed during the current study are not publicly available due to institutional confidentiality agreements and data governance policies. However, they may be made available by the corresponding author upon reasonable request and subject to compliance with applicable data access and ethical approval procedures.

Ethics statement

This study evaluates the performance of an AI-based educational assistant in a Moodle-based learning environment. No personal data were collected, no interventions were performed on participants, and no sensitive or identifiable information was used. According to the Reglamento Sustitutivo del Reglamento para la Aprobación y Seguimiento de Comités de Ética de Investigación en Seres Humanos (CEISH), published in the Registro Oficial No. 118 on August 2, 2022, and specifically under Article 43, this study qualifies as a "no-risk" research project. As such, it is exempt from requiring CEISH approval, as it does not involve human intervention, identifiable data, or sensitive variables. Participants interacted with the assistant as part of regular academic activities, were informed about the nature of the tool, and the data were fully anonymized. The research adhered to ethical principles, including informed participation, confidentiality, and respect for autonomy, and was conducted by the Declaration of Helsinki and national data protection laws. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

RG: Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. WV-C: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. JG: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aguayo, R., Lizarraga, C., and Quiñonez, Y. (2021). Evaluation of academic performance in virtual environments using the NLP model. *Rev. Iber. Sist. Tecnol. Inf.* 2021, 34–49. doi: 10.17013/risti.41.34-49

b18Macháček, D., Dabre, R., and Bojar, O. (2024). "Turning whisper into real-time transcription system," in *Proceedings of the 2023 International Joint Conference on Natural Language Processing (Demo Track)* (Bali: ACL). doi: 10.18653/v1/2023.ijcnlp-demo.3

Barik, K., and Misra, S. (2024). Analysis of customer reviews with an improved VADER lexicon classifier. *J. Big Data* 11, 1–29. doi: 10.1186/s40537-023-00861-x

Barros, P., and Sciutti, A. (2022). Across the universe: biasing facial representations toward non-universal emotions with the face-STN. *IEEE Access* 10, 103932–103947. doi: 10.1109/ACCESS.2022.3210183

Bauer, E., Sailer, M., Seidel, T., Niklas, F., Fischer, F., Greiff, S., et al. (2025). AI-based adaptive feedback in simulations for teacher education: an experimental replication in the field. *J. Comput. Assist. Learn.* 41, 1–20. doi: 10.1111/jcal.13123

Beevi, L. S., Vemuri, V. V., and Unnam, S. S. (2024). Sentiment analysis by using TextBlob and making this into a webpage. *AIP Conf. Proc.* 2802:120009. doi: 10.1063/5.0182186

Chheang, V., Sharmin, S., Marquez-Hernandez, R., Patel, M., Rajasekaran, D., Caulfield, G., et al. (2024). "Towards anatomy education with generative AI-based virtual assistants in immersive virtual reality environments," in *Proceedings - 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality, AIxVR 2024* (Los Angeles, CA: IEEE). doi: 10.1109/AIxVR59861.2024.00011

Doumanis, I., Economou, D., Sim, G. R., and Porter, S. (2019). The impact of multimodal collaborative virtual environments on learning: a gamified online debate. *Comput. Educ.* 130, 121–138. doi: 10.1016/j.compedu.2018.09.017

Elmashhara, M. G., De Cicco, R., Silva, S. C., Hammerschmidt, M., and Silva, M. L. (2024). How gamifying AI shapes customer motivation, engagement, and purchase behavior. *Psychol. Mark.* 41, 134–150. doi: 10.1002/mar.21912

Essel, H. B., Vlachopoulos, D., Essuman, A. B., and Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: receiving instant responses from AI-based conversational large language models (LLMs). *Comput. Educ. Artif. Intell.* 6:100198. doi: 10.1016/j.caeai.2023.100198

Hong, Q. (2024). Statistical vector model for the audience emotional response examination with affective computing based on conductor's expressive. *Int. J. Intell. Syst. Appl. Eng.* 12, 432–449. Retrieved from: https://ijisae.org/index.php/IJISAE/article/view/3987

Huang, S. Y. B., Huang, C. H., and Chang, T. W. (2022). A new concept of work engagement theory in cognitive engagement, emotional engagement, and physical engagement. *Front. Psychol.* 12:663440. doi: 10.3389/fpsyg.2021.663440

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., Mumthas, S., and Joseph, S. (2024). *The Cognitive Paradox of AI in Education: Between Enhancement and Erosion.* Available online at: https://www.ideta.io/blog-posts-english/top---ai-tools- (Accessed March 20, 2025).

Kaiss, W., Mansouri, K., and Poirier, F. (2023). Effectiveness of an adaptive learning chatbot on students' learning outcomes based on learning styles. *Int. J. Emerg. Technol. Learn.* 18, 250–261. doi: 10.3991/ijet.v18i13.39329

Karataş, F., Eriçok, B., and Tanrikulu, L. (2025). Reshaping curriculum adaptation in the age of artificial intelligence: MAPPING teachers' AI-driven curriculum adaptation patterns. *Br. Educ. Res. J.* 51, 154–180. doi: 10.1002/berj.4068

Maaz, N., Mounsef, J., and Maalouf, N. (2025). CARE: towards customized assistive robot-based education. *Front. Robot. AI* 12:1474741. doi: 10.3389/frobt.2025.1474741

Malik, M. A., and Shah, R. (2025). AI teachers (AI-based robots as teachers): history, potential, concerns and recommendations. *Front. Educ.* 10:1541543. doi: 10.3389/feduc.2025.1541543

Orellana, M., García, P. S., Ramon, G. D., Zambrano-Martinez, J. L., Patiño-León, A., Serrano, M. V., et al. (2024). Semantic similarity of common verbal expressions in older adults through a pre-trained model. *Big Data Cogn. Comput.* 8:3. doi: 10.3390/bdcc8010003

Pratama, A. Y., and Razaq, J. A. (2023). Integrasi sistem informasi akademik dan e-learning moodle dengan REST API. *J. Manaj. Inf. Sist. Inform.* 6, 95–107. doi: 10.36595/misi.v6i1.696

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., and Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information* 15:596. doi: 10.3390/info15100596

Sayeed, M. S., Mohan, V., and Muthu, K. S. (2023). BERT: a review of applications in sentiment analysis. *HighTech Innov. J* 4, 453–462. doi: 10.28991/HIJ-2023-04-02-015

Shchedrina, E., Valiev, I., Sabirova, F., and Babaskin, D. (2021). Providing adaptivity in moodle LMS courses. *Int. J. Emerg. Technol. Learn.* 16, 95–107. doi: 10.3991/ijet.v16i02.18813

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., and Vasileva, T. (2024). AI-driven adaptive learning for sustainable educational transformation. *Sustain. Dev.* 33, 1921–1947. doi: 10.1002/sd.3221

Tiezzi, J., Tyler, R., and Sharma, S. (2020). "Lessons learned: a case study in creating a data pipeline using Twitter's API," in *Proceedings of the 2020 Systems and Information Engineering Design Symposium (SIEDS)* (Charlottesville, VA: IEEE). doi: 10.1109/SIEDS49339.2020.9106584

Varghese, N. N., Jose, B., Bindhumol, T., Cleetus, A., and Nair, S. B. (2025). The power duo: Unleashing cognitive potential through human-AI synergy in STEM and non-STEM education. *Front. Educ.* 10:1534582. doi: 10.3389/feduc.2025.1534582

Velander, J., Taiye, M. A., Otero, N., and Milrad, M. (2024). Artificial intelligence in K-12 education: eliciting and reflecting on Swedish teachers' understanding of AI and its implications for teaching and learning. *Educ. Inf. Technol.* 29, 4085–4105. doi: 10.1007/s10639-023-11990-4

Vistorte, A. O. R., Deroncele-Acosta, A., Ayala, J. L. M., Barrasa, A., López-Granero, C., and Martí-González, M. (2024). Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. *Front. Psychol.* 15:1387089. doi: 10.3389/fpsyg.2024.1387089

Wölfel, M., Shirzad, M. B., Reich, A., and Anderer, K. (2024). Knowledge-based and generative-AI-driven pedagogical conversational agents: a comparative study of Grice's cooperative principles and trust. *Big Data Cogn. Comput.* 8, 1–20 doi: 10.3390/bdcc8010002

Yang, Y., and Zeng, F. G. (2024). Syllable-rate-adjusted-modulation (SRAM) predicts clear and conversational speech intelligibility. *Front. Hum. Neurosci.* 18:1324027. doi: 10.3389/fnhum.2024.1324027