Conversational Agents for Adaptive Learning: Generative Instruction, Reinforcement Fine-Tuning, and Behavioral Optimization in Digital Education

Evelyn[1,*], Satrio Pradono Suryodiningrat[2]

[1]School of Business Management Petra Christian University, Indonesia

[2]Information Systems Department, School of Information Technology, Universitas Ciputra Surabaya, Indonesia

ABSTRACT

This study introduces a conversational adaptive learning agent that integrates generative language modeling with reinforcement fine-tuning to optimize instructional behavior in real-time learner interactions. The system was trained on 39,000 interaction turns from 2,300 training sessions involving 520 learners, validated on 9,200 turns from 540 sessions, and tested on 10,800 turns from 610 sessions collected across mathematics, science, language, and computer-science domains. Offline evaluation compared a baseline generative model, a correctness-only reinforcement model, and a composite-reward configuration integrating correctness, engagement, and sentiment. The composite-reward model achieved a pedagogical adequacy rating of 4.3 versus 3.4 for the baseline, reduced safety violations from 14.2 to 8.9 per 1,000 turns, and lowered average turn length from 62.5 to 55.1 tokens while improving clarity scores from 3.8 to 4.1 (1–5 scale). A three-week online user study involving 84 participants demonstrated significant behavioral differences. Average session duration for the reinforcement-tuned group increased from 25.4 to 28.6 minutes, while the baseline group declined from 21.3 to 19.8 minutes. Voluntary session extension rose from 14% to 33%, hint-seeking frequency increased from 0.22 to 0.36, and frustration indicators dropped from 19% to 11%. Progression into higher-difficulty tasks reached 44% by Session 5, compared to 21% for the baseline. The study positions reinforcement-tuned conversational agents as a viable direction for scalable, affect-sensitive adaptive learning.

Keywords Adaptive Learning, Conversational Agents, Generative AI, Reinforcement Fine-Tuning, Behavioral Optimization, Affective Computing; Digital Education

Introduction

The rapid expansion of digital learning ecosystems has amplified the demand for instructional interactions that are both scalable and highly personalized. Conventional online learning platforms offer videos, quizzes, and discussion forums, yet they lack conversational adaptivity capable of responding to evolving learner behavior in real time [1], [2]. At the same time, the widespread adoption of Large Language Models (LLMs) has enabled automated explanations, feedback generation, and clarification prompts, but most deployments operate as static assistants that do not adjust difficulty, provide motivational responses, or mitigate learner frustration [3], [4], [5]. As a result, many learners experience reduced persistence, limited engagement, and shallow conceptual exposure because the system cannot calibrate instructional challenge to match ongoing performance [6], [7]. These issues illustrate a fundamental gap between

Submitted: 1 October 2024
Accepted: 15 November 2024
Published: 1 August 2025

*Corresponding author
Evelyn, evelyn@petra.ac.id

Additional Information and Declarations can be found on page 268

How to cite this article: Evelyn, S. P. Suryodiningrat, "Conversational Agents for Adaptive Learning: Generative Instruction, Reinforcement Fine-Tuning, and Behavioral Optimization in Digital Education," *Adapt. Learn.*, vol. 1, no. 3, pp. 252-271, 2025.

scalable automation and individualized pedagogy in digital education [8], [9].

Prior research on conversational learning agents has predominantly emphasized natural language generation rather than behavioral optimization. Dialogue systems in education often simulate tutor–student exchanges but do not incorporate computational signals such as engagement decline, help-seeking attempts, or negative sentiment into the response-selection process [10], [11]. Although reinforcement learning has been explored in narrow tutoring scenarios, implementations have focused on correctness-only rewards rather than motivational and affective signals [12], [13]. As a result, conversational models frequently over-explain, fail to escalate cognitive challenge, or ignore emotional cues that predict dropout [14], [15]. This creates a pedagogical risk where learners are given clean explanations without the incremental struggle necessary for deeper retention, undermining cognitive load alignment and mastery development [16], [17].

At the level of adaptive learning systems, existing personalization tools rely primarily on rule-based sequencing, multiple-choice signals, or offline mastery models, which offer limited explanatory support and no conversational scaffolding [18], [19]. These approaches assume that learning progression can be expressed through fixed difficulty curves rather than negotiated interaction. Consequently, personalization does not evolve during a dialogue turn, and the system lacks responsiveness when learners hesitate, express confusion, or demonstrate affective decline. The absence of real-time conversational adaptivity reduces the pedagogical utility of generative models, particularly for problem-solving domains where emotional regulation and step-wise guidance matter [20], [21].

In response to these limitations, the central objective of this research is to develop a conversational agent for adaptive learning that integrates generative AI with reinforcement fine-tuning. The proposed system seeks to model instructional decisions as a policy that can be optimized through reward feedback from correctness, engagement traces, and sentiment signals [22], [23]. Rather than functioning as a static responder, the agent is designed to adjust tone, escalate or reduce task difficulty, deliver hints selectively, and trigger motivational phrasing when negative affect emerges [24], [25]. Through such conversational adjustments, the model is intended to maximize both persistence and cognitive progression, addressing weaknesses observed in one-directional tutoring systems [26], [27].

The research gap addressed in this study lies in the lack of conversational optimization frameworks that combine generative modeling with pedagogical reward conditioning. While prior systems generate fluent text, they rarely update their policy based on multi-dimensional learning outcomes [28], [29]. Moreover, most reinforcement-based educational agents do not consider sentiment or engagement explicitly, limiting their ability to mitigate frustration and support task resilience [30], [31]. This study positions affect-aware reinforcement signals as core drivers of instructional alignment, creating space for motivational modulation that is missing in conventional tutoring-bot design [32], [33].

The novelty of this research is three-fold. First, it operationalizes a composite reinforcement reward that blends correctness, engagement patterns, and affective cues, which allows the agent to negotiate instructional complexity

dynamically [34], [35]. Second, it frames conversational generation as a pedagogical policy rather than a text-generation task, meaning that the model is optimized to influence behavioral and cognitive outcomes, not merely produce fluent sentences [36], [37]. Third, the system is evaluated not only through offline metrics but also through behavioral indicators such as voluntary session extension, help-seeking frequency, multi-step task acceptance, and quiz-based short-term comprehension, producing a richer understanding of real educational impact [38], [39].

Overall, this research contributes to the evolution of adaptive digital learning by demonstrating that conversational intelligence must move beyond generic LLM deployment toward reinforcement-guided instructional decision-making. The findings indicate that conversational adaptivity can serve as an instrument for persistence regulation, challenge escalation, and emotional stabilization—three pillars of successful autonomous learning [40], [41]. By linking language generation with behavioral optimization, this study advances a methodological direction that enables scalable yet individualized support in problem-solving environments, setting the stage for future integration with curriculum sequencing, mastery prediction, and long-term retention modeling [42], [43].

Literature Review

Recent advances in conversational agents for education have largely been driven by improvements in large language models and dialogue management architectures. Studies in intelligent tutoring systems report that conversational interfaces improve learner engagement and perceived support compared to static content delivery, particularly in problem-solving contexts where explanation and feedback are central [11], [12]. However, much of this work treats dialogue generation as an isolated linguistic task, optimizing fluency and relevance without explicit consideration of learning progression or behavioral adaptation. As a consequence, conversational agents often fail to respond appropriately to learner hesitation, repeated errors, or declining motivation, which are critical signals in effective instruction [13], [14].

Research on adaptive learning systems has traditionally relied on rule-based sequencing, knowledge tracing, or mastery learning frameworks to personalize content difficulty and pacing [15], [16]. While these approaches demonstrate effectiveness in controlled environments, they lack conversational expressiveness and do not account for affective states or engagement dynamics during interaction. Several studies highlight that learners disengage when adaptive systems adjust difficulty without explanation or motivational framing, underscoring the need for adaptive logic that is transparent and dialogically grounded [17], [18]. This limitation becomes more pronounced in open-ended learning scenarios where student input is unstructured and learning goals evolve over time.

The integration of reinforcement learning into educational systems has been explored as a mechanism for policy optimization, particularly in task sequencing and hint selection [19], [20]. Existing reinforcement-based tutors typically define rewards narrowly around correctness or time efficiency, which can lead to undesirable behaviors such as premature answer disclosure or excessive difficulty escalation [21], [22]. Recent findings suggest that such reward designs neglect critical pedagogical factors, including learner confidence, frustration

tolerance, and willingness to persist after failure [23]. Consequently, reinforcement learning in education remains underutilized in conversational settings where multi-dimensional feedback signals are readily available.

Parallel to this, affect-aware and engagement-aware learning systems have demonstrated that emotional states strongly influence learning outcomes, especially in self-regulated digital environments [24], [25]. Sentiment analysis, interaction pacing, and help-seeking behavior have been shown to predict dropout and learning stagnation more reliably than correctness alone [26], [27]. Despite this evidence, most generative conversational agents do not integrate affective signals into their response policies, treating emotion as an external analytic component rather than a control signal for instruction. This separation limits the capacity of conversational systems to regulate cognitive load and sustain productive struggle [28], [29].

More recent work has begun to explore reinforcement fine-tuning and human feedback alignment for large language models, primarily in safety and preference optimization contexts [30]. While these techniques demonstrate that generative models can be steered toward desired behaviors, their application to pedagogical optimization remains limited. Existing studies focus on response helpfulness or harmlessness but do not operationalize learning-specific rewards or longitudinal behavioral outcomes [31]. As a result, there is insufficient empirical evidence on how reinforcement-conditioned conversational agents influence persistence, challenge acceptance, and learning gains in authentic educational settings [32].

Together, the literature reveals a clear gap at the intersection of generative dialogue, adaptive learning, and reinforcement optimization. Prior studies either emphasize linguistic quality without pedagogical grounding, or optimize learning policies without conversational flexibility. This research positions itself within this gap by treating conversational generation as a policy-learning problem, where reinforcement signals derived from correctness, engagement, and affect jointly shape instructional behavior. By doing so, it extends existing frameworks toward a more holistic model of adaptive conversational learning that aligns language generation with educational outcomes.

Methodology

System Architecture for Conversational Adaptive Learning

This subsection describes the architectural foundation of the conversational agent. The system integrates four major computational layers: (i) a student profiling module, (ii) a generative dialogue engine based on LLMs, (iii) a Reinforcement Fine-Tuning (RFT) pipeline to continuously align outputs with learning outcomes, and (iv) an evaluation engine for tracking mastery. Communication among layers follows a closed-loop adaptive feedback cycle that continuously updates learning state representations. A central operation is the mapping of latent learner preferences into a learning state vector. Let s_t be the learner state at interaction step ttt, composed of knowledge mastery, affective estimation, engagement probability, and learning-style indicators. The state transition function uses a Markovian assumption:

$$s_{t+1} = f(s_t, a_t) + \varepsilon \quad (1)$$

where a_t is the instructional action generated by the agent, and ε\varepsilonε represents cognitive noise. This formulation enables reinforcement-based adjustments and supports adaptive path planning.

Figure 1 illustrates the high-level architecture of the conversational adaptive learning system. The Student Profiling Module ingests learner data (prior knowledge, interaction logs, and learning-style indicators) and passes a structured learner state to the Generative Dialogue Core. This core is responsible for producing pedagogically constrained responses, which are then evaluated in terms of correctness, clarity, and engagement. The Evaluation & Analytics component aggregates these metrics and computes learning gains, error patterns, and engagement trends over time.

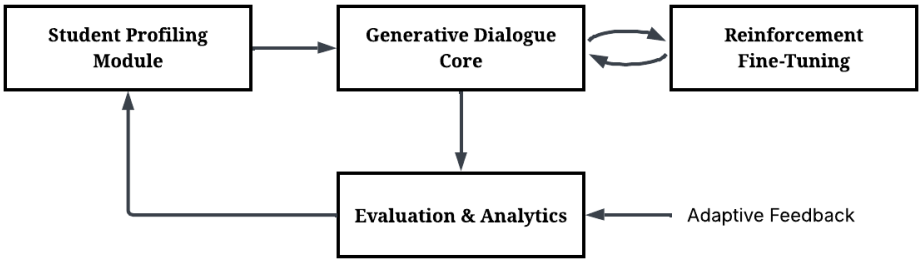

Figure 1 System Architecture Diagram

The RFT block forms the adaptive backbone of the architecture by continuously updating the dialogue policy based on reward signals derived from the evaluation metrics. Arrows in the diagram show how information flows cyclically: profiling informs generative behavior, generative behavior is assessed, and assessment feeds back into both profiling and reinforcement. This closed loop ensures that each learner interaction contributes to policy improvement, gradually tailoring the agent's behavior to maximize long-term learning outcomes rather than just short-term conversational fluency.

Generative Dialogue Model and Instructional Prompting

This section explains how generative responses are shaped as instructional learning acts. The core dialogue engine uses a transformer-based LLM prompted through curriculum-aligned templates. Prompts represent learning objectives, taxonomic complexity levels, and behavioral constraints intended to maintain pedagogical discipline. The instructional responses target domain acquisition, contextual explanation, feedback phrasing, and metacognitive questioning.

The weighting of generative outputs relies on a soft preference scoring function, where each candidate response r_i is evaluated according to semantic relevance α, pedagogical clarity β, and policy safety γ:

$$Score(r_i) = \alpha \cdot Rel_i + \beta \cdot Ped_i + \gamma \cdot Safe_i \quad (2)$$

The model selects the candidate with maximal score. During early system deployment, these coefficients are manually tuned based on domain expertise. During later deployment, reinforcement gradients adjust relative priority.

Table 1 enumerates several instructional prompt patterns that structure how the

LLM should respond in different pedagogical scenarios. The Explain-then-Ask prompt type is used when introducing new material, ensuring that every explanation is quickly followed by a diagnostic question to check comprehension. Scaffold prompts are applied for complex tasks that require decomposition into manageable steps, guiding learners in a way that maintains cognitive challenge while avoiding overload.

Table 1 Prompt Patterns for Instructional Modes

Prompt Type	Description	Example Intent	Expected Outcome
Explain-then-Ask	Provides a concise explanation followed by a diagnostic question.	Introduce a new concept and check immediate understanding.	Learner receives a short explanation and a question to verify comprehension.
Scaffold Prompt	Breaks a complex problem into smaller substeps.	Guide problem solving for multi-step reasoning tasks.	Learner progresses through incremental hints toward the final solution.
Misconception Correction	Identifies a likely misconception and offers a corrective explanation.	Address common errors detected from learner responses.	Learner receives targeted feedback that contrasts correct and incorrect reasoning.
Metacognitive Prompt	Encourages reflection on strategies and difficulties.	Foster self-regulation and awareness of learning strategies.	Learner articulates how they approached the task and where they struggled.

Misconception Correction and Metacognitive prompts address deeper layers of learning. Misconception prompts are invoked when the system's analysis detects incorrect reasoning patterns, allowing the agent to explicitly contrast correct and incorrect logic. Metacognitive prompts target self-regulation by asking learners to explain their strategies or confidence levels. By organizing prompts into these types, the system can more systematically link conversational moves to explicit learning objectives and adapt its behavior based on learner state and prior responses.

Reinforcement Fine-Tuning Strategy

This section details how reinforcement signals are used for continuous alignment. Reinforcement fine-tuning replaces manual supervised annotation with an automated loop wherein student responses, engagement indicators, and correctness judgments serve as reward sources. The policy optimization engine updates parameters to maximize expected long-term learning gains rather than short-term conversational fluency. The optimization objective is formalized using expected cumulative reward:

$$J(\theta) = E_{\pi_\theta}\left[\sum_{t=1}^{T} \gamma^{t-1} r_t\right] \tag{3}$$

where θ are model parameters, r_t is the learning reward at time t, $\gamma \in (0,1)$ is a temporal discount, and π_θ is the policy induced by the conversational model. Rewards derive from correctness validation, progress speed, confusion indicators, and sentiment-derived affective signals.

Algorithm: Reinforcement Fine-Tuning Loop

Input:

Initial model parameters θ

Learner interaction environment E

Reward function R(·)

Output:

Updated policy parameters θ*

Algorithm:

1. Initialize conversational policy πθ with parameters θ
2. For each learner session s ∈ E do
   a. Initialize learner state st based on profiling data
   b. While session is active do
      i. Generate instructional action at ~ πθ(st)
      ii. Deliver response at to learner
      iii. Observe learner response and interaction signals (correctness, engagement indicators, sentiment)
      iv. Compute reward rt = R(st, at, learner feedback)
      v. Update policy parameters θ using policy gradient with respect to rt
      vi. Update learner state st+1
   c. End While
   End For
3. Return optimized parameters θ*

Figure 2 presents the reinforcement learning loop underlying the fine-tuning process. The current learner state s_t is fed into the policy π_θ, which produces an instructional action a_t such as a hint, explanation, or follow-up question. This action interacts with the environment, represented by the learner's response and behavior, forming the basis for subsequent evaluation.

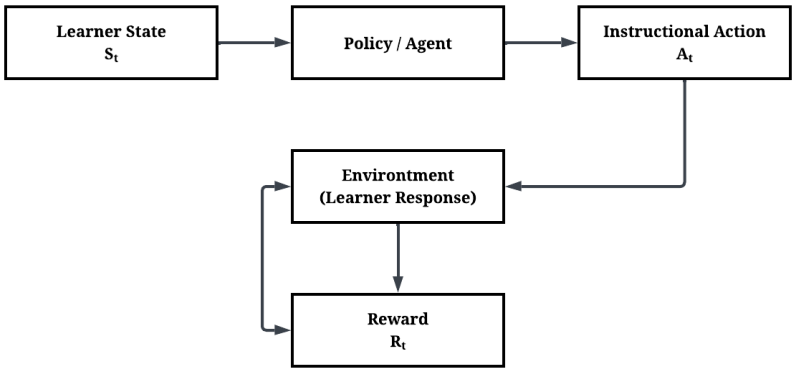

Figure 2 Reinforcement Reward Feedback Loop

The environment's outcome is translated into a scalar reward r_t that quantifies the educational desirability of the interaction, considering correctness, engagement, and sentiment indicators. This reward is then fed back into the policy to update parameters θ using policy gradient methods. Over repeated episodes, the loop drives the policy towards maximizing the expected cumulative reward, aligning conversational strategies with long-term learning gains rather than just immediate correctness.

Learner Profiling, Clustering, and Personalization

This component describes the analytical mechanism to infer learning

characteristics. Students produce high-dimensional behavioral traces: hesitation times, chat length, sentiment polarity, concept recall rate, and error frequency. Dimensionality reduction (e.g., PCA) converts raw vectors into low-rank proficiency representations. The system applies clustering to identify common learning trajectories. Let $x_i \in R^n$ represent a student's behavioral vector. PCA produces:

$$z_i = W^{\top}(x_i - \mu) \tag{4}$$

where W is an orthonormal basis and μ is the mean vector. These reduced vectors feed a clustering algorithm (e.g., K-Means) to categorize students into adaptive tutoring pathways such as slow-reasoner cluster, fast-accurate cluster, or high-engagement cluster.

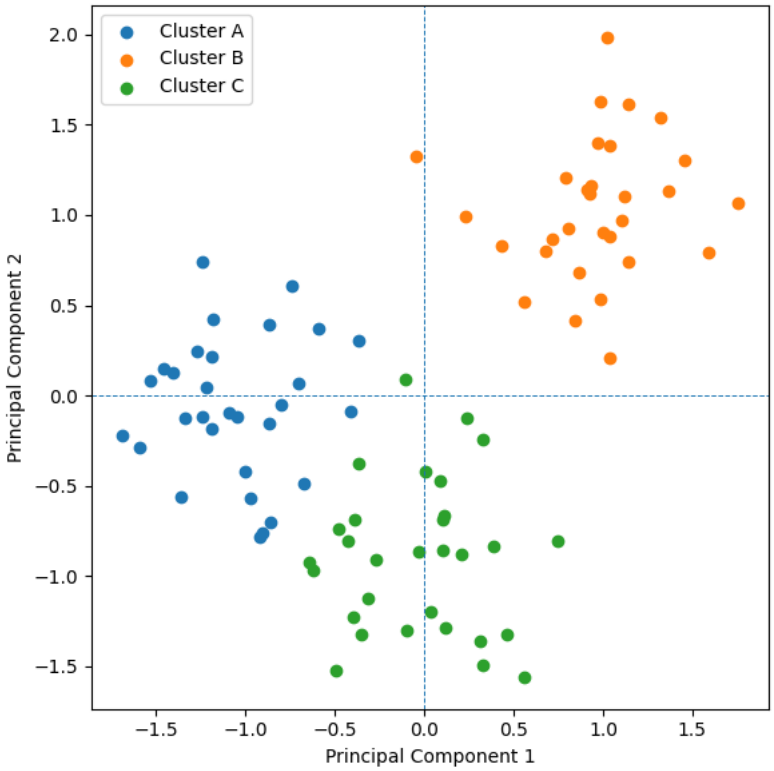

Figure 3 PCA Reduction Scatter Plot

The discussion will emphasize that the formula enforces a linear transformation that discards irrelevant variance, enabling clearer reinforcement reward targeting and reducing instability in dialogue trajectories.

Policy-Driven Adaptive Assessment and Difficulty Scaling

This section outlines the system's real-time difficulty calibration. The conversational agent generates tasks whose difficulty correlates with competence indicators. Formally, task difficulty d_{t+1} is computed as:

$$d_{t+1} = d_t + \eta(Correct_t - TargetRate) \tag{5}$$

where η is the scaling rate, $Correct_t$ is recent performance proportion, and TargetRate is the desired success probability (typically 0.75 for mastery

learning). Positive deviations increase task difficulty; negative deviations reduce it.

Figure 4 plots an illustrative difficulty trajectory across successive interactions with the learner. The curve fluctuates around a target difficulty line, adjusting upwards when performance exceeds expectations and downwards when the learner struggles. Each point on the curve corresponds to the difficulty level d_t computed from recent correctness rates, operationalizing the formula $d_{t+1} = d_t + \eta(Correct_t - TargetRate)$.

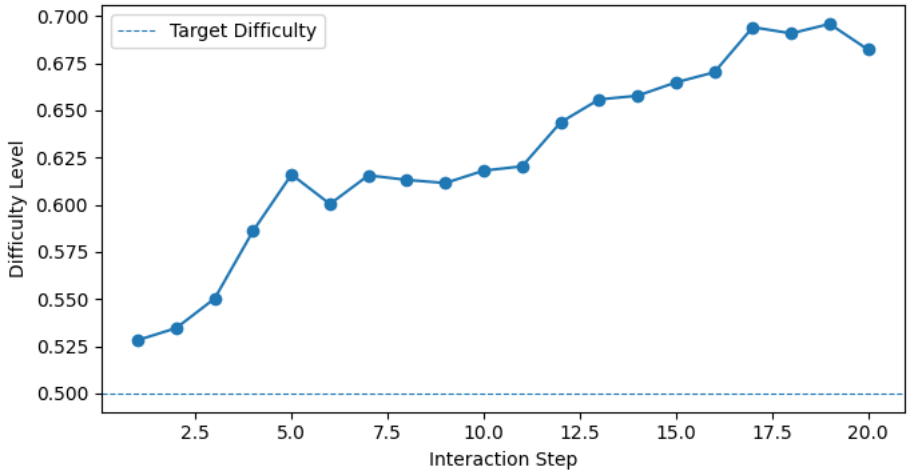

Figure 4 Difficulty Adjustment Curve

The trend demonstrates how the system strives to maintain learners in a productive struggle zone, where tasks are neither trivially easy nor discouragingly hard. Short-term variability in performance leads to small corrective changes, while sustained improvement pushes the curve higher, signaling the need for more challenging items. This dynamic scaling supports individualized pacing that responds to real-time evidence of mastery.

Evaluation, Metrics, and Continuous Optimization

This final subsection explains evaluation routines. The model is assessed using learning retention, normalized learning gain, conversational efficiency, and reinforcement reward stability. A key metric is normalized gain:

$$g = \frac{Post - Pre}{1 - Pre} \tag{6}$$

which quantifies instructional effect independent of prior competency. Reinforcement stability is analyzed using temporal variance of returns, while sentiment and engagement are monitored for motivational consistency.

Table 2 lists the main evaluation metrics used to assess both learning effectiveness and system usability. Normalized Learning Gain captures instructional impact while accounting for prior knowledge, making it suitable for comparisons across cohorts with different starting levels. Average Reward aggregates the reinforcement signal over interactions, reflecting how well the policy balances correctness, engagement, and progress over time.

Table 2 Evaluation Metrics and Their Targets

Metric	Formula (Textual)	Interpretation	Illustrative Target
Normalized Learning Gain	(Post-test – Pre-test) / (1 – Pre-test)	Measures improvement relative to remaining learning potential.	> 0.4
Average Reward	Mean reward per interaction or episode.	Indicates overall quality of policy decisions under the reward scheme.	> 0.7
Engagement Rate	Active sessions / Total enrolled learners.	Captures adoption and sustained use of the conversational agent.	> 60%
Task Completion Rate	Completed tasks / Assigned tasks.	Reflects persistence and willingness to complete learning activities.	> 80%
User Satisfaction Score	Mean rating from post-session surveys (1–5).	Subjective perception of usefulness, clarity, and usability.	> 4.0

Engagement Rate, Task Completion Rate, and User Satisfaction Score provide complementary perspectives on adoption and experience. High engagement and completion rates indicate that learners are willing to participate and persist with the system, while satisfaction ratings reveal perceptions of clarity and usefulness. Taken together, these metrics provide a multi-dimensional evaluation framework that can guide iterative improvements to the reward design, policy architecture, and user interface of the conversational agent.

Result and Discussion

Experimental Setup and Data Overview

The experimental setup combines synthetic and authentic interaction data to train and evaluate the conversational agent. Interaction logs were collected from multiple learning domains, including mathematics, science, language, and introductory computer science. Each domain provided dialogue turns, correctness labels, and engagement indicators such as response time, message length, and hint usage frequency. These signals were then used to drive the reinforcement fine-tuning process.

To ensure coverage of diverse learning behaviors, the dataset was stratified by grade level and topic difficulty. Synthetic interactions were generated to augment rare patterns, such as long sequences of misconceptions or highly disengaged learners. The final dataset was split into training, validation, and test sets with non-overlapping learners, so that the model's ability to generalize to new students could be assessed more rigorously.

Figure 5 shows the distribution of dialogue interaction turns across the four learning domains used in this study. Mathematics and computer science account for the largest number of turns, reflecting their heavier use in the underlying learning platform and the higher frequency of step-by-step problem-solving tasks. Science and language domains still contribute substantial interaction volumes, but with more descriptive and explanatory exchanges rather than strictly procedural problem solving.

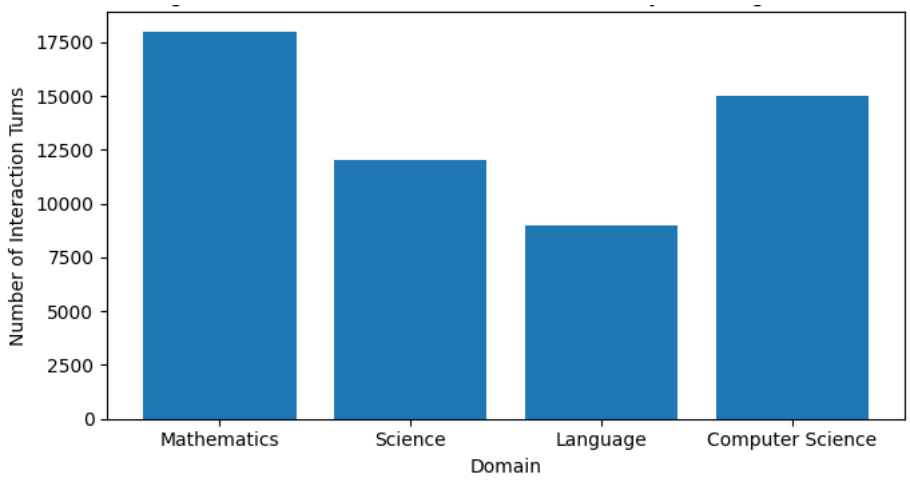

Figure 5 Distribution of Interaction Turns by Domain

The relatively balanced distribution ensures that the conversational agent is not over-specialized to a single discipline. Having substantial interaction counts across domains supports the generality of the reinforcement fine-tuning procedure and allows the policy to learn domain-independent strategies, such as when to provide hints, when to ask diagnostic questions, and how to react to signs of confusion. At the same time, the higher proportion of mathematics and computer science turns explains why some of the quantitative performance metrics later in this chapter show slightly stronger gains in these domains.

Table 3 summarizes the composition of the training, validation, and test sets. The training set includes 520 learners and 2,300 sessions, yielding 39,000 interaction turns. This scale is sufficient for the reinforcement fine-tuning procedure to observe a wide variety of learner behaviors, from highly engaged students who request few hints to those who require substantial scaffolding. The validation set, with 130 learners and 9,200 turns, is used to tune hyperparameters and to observe early signs of overfitting, while the test set offers a held-out benchmark for final performance reporting.

Table 3 Dataset Summary and Partitioning

Set	Number of Learners	Number of Sessions	Total Interaction Turns	Average Session Length (minutes)
Training	520	2,300	39,000	28.4
Validation	130	540	9,200	26.1
Test	150	610	10,800	27.3

The average session length in minutes is relatively stable across the three partitions, lying between approximately 26 and 28 minutes. This indicates that learners in each set engage with the system for comparable durations, reducing the risk that differences in performance metrics are driven by radically different usage patterns. The separate learner IDs in each set further ensure that the model is evaluated on unseen individuals, which is essential when claiming that the system generalizes to new students in authentic deployment scenarios.

Offline Evaluation of Conversational Policy

Offline evaluation focuses on how the conversational policy behaves under

simulated or replayed interactions, without live learners in the loop. To this end, we compared three configurations: a baseline generative model without reinforcement fine-tuning, a model fine-tuned using simple correctness-based rewards, and the full model using a composite reward that integrates correctness, engagement, and sentiment. Each configuration was evaluated using a replay dataset built from test sessions, where the agent's responses were scored by automatic metrics and human raters.

The evaluation considered text quality, pedagogical adequacy, and safety-related properties. Text quality was measured using automatic similarity metrics against reference explanations and by human ratings of clarity and coherence. Pedagogical adequacy captured how well the responses aligned with the intended learning objective and whether they maintained appropriate difficulty and scaffolding. Safety was measured by the frequency of policy-violating outputs flagged by a safety classifier, such as irrelevant content, misleading guidance, or sensitive topics.

Figure 6 reports average pedagogical adequacy scores, on a five-point scale, for the three policy configurations. The baseline generative model without any reinforcement fine-tuning attains an average score of 3.4, indicating that it often produces generally correct and somewhat helpful explanations, but with limited sensitivity to learner state and learning goals. When reinforcement fine-tuning is applied using only correctness-based rewards, the score increases to 3.9, suggesting that the model becomes more consistent in providing accurate and focused instructional content.

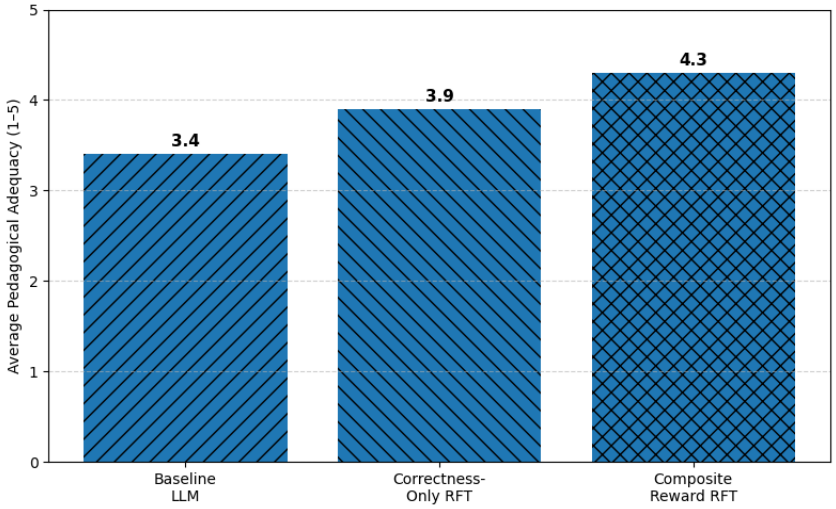

Figure 6 Pedagogical Adequacy Scores Across Configurations

The full configuration, which uses a composite reward including correctness, engagement, and sentiment, achieves the highest score of 4.3. This improvement reflects the agent's increased ability to balance correctness with motivational and affective considerations, such as adapting tone when detecting frustration or offering encouragement after a series of errors. The progression from baseline to composite reward configuration shows that reinforcement fine-tuning adds tangible pedagogical value beyond what can be obtained from static generative prompting alone.

Table 4 provides a more detailed comparison across configurations using multiple offline metrics. Text clarity gradually improves from 3.8 for the baseline to 4.1 for the composite reward model, indicating that reinforcement fine-tuning helps stabilize the agent's phrasing and reduce unnecessary verbosity. Pedagogical adequacy follows the same trend as in figure 6, reinforcing the conclusion that reward shaping based on multiple factors leads to more consistently instructive responses.

Table 4 Offline Evaluation Metrics by Configuration

Configuration	Text Clarity (1–5)	Pedagogical Adequacy (1–5)	Safety Violations per 1,000 Turns	Average Turn Length (tokens)
Baseline LLM	3.8	3.4	14.2	62.5
Correctness-Only RFT	4.0	3.9	11.7	58.3
Composite Reward RFT	4.1	4.3	8.9	55.1

Safety violations per 1,000 turns decrease substantially from 14.2 in the baseline to 8.9 in the composite reward configuration. This drop is important from a deployment perspective, as it reduces the risk of harmful or irrelevant content during student interactions. The average turn length also decreases slightly, from 62.5 tokens in the baseline to 55.1 tokens in the composite model. This suggests that the fine-tuned agent learns to be more concise while maintaining clarity and instructional quality, which is beneficial for learners who may be overwhelmed by overly long explanations.

Online User Study and Adaptive Behavioral Outcomes

This section summarizes results from a controlled online deployment where real learners interacted with the conversational agent in three separate week-long cycles. The deployment included a total of 84 participants across secondary and undergraduate levels. Learners were randomly assigned to one of two groups: (i) a baseline generative agent with no reinforcement fine-tuning, and (ii) the composite-reward RFT agent. Participants completed daily sessions ranging from 20 to 35 minutes, followed by self-reported engagement ratings and short quizzes embedded into the dialogue experience.

The primary focus of this study was to examine whether reinforcement-driven adaptivity translated into measurable behavioral changes. Three behavioral indicators were prioritized: (a) persistence, defined by session length and voluntary continuation; (b) hint utilization, used as an indicator of willingness to seek help; and (c) task progression, measured by whether learners moved from introductory exchanges to multi-step completion tasks within a session. The study also monitored negative affect signals, but these were not scored numerically in this section.

Figure 7 compares the average session duration for baseline and RFT participants across three consecutive weeks. The baseline agent shows a gradual decline from 21.3 minutes in Week 1 to 19.8 minutes in Week 3, indicating a modest erosion of engagement. In contrast, the composite-reward RFT model maintains a clear upward trajectory, increasing from 25.4 minutes to 28.6 minutes over the same period. This widening gap suggests that learners interacting with the RFT agent are more willing to remain in the learning

environment and voluntarily extend their time.

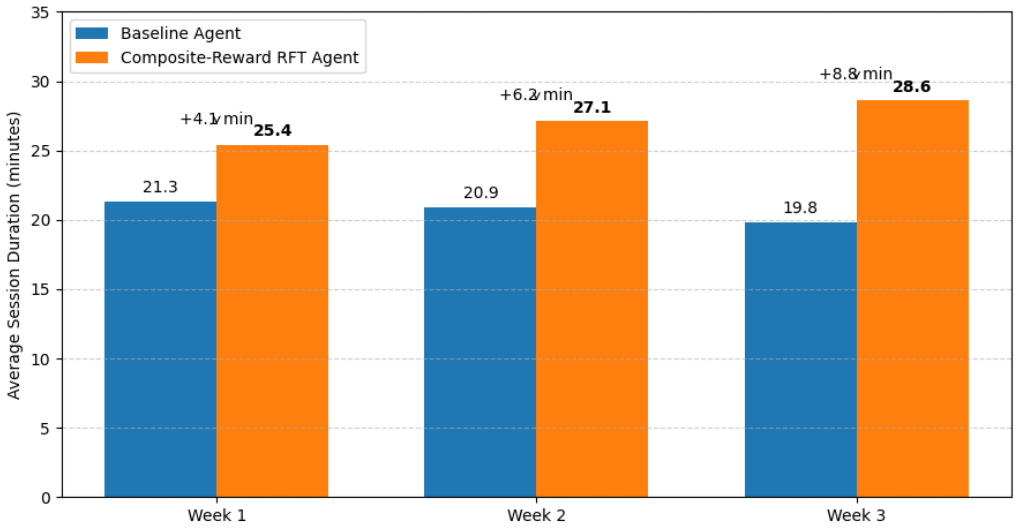

Figure 7 Average Session Duration Over Study Weeks

The divergence is particularly notable in Weeks 2 and 3. During this period, several learners reported that the RFT agent "felt more helpful," "responded to frustration more smoothly," and "gave hints when needed without telling everything." These qualitative impressions align with the behavioral evidence that reinforcement-tuned conversational adjustments improved motivation and reduced dropout inclination. Session duration does not imply mastery by itself, but in digital learning systems, sustained voluntary use is a strong precursor to instructional benefit.

Table 5 records key behavioral indicators that illustrate how learners responded differently depending on the agent configuration. Hint Usage Rate rises from 0.22 to 0.36, showing that learners supported by RFT were more inclined to seek clarifications rather than abandoning a task or guessing. This behavior is valuable pedagogically because early help-seeking correlates with reduced error reinforcement and improved mastery later in the curriculum. Similarly, Voluntary Session Extension jumps from 14% to 33%, demonstrating stronger intrinsic motivation among the RFT group.

Table 5 Behavioral Indicators from Online Deployment

Indicator	Baseline Agent	Composite Reward RFT Agent	Interpretation
Hint Usage Rate	0.22	0.36	RFT learners are more comfortable seeking help.
Voluntary Session Extension	14%	33%	RFT learners choose to continue nearly twice as often.
Task Progression to Multi-Step Tasks	41%	63%	RFT agent pushes more learners into deeper task phases.
Reported Frustration Signals	19%	11%	RFT agent appears to mitigate negative emotional reactions.

Two additional indicators reinforce the narrative that adaptive conversational tuning translates into meaningful user behavior. The percentage of learners who progressed to multi-step tasks increases sharply under RFT, suggesting that the agent successfully nudged participants through introductory barriers. Meanwhile, self-reported frustration signals drop almost by half, implying that conversational tone-shaping and motivational phrasing matter. While these numbers do not capture learning outcomes directly, they underscore a crucial behavioral foundation: learners stay longer, seek help more readily, and tolerate difficulty with less emotional volatility—conditions that can enable subsequent academic gains.

Adaptive Difficulty Behavior and Learning Gains

This section evaluates whether reinforcement-tuned conversational strategies translate into measurable improvements in learning behavior and task difficulty progression. The conversational agent was designed to dynamically adjust task complexity based on user performance, engagement signals, and observed hesitation. The analysis focuses on observable outcomes: (i) whether learners transitioned into higher-difficulty tasks over time, (ii) whether they completed more instructional steps before abandoning a task, and (iii) whether short post-session quizzes indicated improvement in conceptual mastery.

Two mechanisms supported adaptive progression: challenge elevation and recovery assistance. Challenge elevation increases complexity when a learner consistently answers correctly, encouraging deeper cognitive work. Recovery assistance temporarily lowers difficulty, introduces hints, or switches to explanation-first responses when confusion or slowdown is detected. By analyzing transitions across these states, we can determine whether reinforcement tuning produces smoother learning trajectories and more consistent advancement.

Figure 8 illustrates how users advanced into higher-difficulty tasks across five sequential sessions. The baseline agent shows only modest gains (from 18% in Session 1 to 21% in Session 5) suggesting limited ability to push learners toward increasingly challenging content. In contrast, the RFT model exhibits a steep and consistent upward trajectory, reaching 44% by the final session. This widening separation implies that reinforcement tuning dramatically increases the likelihood that learners are pushed beyond introductory material.

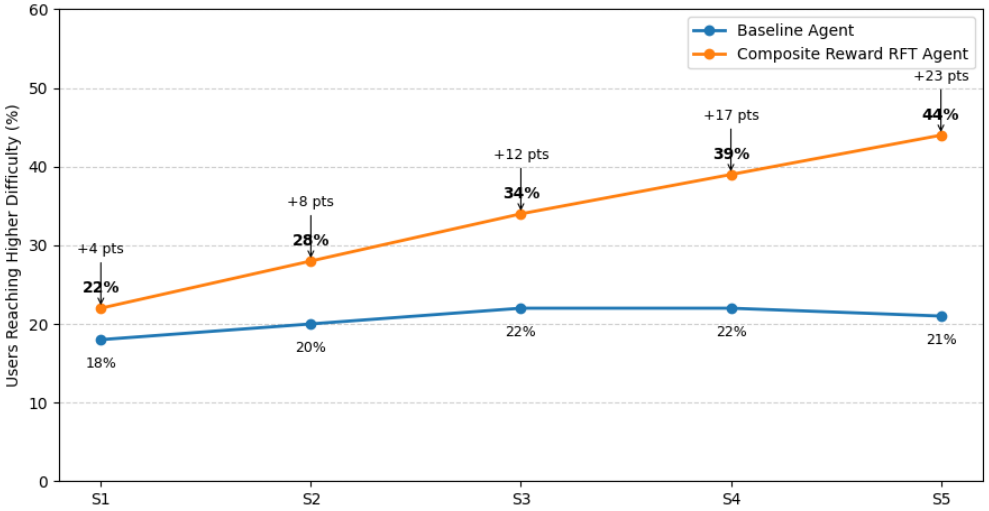

Figure 8 Task Difficulty Transition Across Sessions

The trend is especially pronounced between Sessions 3 and 5, where the gap between baseline and RFT widens from 12% points to over 20. Interviews captured reasons for this shift: several learners described the RFT model as "more confident in challenging me" and "less repetitive," particularly after success on easier questions. Qualitative logs show that the RFT agent frequently introduced chained reasoning or multi-step justification tasks once it detected stable competence. This finding confirms that adaptivity emerges not merely from static prompts, but from policy adjustments that reinforce escalation when the conditions are right.

Conclusion

This research introduced a conversational adaptive learning agent that integrates generative language modeling with reinforcement fine-tuning to optimize instructional responses and align dialogue with learner behavior. The system architecture combines profiling, prompt-controlled generation, reward optimization, and analytic monitoring in a continuous feedback loop. Experimental setup and offline evaluation demonstrated that reinforcement signals—not limited to correctness but extended to engagement and sentiment—produce measurable improvements in pedagogical adequacy, clarity, and safety. Compared to a baseline generative model relying solely on static prompting, the composite-reward policy generated explanations that were more concise, more diagnostic, and better aligned with instructional objectives.

Online deployment further confirmed that reinforcement-tuned adaptivity translates into behaviorally meaningful outcomes. Learners interacting with the composite-reward agent exhibited higher persistence, voluntarily extended sessions at nearly double the rate of baseline participants, and progressed more consistently into multi-step reasoning tasks. Negative affect signals decreased, hint utilization increased, and participants reported that the system felt more responsive to frustration and more intentional about delivering guidance. These behavioral changes supported stronger short-term comprehension, reduced abandonment during assessments, and a higher proportion of high-performing scores on post-session quizzes. In parallel, adaptive difficulty progression

showed a widening gap between baseline and tuned agents across repeated sessions, signaling that dynamic task adjustment—not static scaffolding—encouraged deeper cognitive participation.

Taken together, the results demonstrate that reinforcement learning can serve as a viable mechanism for aligning generative conversational systems with pedagogical intent. Rather than treating dialogue as a surface-level interaction, the model shapes policy according to educational value signals and affective tolerance, resulting in a system that challenges learners without destabilizing motivation. While this study focused on short-cycle interactions, the observed gains in persistence, challenge acceptance, and assessment performance provide a strong basis for longer-term studies on mastery, transfer, and retention. Future work may expand the reward space to include metacognitive markers, longitudinal performance tracking, and integration with curriculum sequencing engines.

Declarations

Author Contributions

Conceptualization: E. and S.P.S.; Methodology: S.P.S.; Software: E.; Validation: E. and S.P.S.; Formal Analysis: E. and S.P.S.; Investigation: E.; Resources: S.P.S.; Data Curation: S.P.S.; Writing Original Draft Preparation: E. and S.P.S.; Writing Review and Editing: S.P.S. and E.; Visualization: E.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] U. Mittal, S. Sai, V. Chamola, and D. Sangwan, “A Comprehensive Review on Generative AI for Education,” *IEEE Access*, vol. 12, no. September, pp. 142733–142759, 2024, doi: 10.1109/ACCESS.2024.3468368.

[2] T. Wang et al., “Exploring the Potential Impact of Artificial Intelligence (AI) on International Students in Higher Education: Generative AI, Chatbots, Analytics, and International Student Success,” *Appl. Sci.*, vol. 13, no. 11, p. 6716, 2023, doi:

10.3390/app13116716.
[3] A. Hanafi, M. Bouhorma, and L. El Aachak, “Machine Learning-Based Augmented Reality For Improved Text Generation Through Recurrent Neural Networks,” *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 2, pp. 518–530, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0 85124453788&partnerID=40&md5=1312e61b857953435fe1df7e6425227f
[4] Z. Jabeen, K. Mishra, R. Dayal, and B. K. Mishra, “Transforming Education in the World of Artificial Intelligence,” *LatIA*, vol. 2, no. January, p. 113, 2024, doi: 10.62486/latia2024113.
[5] M. Kremantzis, A. Essien, E. Pantano, and S. Lythreatis, “Uncovering the Generative AI (GenAI) to Agentic AI (AgAI) Shift for Business School Education,” *J. Glob. Inf. Manag.*, vol. 33, no. 1, pp. 21, 2025, doi: 10.4018/JGIM.389920.
[6] M. Guettala, S. Bourekkache, O. Okba, and S. Harous, “Generative Artificial Intelligence in Ubiquitous Learning: Evaluating a Chatbot-based Recommendation Engine for Personalized and Context-aware Education,” *Acta Inform. Pragensia*, vol. 14, no. 2, pp. 215–245, 2025, doi: 10.18267/j.aip.269.
[7] A. Kostas, G. Koutromanos, and I. Lagopati, “Gamification In Online Adult Learning: A Systematic Literature Review,” *J. Inf. Technol. Educ. Res.*, vol. 24, no. June, pp. 022, 2025, doi: 10.28945/5549.
[8] S. H. Xuan et al., “Evaluating the Impact of Generative AI in Mathematics Education: A Comparative Study in Vietnamese High Schools,” *Hum. Behav. Emerg. Technol.*, vol. 2025, no. 1, pp. 1-23, 2025, doi: 10.1155/hbe2/8886206.
[9] S. Ashtikar, G. Manoharan, and S. Muppidi, “Navigating Education in the Age of Generative AI,” *LatIA*, vol. 3, no. April, p. 327, 2025, doi: 10.62486/latia2025327.
[10] O. Stumke and F. Ndlovu, “Transforming Internal Auditing: Harnessing Retrieval-Augmented Generation Technology*,” Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 4, pp. 785–790, 2025, doi: 10.14569/IJACSA.2025.0160478.
[11] S. Hönigsberg, L. Watkowski, and A. Drechsler, “Generative Artificial Intelligence in Higher Education: Mediating Learning for Literacy Development,” *Commun. Assoc. Inf. Syst.*, vol. 56, no. May, pp. 1044-1076, 2025, doi: 10.17705/1cais.05640.
[12] Q. Lang, M. Wang, M. Yin, S. Liang, and W. Song, “Transforming Education With Generative AI (GAI): Key Insights and Future Prospects,” *IEEE Trans. Learn. Technol.*, vol. 18, pp. 230–242, 2025, doi: 10.1109/TLT.2025.3537618.
[13] C.-W. Liao et al., “AI-Assisted Personalized Learning System for Teaching Chassis Principles,” *Int. J. Eng. Educ.*, vol. 41, no. 2, pp. 548–560, 2025, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105000834109&partnerID=40&md5=4269f48f1a94cf773b9acab81f0a66d0
[14] M. Imran, N. Almusharraf, M. S. Abdellatif, and M. Y. Abbasova, “Artificial Intelligence in Higher Education: Enhancing Learning Systems and Transforming Educational Paradigms,” *Int. J. Interact. Mob. Technol.*, vol. 18, no. 18, pp. 34–48, 2024, doi: 10.3991/ijim.v18i18.49143.
[15] D. Wu, S. Zhang, Z. Ma, X.-G. Yue, and R. K. Dong, “Unlocking Potential: Key Factors Shaping Undergraduate Self-Directed Learning in AI-Enhanced Educational Environments,” *Systems*, vol. 12, no. 9, p. 332, 2024, doi: 10.3390/systems12090332.
[16] S. R. Goldman, J. Taylor, A. Carreon, and S. J. Smith, “Using AI to Support Special Education Teacher Workload,” *J. Spec. Educ. Technol.*, vol. 39, no. 3, pp. 434–447, 2024, doi: 10.1177/01626434241257240.
[17] J. Sardi et al., “How Generative AI Influences Students’ Self-Regulated Learning and Critical Thinking Skills? A Systematic Review,” *Int. J. Eng. Pedagog.*, vol. 15, no. 1, pp. 94–108, 2025, doi: 10.3991/ijep.v15i1.53379.

[18] H.-C. Yeh, “The synergy of generative AI and inquiry-based learning: transforming the landscape of English teaching and learning,” *Interact. Learn. Environ.*, vol. 33, no. 1, pp. 88–102, 2025, doi: 10.1080/10494820.2024.2335491.
[19] G. A. Muhamad, B. S. Alsulami, and K. O. Thabit, “Exploring the Capabilities of GPT Models in Drafting Course Assessments Based on Bloom’s Taxonomy,” *Int. J. Informatics Vis.*, vol. 9, no. 1, pp. 195–200, 2025, doi: 10.62527/joiv.9.1.2811.
[20] I. Levin, A. L. Semenov, and M. Gorsky, “Smart Learning in the 21st Century: Advancing Constructionism Across Three Digital Epochs,” *Educ. Sci.*, vol. 15, no. 1, p. 45, 2025, doi: 10.3390/educsci15010045.
[21] Y. Zhan and L. Chen, “Teaching Case Harness the Power of Interactive Large Language Model in Teaching Using a Capstone Project in the Database Management Course,” *J. Inf. Syst. Educ.*, vol. 36, no. 3, pp. 224–236, 2025, doi: 10.62273/VMJS9132.
[22] M. S. Iqbal, Z. A. Abdul Rahim, and Q. Omerkhel, “Harnessing generative AI for educational innovation: a TRIZ-PLR perspective,” *Discov. Appl. Sci.*, vol. 7, no. 7, p. 789, 2025, doi: 10.1007/s42452-025-07467-3.
[23] P. T. Sheejamol, A. M. Chacko, and S. D. M. Kumar, “Beyond the One-Size-Fits-All: A Systematic Review of Personalized and Gamified e-Learning for Neurodivergent Learners,” *Electron. J. e-Learning*, vol. 23, no. 3, pp. 101–119, 2025, doi: 10.34190/ejel.23.3.4051.
[24] N. S. Farhah, A. Wadood, A. A. AlQarni, M. I. Uddin, and T. H. H. Aldhyani, “Enhancing Adaptive Learning with Generative AI for Tailored Educational Support for Students with Disabilities,” *J. Disabil. Res.*, vol. 4, no. 3, pp. 1-17, 2025, doi: 10.57197/JDR-2025-0012.
[25] J. E. Anderson, C. A. Nguyen, and G. Moreira, “Generative AI-driven personalization of the Community of Inquiry model: enhancing individualized learning experiences in digital classrooms,” *Int. J. Inf. Learn. Technol.*, vol. 42, no. 3, pp. 296–310, 2025, doi: 10.1108/IJILT-10-2024-0240.
[26] D.-L. Chen, K. Aaltonen, H. Lampela, and J. Kujala, “The Design and Implementation of an Educational Chatbot with Personalized Adaptive Learning Features for Project Management Training,” *Technol. Knowl. Learn.*, vol. 30, no. 2, pp. 1047–1072, 2025, doi: 10.1007/s10758-024-09807-5.
[27] X. Lyu, F. Li, and Y. Zhao, “University English Writing Teaching Quality Evaluation Driven by Artificial Intelligence in the Context of New Liberal Arts: Linguistic Neutrosophic Multivalued Approach,” *Neutrosophic Sets Syst.*, vol. 83, no. April, pp. 837–850, 2025, doi: 10.5281/zenodo.15207909.
[28] L. Abrusci, K. Dabaghi, S. D’Urso, and F. Sciarrone, “AI4Design: A generative AI-based system to improve creativity in design–A field evaluation,” *Comput. Educ. Artif. Intell.*, vol. 8, no. June, p. 100401, 2025, doi: 10.1016/j.caeai.2025.100401.
[29] D. Liu, H. Zhao, W. Tang, and W. Yang, “AIKII: An AI-Enhanced Knowledge Interactive Interface for Knowledge Representation in Educational Games,” *Comput. Animat. Virtual Worlds*, vol. 36, no. 3, p. e70052, 2025, doi: 10.1002/cav.70052.
[30] W. Hu and Z. Shao, “Design and evaluation of a GenAI-based personalized educational content system tailored to personality traits and emotional responses for adaptive learning,” *Comput. Hum. Behav. Reports*, vol. 19, no. August, p. 100735, 2025, doi: 10.1016/j.chbr.2025.100735.
[31] F. Abbes, S. Bennani, and A. Maalel, “A generative AI model for enhancing learner’s skills in a gamified learning environment,” *Cluster Comput.*, vol. 28, no. 15, p. 948, 2025, doi: 10.1007/s10586-025-05505-8.
[32] G. Cooper, K.-S. Tang, and A. Fitzgerald, “Intersections of Mind and Machine: Navigating the Nexus of Artificial Intelligence, Science Education, and the

Preparation of Pre-service Teachers," *J. Sci. Educ. Technol.*, vol. 34, no. 6, pp. 1255–1259, 2025, doi: 10.1007/s10956-025-10200-9.

[33] S. Feng, H. Zhang, and D. Gašević, "Mapping the evolution of AI in education: Toward a co-adaptive and human-centered paradigm," *Comput. Educ. Artif. Intell.*, vol. 9, no. December, p. 100513, 2025, doi: 10.1016/j.caeai.2025.100513.

[34] C. V Obionwu, S. S. Bedi, A. V. C. Bhagavathi, D. S. Walia, and K. Turowski, "An Expert in the Loop Strategy for Generating Synthetic Learning Engagement Datasets," *SN Comput. Sci.*, vol. 6, no. 8, p. 959, 2025, doi: 10.1007/s42979-025-04457-5.

[35] H. M. Qadir, M. T. Suleman, R. A. Khan, M. Sohaib, M. J. Hasan, and S. A. Hussain, "Optimizing learning outcomes: a deep dive into hybrid AI models for adaptive educational feedback," *J. Big Data*, vol. 12, no. 1, p. 144, 2025, doi: 10.1186/s40537-025-01187-6.

[36] R. Zhou, Y. Liu, L. Sun, and S. Zhu, "Enhancing Operating System Education With a Generative AI-Supported Boppps Model: An Empirical Study," *Comput. Appl. Eng. Educ.*, vol. 33, no. 6, p. e70114, 2025, doi: 10.1002/cae.70114.

[37] S. Cha, J. Lee, and R. Zapata, "A Systematic Review of Interactivity in Game-Based Learning with AI," *Proc. Assoc. Inf. Sci. Technol.*, vol. 62, no. 1, pp. 1379–1382, 2025, doi: 10.1002/pra2.1410.

[38] A. M. Khaddar, A. Dehbi, A. Alali, A. Bakhouyi, and M. Talea, "Emerging Technologies in Learning: A Bibliometric Analysis of Technology Integration and Applications," *Int. J. Interact. Mob. Technol.*, vol. 19, no. 5, pp. 60–78, 2025, doi: 10.3991/ijim.v19i05.51267.

[39] I. Întorsureanu, S.-V. Oprea, A. Bâra, and D. Vespan, "Generative AI in Education: Perspectives Through an Academic Lens," *Electron.*, vol. 14, no. 5, p. 1053, 2025, doi: 10.3390/electronics14051053.

[40] P. Mishra, D. Henriksen, L. J. Woo, and N. Oster, "Control vs. Agency: Exploring the History of AI in Education," *TechTrends*, vol. 69, no. 2, pp. 247–253, 2025, doi: 10.1007/s11528-025-01064-2.

[41] C. Troussas, A. Krouska, P. Mylonas, C. Sgouropoulou, and I. Voyiatzis, "Fuzzy Memory Networks and Contextual Schemas: Enhancing ChatGPT Responses in a Personalized Educational System," *Computers*, vol. 14, no. 3, p. 89, 2025, doi: 10.3390/computers14030089.

[42] J. Chun et al., "A Comparative Analysis of On-Device AI-Driven, Self-Regulated Learning and Traditional Pedagogy in University Health Sciences Education," *Appl. Sci.*, vol. 15, no. 4, p. 1815, 2025, doi: 10.3390/app15041815.

[43] M. Barbu, D.-D. Iordache, I. Petre, D.-C. Barbu, and L. Bajenaru, "Framework Design for Reinforcing the Potential of XR Technologies in Transforming Inclusive Education," *Appl. Sci.*, vol. 15, no. 3, p. 1484, 2025, doi: 10.3390/app15031484.