PersonaGPT: A Context-Aware Personalization Engine for Educational Chatbots Using Dynamic Learner Personas and Reflexive Dialogue

Christos Troussas, Akrivi Krouska, Phivos Mylonas, Cleo Sgouropoulou
Department of Informatics and Computer Engineering
University of West Attica
Egaleo, Greece
{ctrouss, akrouska, mylonasf, csgouro}@uniwa.gr

***Abstract*—Artificial intelligence has transformed educational technology, with conversational agents becoming increasingly prominent in supporting student learning. While these chatbots offer accessibility and scalability, many still lack the capacity to personalize interactions based on the learner's evolving cognitive and emotional state. Current systems often rely on static profiles or shallow adaptation mechanisms that fail to account for real-time fluctuations in confidence, engagement, and comprehension. This paper presents PersonaGPT, a novel personalization engine for educational chatbots that constructs and continuously updates dynamic learner personas using behavioral signals, linguistic cues, and meta-dialogue patterns. The system integrates these personas into a reflexive dialogue engine, enabling the chatbot to adapt its tone, explanation depth, and instructional strategy in real time. PersonaGPT was implemented as a modular, web-based tutoring system using GPT-4, a Flask backend, and React.js interface, designed to teach SQL programming. In a controlled experiment with 81 university students, PersonaGPT significantly outperformed static and non-personalized chatbots in task completion, knowledge retention, and learner engagement. The study demonstrates that real-time persona modeling and reflexive dialogue enhance both the effectiveness and user experience of AI-powered education.**

Keywords—Adaptive Educational Chatbots; Learner Modeling; Personalized Learning Systems; Reflexive Dialogue in AI; Human-Centered Artificial Intelligence

I. INTRODUCTION

Over the past decade, artificial intelligence (AI) has begun to significantly influence the landscape of educational technology, transforming the way in which learners can gain access to information, receive feedback, and interact with educational materials [1, 2]. Perhaps the most notable phenomenon has been the use of AI-based conversational agents (also referred to as educational chatbots) that provide students with the capability for natural language conversation while doing academic work [3,4]. These systems are particularly attractive because they may be accessed, scaled, and replicated at low cost to simulate personalized tutoring. They have the potential to be applied in numerous contexts, including elementary school through university, and across mathematics, language teaching, and computer science; AI chatbots have been discovered to be able to offer step-by-step guidance, immediate corrective feedback, and sustain learner interest in self-paced environments [5]. Their creation has also been accelerated by the advancement of large language models (LLMs), which now enable chatbots to handle open-ended questions, understand subtle student statements and even generate context-dependent learning feedback [6]. This innovation is a significant advancement in human-computer interaction within learning settings, particularly with the rising need for flexible, remote, and asynchronous learning support.

Despite the functional and technological promise of these systems, their pedagogical effectiveness is not solely determined by their ability to answer questions or deliver instructional content [7]. Rather, the quality of learner experience – especially in terms of engagement, trust, and sustained interaction – hinges on the chatbot's capacity to personalize its responses to the individual learner's needs and state. Personalization in educational contexts encompasses a range of adaptive behaviors, including the ability to adjust content difficulty, modify the mode of explanation, adapt the pacing of instruction, and shift the communicative tone in response to learners' affective and cognitive cues [8, 9]. However, many existing chatbots rely on surface-level personalization strategies, such as allowing users to select their preferred learning style or adapting based on a fixed pre-session profile [10-12]. These approaches often assume that learners are static entities whose preferences or knowledge states remain constant across time. Yet educational research has shown that learners' cognitive load, motivation, and confidence can fluctuate significantly within a single learning session, influenced by success or failure on a task, emotional response to feedback, or evolving interest in the content [13, 14]. As such, personalization that does not take into account these real-time dynamics risks being ineffective or, worse, counterproductive – misaligning the support provided with the learner's actual state and thereby reducing learning efficacy.

What is currently missing from most educational chatbot architectures is a mechanism for continuously sensing and adapting to the learner's evolving behavioral and emotional context [15-17]. While some systems incorporate basic analytics or surface-level interaction tracking, few are capable of constructing a nuanced, dynamic model of the learner that integrates both performance data and linguistic indicators of affect or engagement [18]. Moreover, even fewer systems are capable of using such a model to modulate their dialogue behavior in ways that are pedagogically responsive and emotionally intelligent [19]. This gap is critical because learner trust, willingness to persist through difficulty, and satisfaction with the learning experience are deeply influenced by the perceived responsiveness of the instructional agent. When learners encounter static or generic responses that fail to reflect their current struggle or progress, the educational experience can feel impersonal and detached – more like interacting with a FAQ database than engaging with a tutor. Furthermore, learners' trust in AI systems is known to be fragile and context-dependent; when a chatbot fails to acknowledge confusion or misinterprets a learner's intent, the resulting breakdown in interaction can lead to disengagement or even rejection of the system [20]. There is thus a pressing need for AI systems in education to become more reflexive – able not just to respond but to self-adjust, rephrase, and offer alternatives based on the learner's trajectory of understanding and affect. This need points to a broader research gap in the field of AI in education: the lack of scalable systems that integrate real-time learner modeling with adaptive dialogue grounded in both pedagogy and human-centered communication principles.

The literature on adaptive learning systems has explored various strategies to address personalization, ranging from rule-based engines to probabilistic modeling frameworks [21-30]. For example, Bayesian Knowledge Tracing (BKT) and its successors have been widely used to model learner knowledge states across time, estimating the probability that a learner has mastered a particular concept based on task performance. While effective for knowledge tracking, BKT does not account for emotional state, engagement, or metacognitive reflection. More recent approaches, such as Deep Knowledge Tracing (DKT), leverage recurrent neural networks to track student learning trajectories, offering more predictive power but at the cost of reduced interpretability [31]. Works on Fuzzy Cognitive Maps (FCMs) [32] and Human Plausible Reasoning (HPR) [23] has sought to model learner reasoning processes and behavior patterns in a more symbolic and interpretable manner, enabling explicit representation of misconceptions and learning pathways. These systems provide a rich foundation for personalization but often lack integration with conversational interaction or are too rigid to adapt to the fluidity of natural dialogue. In the domain of educational chatbots, most implementations still rely on scripted dialogues, keyword matching, or limited dialogue state management. Even those systems that leverage LLMs such as GPT-3 or GPT-4 often do so in a stateless or context-light fashion, limiting the system's ability to "remember" the learner's prior difficulties, preferred communication style, or evolving affective state [3, 5]. Moreover, very few systems employ reflexive dialogue strategies – utterances where the system acknowledges miscommunication, adjusts its tone, or offers alternative forms of explanation to rebuild trust and enhance rapport [33-35]. The absence of such reflexivity represents a major gap in current conversational AI designs for education, especially given growing awareness of the importance of socio-emotional dimensions in learning.

In response to this critical need, this paper presents PersonaGPT, a novel personalization engine for educational chatbots that integrates real-time learner persona modeling with reflexive dialogue capabilities. At the core of PersonaGPT is a dynamic learner modeling pipeline that constructs a continuously evolving persona based on multiple input streams: behavioral indicators such as task completion success and error patterns; linguistic signals such as expressions of uncertainty, confusion, or confidence; and meta-dialogue behaviors such as requests for re-explanation or clarification. This persona is not a static profile but a moment-to-moment representation of the learner's cognitive, affective, and engagement state. It is used to guide response generation through prompt conditioning, enabling the chatbot to adjust not only the content and complexity of its explanations but also its communicative tone, offering empathy, encouragement, or challenge as appropriate. In addition, the system incorporates reflexive dialogue mechanisms designed to repair interaction breakdowns, acknowledge learner frustration, and offer metacognitive scaffolding. For example, if a learner repeatedly struggles with a concept, the chatbot might say: "Let's look at this in a different way" or "No worries, many students find this tricky at first". These interventions are informed by research on trust-building in AI and pedagogical best practices for fostering learner persistence.

PersonaGPT is implemented as a lightweight, modular system capable of real-world deployment in educational settings. The architecture consists of a React.js frontend for chat interaction, a Flask-based backend for dialogue management, and integration with OpenAI's GPT-4 API for response generation. A persona inference engine aggregates input signals into a vector representation of learner traits – such as cognitive load, engagement, and confidence – which is used to condition GPT-4 prompts in real time. The system supports both short-term personalization (during a session) and session-based memory for extended learning. To evaluate its effectiveness, a two-week experimental study was conducted involving 81 undergraduate students learning SQL through chatbot-mediated interaction. Participants were randomly assigned to one of three conditions: PersonaGPT, a static-profile chatbot, or a generic non-personalized chatbot. All participants completed the same SQL tasks, and their performance, engagement, and survey feedback were analyzed. The results showed that learners in the PersonaGPT condition had significantly higher task success rates, better retention of material, and more positive perceptions of adaptivity and engagement. Moreover, open-ended feedback suggested that learners felt "seen" by the chatbot and appreciated its ability to adjust explanations and tone dynamically. These findings contribute to the growing body of evidence that personalization in educational AI must go beyond content adaptation to include real-time, responsive dialogue strategies grounded in learner context.

II. System Architecture and Implementation

The architectural design of PersonaGPT embodies an integrated approach to adaptive AI tutoring, where real-time learner modeling, natural language interaction, and reflexive personalization converge into a scalable and deployable chatbot system. The novelty of the system lies in its ability to construct dynamic learner personas using observable signals, such as performance data, linguistic cues, and interaction behaviors, and to generate instructional dialogue that adapts not only pedagogically but also emotionally in response to these evolving personas.

This section offers a detailed exposition of the system's design principles, key components, adaptive mechanisms, and implementation technologies, concluding with its ethical safeguards and real-time personalization cycle. A visual representation of the system's logical structure is provided in Figure 1.

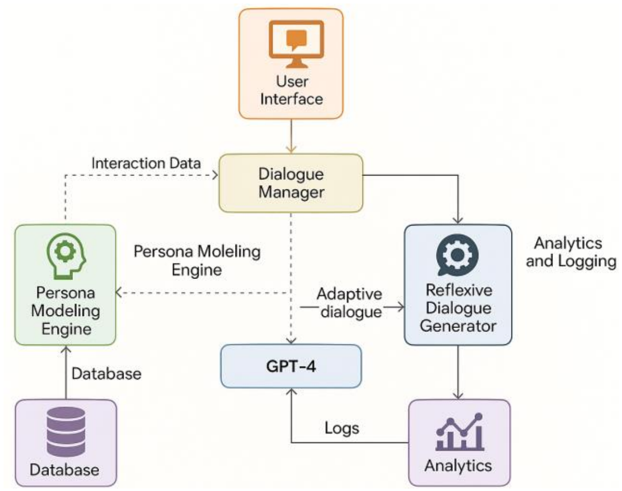

Fig. 1. Logical Architecture of the PersonaGPT system.

From the outset, the development of PersonaGPT was governed by five core design principles. First, the system must furnish context-aware personalization not emanating from static learner profiles but from data continually updated to reflect fluctuating learner confidence, engagement, and cognitive load. Second, the architecture was to be modular and scalable so as to allow independently deployable components to be used by multiple concurrent users. Third, there was to be reflexivity in the conversations so that the system could say it is confused, rephrase instructions, or even attempt to build trust—something very much needed in human tutoring. Fourth, there was to be pedagogical alignment: the chatbot would adapt its mode of discourse and instructional strategies in response to the learner's performance and task history. Fifth, an ethical and transparent system had to be put in place so that explainable personalization could be offered and data handling could be made GDPR compliant.

PersonaGPT follows a layered and service-oriented architecture. At the user interface level, learners interact with a Frontend UI module. This interface allows learners to engage in natural conversation, pose questions, and solve tasks such as SQL challenges. It handles message exchange, streaming of responses, and also logs implicit behavioral signals (e.g., typing pauses, input length, and rephrasing frequency).

The Dialogue Manager, implemented in Flask (Python), forms the operational core. It coordinates the interaction loop: managing dialogue states, interpreting intents, updating learner progress, and orchestrating requests to internal services. It also determines when to switch between instructional modes (e.g., problem-solving, explanation, clarification) and when to trigger personalization updates. This manager acts as the conductor of all data flow between system components.

One of its key responsibilities is invoking the Persona Modeling Engine, which continuously updates a learner model – called the persona – based on multi-source input:

- Task performance: error types (syntax vs. logic), number of retries, time-on-task, success rate.
- Behavioral signals: help requests, long pauses, skipped tasks.
- Linguistic analysis: expressions of confusion ("I don't get it"), confidence ("That was easy"), or hesitation ("maybe…"), extracted using spaCy and VADER.
- Interaction preferences: learner reactions to analogies, definitions, or examples are used to infer optimal explanation format.

These signals form a real-time vector representation of the learner's state. The four primary traits tracked are:

- Cognitive Load (CL) – mental effort required,
- Confidence (CONF) – emotional readiness and self-efficacy,
- Engagement (ENG) – level of interaction, participation, and enthusiasm,
- Explanation Preference (XPREF) – inferred based on format responsiveness.

Each trait is updated using a time-decay model:

$$\text{trait}_t = \alpha \cdot \text{signal}_{\text{current}} + (1 - \alpha) \cdot \text{trait}_{t-1}$$

with α typically set at 0.6, striking a balance between sensitivity and stability. The α value (0.6) was chosen based on empirical tuning to balance sensitivity to new input with memory of prior behavior.

This persona vector is directly fed into the Reflexive Dialogue Generator, which uses a technique known as prompt conditioning – the process of embedding learner-specific traits into the input prompt of the language model – to guide its responses. By conditioning the prompt with real-time updates on cognitive load, confidence, and engagement, the system ensures that generated replies are both pedagogically appropriate and emotionally attuned. Instead of fixed system prompts, GPT-4 receives personalized directives such as:

"You are a friendly tutor. The learner is showing signs of low confidence and high cognitive load. Use supportive tone, simpler language, and step-by-step examples."

This process – called prompt conditioning – ensures that responses are not only content-appropriate but also emotionally adaptive. These prompt descriptors are constructed dynamically at each learner turn, based on the current state of the persona vector (e.g., confidence, cognitive load, explanation preference). They are embedded directly into the GPT-4 prompt sent via the OpenAI API to influence the style, depth, and tone of the response. The following example illustrates a typical prompt constructed by the system:

- System prompt:

"You are a supportive and adaptive SQL tutor. The learner currently shows low confidence, high cognitive load, and prefers analogical explanations. Use simple language, short sentences, and include analogies where appropriate."

- User input:

"What's the difference between INNER JOIN and LEFT JOIN?"

In cases of misunderstanding, fatigue, or disengagement, PersonaGPT deploys reflexive dialogue strategies. These include metacognitive interventions such as:

- "Let's take another look – maybe in a different way."
- "No worries – this is a tricky topic for many learners."
- "Would you like an example, a hint, or a simpler explanation?"

These reflexive utterances are generated when the Dialogue Manager detects risk indicators (e.g., repeated errors or affective signals) and seeks to rebuild trust and motivation.

Once generated, responses are returned to the User Interface and simultaneously logged. Every interaction triggers updates not just in the persona but in session metadata. These updates are stored in the Database Layer, which uses PostgreSQL for persistent session storage and Redis for rapid access to persona vectors and task records. This storage layer also enables session resumption and longitudinal tracking across multiple learning events.

Parallel to this, the system feeds data into the Analytics and Logging Module. This component records interaction histories, including timestamps, trait evolution, and dialogue outcomes. These logs are used for research analytics, debugging, future training sets, and instructor-facing dashboards in extended versions. Logs are anonymized and encrypted to preserve learner privacy and comply with institutional data handling policies.

Working as an asynchronous layer between the system and GPT-4, the API layer for Natural Language ensures a modularity that separates LLM access: institutions may replace GPT-4 for other providers or custom private models in a future deployment. Error-checking, retry policies, and rate-limiting mechanisms are also in place within the API layer for system stability and scalability.

Taken together, all these modules form a closed-loop adaptive learning cycle:

1. The learner submits input via the UI.
2. The Dialogue Manager interprets the query and calls for updates to the persona.
3. The updated persona is used to condition GPT-4's system prompt.
4. The Reflexive Dialogue Generator creates the final output.
5. The response is streamed back to the learner and logged.
6. Analytics and storage systems capture new signals, restarting the loop.

This cycle allows PersonaGPT to adapt at the granular level of each turn, rather than at the lesson or unit level, fostering a highly personalized dialogue experience.

To ensure robustness and learner autonomy, PersonaGPT incorporates several fallback strategies. When learner intent is ambiguous, the bot asks clarifying questions. When model confidence is low (e.g., due to missing persona data or anomalous inputs), it defaults to safe explanations. And when learners explicitly ask for help, the system offers flexible options such as hints, scaffolds, or simplified analogies – depending on both context and persona state.

From an engineering perspective, the architecture is designed for containerized deployment. All components (UI, API, dialogue manager, persona engine) are dockerized and deployable on platforms like Heroku or AWS. OAuth2 authentication is used to enable LMS integration and user account control, and the system is prepared for institutional pilots at scale.

Crucially, PersonaGPT is built with ethical AI design in mind. Users are informed about data logging. All session data is anonymized. Adaptive behaviors are sometimes explained with transparency utterances like:

"I'm offering an example here because you seemed to prefer it in previous questions."

These design decisions are vital to building trust with learners and meeting the growing demand for AI accountability in education.

In conclusion, the architecture of PersonaGPT is not merely a technical configuration – it is the material realization of a pedagogical philosophy: that personalization must be continuous, interpretable, and respectful. By integrating real-time persona modeling, reflexive conversational strategies, and modular cloud-based components, the system presents a robust, scalable, and human-centered model for next-generation educational chatbots.

III. EXPERIMENTAL DESIGN

To rigorously evaluate the pedagogical effectiveness of PersonaGPT as a personalization engine for educational chatbots, we conducted a mixed-method experimental study embedded within a real-world university environment. The goal was to investigate how dynamically generated learner personas and reflexive dialogue strategies affect measurable learning outcomes, learner engagement, and the overall perception of personalization in AI-mediated tutoring systems. By combining quantitative and qualitative methods in a carefully structured design, this study provides a strong empirical foundation to assess the impact of adaptive conversational AI on learning performance in an authentic instructional context.

The study was structured around three key research questions. First, we examined whether the use of PersonaGPT led to improvements in learners' task performance when compared to alternative chatbot configurations – specifically those using static profiles or offering no personalization at all. Second, we explored whether the system enhanced learners' retention of the material, measured through delayed post-task assessments. Third, we aimed to understand how learners perceived the chatbot's adaptivity, trustworthiness, and overall usefulness in supporting their learning journey. From these questions, we derived three testable hypotheses: that learners supported by PersonaGPT would complete more learning tasks correctly, retain more of the material after a delay, and report greater satisfaction and perceived personalization than learners in the comparison conditions.

Participants were drawn from an undergraduate population enrolled in introductory database courses at a public university. Recruitment was conducted through in-class announcements and follow-up emails. Eligibility was limited to students who had not previously taken advanced SQL coursework, ensuring that all participants had roughly equal levels of prior knowledge. Participation was voluntary, with a minor academic incentive – 2% bonus credit – offered for full completion of the study protocol. A total of 81 students participated, and they were randomly assigned into three equal-sized groups of 27 using a simple Python-based random assignment script. This ensured group balance while minimizing selection bias and supporting internal validity for statistical comparison.

Each group was exposed to a distinct version of the chatbot system. Group A interacted with PersonaGPT, which featured dynamic learner modeling and real-time reflexive dialogue strategies. Group B used a chatbot configured with a static profile: at the beginning of their session, participants selected their preferred explanation format (e.g., example-based, definition-based, visual/textual) and rated their confidence on a five-point scale. These initial selections guided the chatbot's behavior throughout the session, but the system did not update or adapt its strategy in real time. Group C served as the control group, interacting with a generic chatbot that delivered consistent responses regardless of learner behavior, performance, or expressed preferences. This configuration allowed us to isolate the effects of real-time personalization and adaptivity from those of baseline interactivity or interface design.

All groups interacted with the chatbot in the same interface environment. The frontend UI, developed in React.js, was identical across conditions to control for layout, usability, and aesthetic influence. The primary instructional task focused on a progression of five SQL programming problems, each increasing in difficulty and designed to assess the learner's grasp of key relational database concepts. These tasks were crafted in collaboration with course instructors to ensure alignment with curricular objectives. The problems required learners to construct SQL queries involving INNER JOINs, LEFT JOINs with NULL filtering, GROUP BY clauses with aggregations, JOIN conditions with WHERE clauses, and nested SELECT queries. Participants completed the tasks independently in a supervised computer lab setting. They had up to 40 minutes to engage with the chatbot, complete the tasks, and provide feedback, though many completed the session earlier.

The PersonaGPT condition featured the most sophisticated backend logic. The chatbot responses were conditioned on the learner's evolving persona, which was updated in real time based on behavioral and linguistic input. The system continuously interpreted performance metrics – such as task accuracy, number of retries, and completion time – as well as linguistic indicators like expressions of confusion or confidence. These signals were synthesized into a dynamic learner profile that guided the depth of explanation, tone, and response presentation format of chatbot. Some reflexive dialog strategies were implemented as well: if signs of struggle or disengagement were detected, the system would intervene with empathetic utterances and rephrasings, as well as opportunities for metacognitive reflection. Whereas in the static-profile group, responses aligned with the learner's initial self-declared preferences but no further update to personalization was applied during the session. The control group was given fixed explanations regardless of context or learner input.

To evaluate the effects of this prototype, we used a variety of data collection instruments. The first, which was automatically collected by the system, was a set of objective performance measures. These were: how many tasks the learner successfully completed, total duration (time spent on the task), the number of help requests made, and the number of attempts that were made with corrections before arriving at a valid solution. Each request was timestamped, and included in PostgreSQL, which also allowed for cross-temporal and behavioral analysis. Requesting help would consist of either measurable language (e.g., "I need help", "I don't get this") or requesting help via help request buttons embedded in the interface. Correction attempts were based on the number of distinct submissions that preceded the final correct answer. This would give us a clear understanding of the learner's trial and error activity, or confusion.

To measure learning retention, participants completed a delayed post-test 24 hours after their chatbot interaction. This assessment was administered via the university's learning management system and included five multiple-choice questions and three open-ended questions, covering the SQL topics encountered during the main task. Scores were normalized to a 10-point scale. This retention test allowed us to go beyond immediate task performance and evaluate whether the chatbot's support strategies led to lasting learning gains.

Participants also completed a post-session survey that captured their subjective experiences. The instrument included items grouped into three key subscales: perceived personalization, engagement, and trust/satisfaction. Statements such as “The chatbot adapted its explanations to my needs,” “The interaction kept me interested and involved,” and “I felt the chatbot understood my difficulties” were rated on a five-point Likert scale ranging from strongly disagree to strongly agree. The survey was adapted from validated instruments in prior HCI and educational technology research. To ensure reliability, internal consistency was measured using Cronbach’s alpha, which returned a value of 0.86 – indicating strong cohesion among items.

In addition to closed-ended items, the survey included open-ended prompts. Learners were asked what they liked most about the chatbot, whether anything frustrated them during the session, and how the chatbot could be improved. These qualitative responses were later subjected to thematic analysis using an inductive coding approach. Two independent researchers reviewed the responses, created initial codes, and refined these into themes through iterative comparison. Interrater reliability exceeded 85%, indicating a high level of agreement and thematic coherence.

All participants followed the same sequence of steps during the session. After signing a consent form and completing a brief demographic questionnaire, learners engaged in a short practice round designed to familiarize them with the chatbot interface and the types of SQL questions they would encounter. This warm-up lasted about five minutes. They then entered the main session, which was limited to a maximum of 40 minutes. Upon completion, learners filled out the experience survey and submitted open-ended feedback. The delayed retention test was emailed to them the next day with a 24-hour completion window.

To preserve the study’s internal validity, we implemented a single-blind protocol. Participants were unaware of the different versions of the chatbot and were not informed of the study’s hypotheses. This helped to control for expectancy effects and social desirability bias. The research team took steps to ensure that all learners received similar conditions in terms of lab setup, task instructions, and availability of technical support. All tasks and explanations were delivered in English, and no participant received personalized assistance from human facilitators during the task phase.

Ensuring methodological robustness required attention to multiple forms of validity and reliability. Construct validity was supported by the triangulation of performance metrics, survey responses, and qualitative feedback. Internal validity was reinforced by random group assignment and by standardizing the interface and instructional materials across all conditions. Ecological validity was addressed through the use of real course content, authentic SQL tasks, and a deployment environment (i.e., a university computer lab) that mirrored the setting in which such a system might be deployed in practice. Reliability was tested both quantitatively (e.g., survey alpha scores) and qualitatively (e.g., interrater agreement in feedback coding).

For the study, participants were informed of their rights, including the right to withdraw at any point without penalty. No personally identifying information was recorded; instead, each learner was assigned a random user ID. All chatbot logs and performance data were stored securely on encrypted servers and used only for research purposes. To prevent the risk of overreliance on AI, the chatbot was configured to encourage independent problem-solving. For example, when a learner asked for a direct answer, the chatbot might instead respond with, “Let’s think through this together. What do you think would happen if we tried this JOIN instead?” A full debriefing followed the study, where participants were informed of the different conditions and the study goals.

IV. Results and Analysis

The findings of the experimental evaluation of PersonaGPT offer substantial empirical evidence to substantiate the effectiveness of adaptive learner personas and reflexive dialogue for improving learning in mediated chatbot environments. This section includes a discussion about the study's findings, including quantitative task performance, retention, engagement, perceived personalization, behavioral measures, and qualitative learner feedback. The data or evidenced important differences among the three experimental groups (PersonaGPT, Static Chatbot, and Control) and we offer a multi-dimensional view of the system's pedagogical impact.

The first and primary learning metric evaluated was task completion rate, defined as the number correctly submitted SQL queries over the five tasks given to each participant. As shown in Table I, learners in the PersonaGPT condition achieved a mean task completion rate of 92.3%, which was significantly higher than the Static Chatbot condition (83.7%) and the Control condition (76.4%). Standard deviations were lowest in the PersonaGPT group, suggesting more consistent performance. A one-way ANOVA confirmed a statistically significant difference in means across groups ($F(2,78) = 16.14$, $p < 0.001$). Post hoc Tukey tests showed that the differences between PersonaGPT and each of the other two groups were significant at $p < 0.01$, while the Static vs. Control comparison yielded a smaller yet significant difference ($p = 0.045$). These findings provide strong support for H1 and reinforce the conclusion that real-time adaptation contributes to task success.

TABLE I. Task Completion Rates by Group

	Mean (%)	**Std. Dev.**
PersonaGPT	92.3	5.2
Static Chatbot	83.7	7.9
Control (No Personalization)	76.4	8.6

To evaluate knowledge retention beyond immediate task performance, a follow-up assessment was administered 24 hours after the main chatbot session. This retention test, scored out of 10 points, included both multiple-choice and open-ended questions covering the SQL concepts practiced in-session. As presented in Table II, the PersonaGPT group scored highest ($M = 8.4$), followed by the Static Chatbot group ($M = 7.1$) and the Control group ($M = 6.3$). Differences among the groups were significant ($F(2,78) = 13.57$, $p < 0.001$), with post hoc comparisons confirming that learners in the PersonaGPT condition retained significantly more knowledge ($p < 0.01$ vs. both groups). These results confirm H2 and demonstrate that the adaptive and reflective support provided by PersonaGPT was not only effective in-session but also contributed to longer-term understanding of SQL concepts.

TABLE II. RETENTION TEST SCORES BY GROUP

	Mean Score (out of 10)	Std. Dev.
PersonaGPT	8.4	1.1
Static Chatbot	7.1	1.3
Control	6.3	1.6

Session duration was also analyzed to understand how long learners remained engaged with the chatbot. Although the session was capped at 40 minutes, participants could finish early. As shown in Table III, the average session time was highest for the PersonaGPT group (M = 21.5 minutes), followed by Static Chatbot (18.9) and Control (17.8). While longer duration alone may not be desirable, in this context it was associated with greater depth of interaction and aligned with higher task completion and retention. Informal observation suggested that PersonaGPT learners stayed engaged due to the system's responsive explanations and adaptive pacing.

TABLE III. AVERAGE SESSION DURATION

	Mean Time (min)	Std. Dev.
PersonaGPT	21.5	3.2
Static Chatbot	18.9	2.7
Control	17.8	2.9

In terms of subjective experience, participants completed a post-task survey measuring perceived personalization and engagement using a 5-point Likert scale. As shown in Table IV, the PersonaGPT group reported significantly higher levels of personalization (M = 4.6) and engagement (M = 4.5) than either comparison group. Kruskal-Wallis tests indicated significant differences ($p < 0.001$), and Mann-Whitney U tests confirmed that all pairwise comparisons involving PersonaGPT were significant at $p < 0.01$. These results provide strong support for H3 and confirm that learners not only performed better but also felt more supported and involved.

TABLE IV. LEARNER EXPERIENCE RATINGS

	PersonaGPT	Static Chatbot	Control
Perceived Personalization	4.6	3.7	2.9
Engagement	4.5	3.6	3.1

Behavioral indicators further corroborated these findings. One of the clearest signals was the number of explicit help requests submitted during the session. These requests were either typed manually (e.g., "Can I get a hint?") or initiated through the help button. As seen in Table V, learners using PersonaGPT requested help less frequently (M = 2.1) than those using the Static Chatbot (M = 3.8) or the Control chatbot (M = 4.2). This suggests that adaptive scaffolding and reflexive clarification may have preemptively addressed learner confusion, thereby increasing confidence and reducing dependency.

TABLE V. HELP REQUESTS DURING SESSION

	Mean Requests	Std. Dev.
PersonaGPT	2.1	1.4
Static Chatbot	3.8	2.0
Control	4.2	2.3

A second behavioral indicator was the number of correction attempts per task, defined as the number of incorrect submissions prior to a correct solution. Table VI reveals that PersonaGPT users averaged only 1.7 correction attempts, versus 2.5 in the Static Chatbot group and 2.9 in the Control group. This suggests that PersonaGPT learners benefited from clearer guidance and better internalized task expectations, leading to fewer errors and more efficient learning.

TABLE VI. ERROR CORRECTION ATTEMPTS

	Mean Attempts	Std. Dev.
PersonaGPT	1.7	1.1
Static Chatbot	2.5	1.4
Control	2.9	1.6

Learners in the Control and Static groups often submitted queries with incorrect JOIN conditions or failed to handle NULLs properly. In contrast, PersonaGPT users made fewer conceptual errors, likely due to timely clarification and targeted examples.

These quantitative patterns – derived from task performance, retention scores, session duration, learner experience ratings, and behavioral indicators – are visually summarized in Figure 2. The radar plot presents a normalized comparison across all major metrics and illustrates the multidimensional effectiveness of PersonaGPT. Compared to the Static Chatbot and Control groups, PersonaGPT consistently achieves higher values across performance and engagement dimensions, and lower values on inverse metrics such as help requests and error correction attempts (reflected as higher normalized scores).

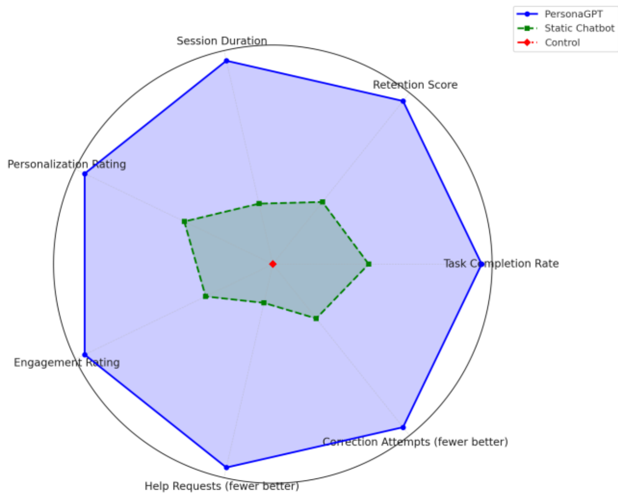

Fig. 2. Comparative Performance and Experience Metrics

The visual representation reinforces the conclusion that PersonaGPT's adaptive and reflexive dialogue mechanisms not only enhanced cognitive outcomes but also promoted a more engaging and efficient learning experience.

Quantitative results were complemented by qualitative thematic analysis of learners' open-ended feedback. Several recurring themes emerged across conditions. In the PersonaGPT group, many learners described the system as "understanding," "encouraging," and "like a real tutor." Representative comments included: "It felt like the bot knew when I was confused – it rephrased without me even asking," and "It encouraged me when I got stuck, like a personal coach." These responses confirm that learners not only noticed the system's adaptivity but appreciated its relational qualities.

By contrast, feedback from Static Chatbot users was more ambivalent. While some learners appreciated the initial customization ("I liked choosing my preferences"), others noted the system's rigidity ("It didn't change much even after I made mistakes"). This suggests that static profiling is perceived as less responsive and may result in plateaued engagement. Control group learners were most likely to express frustration or boredom, with comments such as "It was too repetitive," or "It didn't really respond to me." These reflections underscore the importance of real-time adaptation in maintaining learner motivation.

Interestingly, a few participants in the PersonaGPT group mentioned instances of overpersonalization, such as "It sometimes overexplained things I already knew." This highlights the challenge of striking the right balance in adaptive support – ensuring responsiveness without redundancy. Nevertheless, such cases were rare and did not diminish the overall positive perception of the system.

When we examine the results holistically, several important patterns emerge. First, all three hypotheses (H1–H3) are supported by statistically significant and triangulated evidence: PersonaGPT improved task performance, retention, and learner experience. Second, the system's gains were consistent across cognitive, behavioral, and affective dimensions. This suggests that personalization in PersonaGPT was not superficial but rather deeply embedded in its interaction design and instructional strategy.

The study also yields valuable insights into the nature of personalization. It is not simply a matter of selecting the right explanation format or controlling pacing. What matters most is the system's ability to recognize changes in learner state and respond with emotionally appropriate, pedagogically effective interventions. Reflexive dialogue – such as offering encouragement, rephrasing confusing concepts, and inviting clarification – proved critical in maintaining trust and engagement. The data further suggest that adaptive dialogue can be just as important as adaptive content in achieving learning outcomes.

Nevertheless, it is important to acknowledge several limitations. While the sample size was sufficient for medium-effect statistical power, replication in larger and more diverse populations would help verify generalizability. The study was also restricted to one domain – SQL programming. It remains to be seen how the same personalization mechanisms would perform in other subjects, such as writing or calculus. Additionally, the system's reliance on the GPT-4 API introduces dependencies that may pose barriers for cost-sensitive institutions. Future work should explore fine-tuned models that offer similar capabilities with fewer constraints.

Another potential limitation concerns behavioral bias. It is possible that learners behave differently when interacting with an AI tutor than they would with a human instructor. Some students may feel more comfortable taking risks; others may struggle to interpret cues from a non-human agent. Although the positive results in this study suggest that PersonaGPT overcame such issues, it would be valuable to compare AI-human hybrid models in future research.

Despite these limitations, the robustness of the findings points to a clear conclusion: PersonaGPT provides an effective, engaging, and scalable solution for AI-driven personalization in education. Its hybrid approach – combining behavioral modeling, language analysis, and adaptive generation – offers a promising blueprint for future systems that aim to replicate the nuance of human tutoring in a digital form.

V. Discussion and Conclusions

This study marks a meaningful shift in how we design and think about conversational agents in education. While adaptive learning systems and intelligent tutors have come a long way, the combination of real-time learner modeling with responsive, reflexive dialogue – seen in PersonaGPT – opens new doors. What sets PersonaGPT apart isn't just the accuracy of its responses, but the way it actively listens and adjusts to the learner's state, much like a skilled human tutor. This kind of personalization, based on behavioral and emotional cues, represents a more relational and reflective approach to AI in education.

One of the clearest findings is how much tone, wording, and pacing matter. Learners did not just receive correct answers – they felt understood. Unlike older systems that personalized only through content filtering or level adjustment, PersonaGPT shows that the way a system communicates can deeply affect engagement and learning outcomes. Its ability to detect confusion or hesitation without being directly asked adds a layer of authenticity and care often missing in digital tools.

The implications go beyond technical subjects like SQL. If AI systems can adapt not just what they teach but how they teach—and do so transparently – they become more than instructional tools. They become collaborators in learning, capable of motivating students, encouraging persistence, and offering support when human teachers can't. This is especially important in large or asynchronous learning environments where students often feel alone.

Technically, PersonaGPT demonstrates that hybrid systems combining rule-based logic with large language models can remain both adaptive and interpretable. Its use of open behavioral cues, rather than intrusive data, keeps it explainable and privacy-conscious. Its modular design also makes it scalable and institution-friendly, ready for integration into a variety of educational settings.

More broadly, the study invites us to rethink what we want from AI in education. Instead of focusing solely on efficiency or performance metrics, we should aim to nurture confident, independent learners. PersonaGPT shows that learners appreciate systems that respond to them genuinely and attentively. This brings empathy to the forefront—raising important questions about how future AI might balance emotional intelligence with academic rigor.

Of course, these advances also raise concerns. While learners benefited from PersonaGPT's guidance, we must examine the long-term impact of relying on AI for emotional and cognitive support. There's a risk of fostering dependence or reducing self-regulation. Future research should explore how to maintain a healthy balance, monitor changes in system behavior over time, and ensure that personalization continues to serve educational goals in an ethical and transparent way.

In conclusion, PersonaGPT represents a promising evolution in the design of AI-driven educational systems. It offers a working example of how learner modeling, conversational adaptivity, and affective sensitivity can be woven into a cohesive, responsive, and pedagogically meaningful interaction paradigm. By moving beyond static profiling and embracing dialogue as a dynamic, context-aware process, the system provides not only measurable improvements in learning but also a more humane and personalized experience of AI-mediated instruction. The findings of this study reinforce the potential of AI not simply to deliver content, but to participate in the learning process as an adaptive partner – attuned to the learner's needs, responsive in its feedback, and supportive of deeper engagement with the material. As the field moves forward, systems like PersonaGPT can serve as both a technical model and a conceptual inspiration for the future of human-centered educational technology.

References

[1] Z. Almahasees, A. Al-Harahsheh, and Y. Albudairi, "AI-powered education: Revolutionizing teaching and learning through artificial intelligence in Jordan," *Research Journal in Advanced Humanities*, vol. 5, no. 4, Sept. 2024. [Online]. Available: https://doi.org/10.58256/dqpagd83

[2] A. Krouska, C. Troussas, and M. Virvou, "A literature review of social networking-based learning systems using a novel ISO-based framework," *Intelligent Decision Technologies*, vol. 13, no. 1, pp. 23–39, 2019. [Online]. Available: https://doi.org/10.3233/IDT-190362

[3] W. Wiboolyasarin, K. Wiboolyasarin, P. Tiranant, N. Jinowat, and P. Boonyakitanont, "AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications," *Ampersand*, vol. 14, p. 100224, 2025. [Online]. Available: https://doi.org/10.1016/j.amper.2025.100224

[4] C. Troussas, C. Papakostas, A. Krouska, P. Mylonas, and C. Sgouropoulou, "Personalized feedback enhanced by natural language processing in intelligent tutoring systems," in *Augmented Intelligence and Intelligent Tutoring Systems. ITS 2023*, C. Frasson, P. Mylonas, and C. Troussas, Eds., *Lecture Notes in Computer Science*, vol. 13891. Cham, Switzerland: Springer, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-32883-1_58

[5] N. F. Davar, M. A. A. Dewan, and X. Zhang, "AI chatbots in education: Challenges and opportunities," *Information*, vol. 16, no. 3, p. 235, 2025. [Online]. Available: https://doi.org/10.3390/info16030235

[6] A. Krouska, C. Troussas, I. Voyiatzis, P. Mylonas, and C. Sgouropoulou, "ChatGPT-based recommendations for personalized content creation and instructional design with a tailored prompt generator," in *Proc. 2024 2nd Int. Conf. Foundation and Large Language Models (FLLM)*, Dubai, United Arab Emirates, 2024, pp. 295–299. [Online]. Available: https://doi.org/10.1109/FLLM63129.2024.10852487

[7] C. Troussas, C. Papakostas, A. Krouska, P. Mylonas, and C. Sgouropoulou, "Evaluating ChatGPT-driven automated test generation for personalized programming education," in *Proc. 2024 2nd Int. Conf. Foundation and Large Language Models (FLLM)*, Dubai, United Arab Emirates, 2024, pp. 194–200. [Online]. Available: https://doi.org/10.1109/FLLM63129.2024.10852510

[8] O. Avidov-Ungar and S. Zamir, "Personalization in education," in *IntechOpen*, 2024. [Online]. Available: https://doi.org/10.5772/intechopen.113380

[9] K. Chrysafiadi, C. Troussas, and M. Virvou, "Combination of fuzzy and cognitive theories for adaptive e-assessment," *Expert Systems with Applications*, vol. 161, p. 113614, 2020. [Online]. Available: https://doi.org/10.1016/j.eswa.2020.113614

[10] W. Kaiss, K. Mansouri, and F. Poirier, "Effectiveness of an adaptive learning chatbot on students' learning outcomes based on learning styles," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 13, pp. 250–261, 2023. [Online]. Available: https://doi.org/10.3991/ijet.v18i13.39329

[11] W. Kaiss, K. Mansouri, and F. Poirier, "Towards a chatbot-based learning object recommendation: A comparative experiment," in *Augmented Intelligence and Intelligent Tutoring Systems. ITS 2023*, C. Frasson, P. Mylonas, and C. Troussas, Eds., *Lecture Notes in Computer Science*, vol. 13891. Cham, Switzerland: Springer, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-32883-1_14

[12] J. N. Davies, M. Verovko, O. Verovko, and I. Solomakha, "Personalization of e-learning process using AI-powered chatbot integration," in *Mathematical Modeling and Simulation of Systems (MODS'2020)*, S. Shkarlet, A. Morozov, and A. Palagin, Eds., *Advances in Intelligent Systems and Computing*, vol. 1265. Cham, Switzerland: Springer, 2021. [Online]. Available: https://doi.org/10.1007/978-3-030-58124-4_20

[13] E. Gkintoni, H. Antonopoulou, A. Sortwell, and C. Halkiopoulos, "Challenging cognitive load theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy," *Brain Sciences*, vol. 15, no. 2, p. 203, 2025. [Online]. Available: https://doi.org/10.3390/brainsci15020203

[14] P. Hinduja, R. Fakir Mohammad, and S. Siddiqui, "Factors influencing students' academic self-efficacy in related domains," *SAGE Open*, vol. 14, no. 4, 2024. [Online]. Available: https://doi.org/10.1177/21582440241289738

[15] D. Ramandanis and S. Xinogalos, "Designing a chatbot for contemporary education: A systematic literature review," *Information*,

vol. 14, no. 9, p. 503, 2023. [Online]. Available: https://doi.org/10.3390/info14090503

[16] C. Kooli, "Chatbots in education and research: A critical examination of ethical implications and solutions," *Sustainability*, vol. 15, no. 7, p. 5614, 2023. [Online]. Available: https://doi.org/10.3390/su15075614

[17] P. Pergantis, V. Bamicha, C. Skianis, and A. Drigas, "AI chatbots and cognitive control: Enhancing executive functions through chatbot interactions—A systematic review," *Brain Sciences*, vol. 15, no. 1, p. 47, 2025. [Online]. Available: https://doi.org/10.3390/brainsci15010047

[18] A. Uzoka, E. Cadet, and P. U. Ojukwu, "Leveraging AI-powered chatbots to enhance customer service efficiency and future opportunities in automated support," *Computer Science & IT Research Journal*, vol. 5, no. 10, pp. 2485–2510, 2024. [Online]. Available: https://doi.org/10.51594/csitrj.v5i10.1676

[19] C. Zhai and S. Wibowo, "A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: The case of second language acquisition," *Heliyon*, vol. 8, no. 12, p. e12056, 2022. [Online]. Available: https://doi.org/10.1016/j.heliyon.2022.e12056

[20] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI chatbots in education: Systematic literature review," *International Journal of Educational Technology in Higher Education*, vol. 20, p. 56, 2023. [Online]. Available: https://doi.org/10.1186/s41239-023-00426-1

[21] J. Llanes-Jurado, L. Gómez-Zaragozá, M. E. Minissi, M. Alcañiz, and J. Marín-Morales, "Developing conversational virtual humans for social emotion elicitation based on large language models," *Expert Systems with Applications*, vol. 246, p. 123261, 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.123261

[22] I. Gligorea, M. Cioca, R. Oancea, A.-T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: A literature review," *Education Sciences*, vol. 13, no. 12, p. 1216, 2023. [Online]. Available: https://doi.org/10.3390/educsci13121216

[23] C. Papakostas, C. Troussas, A. Krouska, and C. Sgouropoulou, "Integrating Bayesian knowledge tracing and human plausible reasoning in an adaptive augmented reality system for spatial skill development," *Information*, vol. 16, no. 6, p. 429, 2025. [Online]. Available: https://doi.org/10.3390/info16060429

[24] M. Iqbal, S. Sarwar, M. Safyan, and M. Nasralla, "Personalized and adaptive e-learning systems for semantic Web: A systematic review and roadmap," *International Journal of Web Information Systems*, vol. ahead-of-print, no. ahead-of-print, 2024. [Online]. Available: https://doi.org/10.1108/IJWIS-01-2024-0026

[25] F. Li, Y. He, and Q. Xue, "Progress, challenges and countermeasures of adaptive learning: A systematic review," *Educational Technology & Society*, vol. 24, no. 3, pp. 238–255, 2021. [Online]. Available: https://www.jstor.org/stable/27032868

[26] B. Albreiki, T. Habuza, Z. Shuqfa, M. A. Serhani, N. Zaki, and S. Harous, "Customized rule-based model to identify at-risk students and propose rational remedial actions," *Big Data and Cognitive Computing*, vol. 5, no. 4, p. 71, 2021. [Online]. Available: https://doi.org/10.3390/bdcc5040071

[27] N. Shabani, A. Beheshti, H. Farhood, M. Bower, M. Garrett, and H. Alinejad-Rokny, "A rule-based approach for mining creative thinking patterns from big educational data," *AppliedMath*, vol. 3, no. 1, pp. 243–267, 2023. [Online]. Available: https://doi.org/10.3390/appliedmath3010014

[28] K. Kabassi, "A framework for evaluating distance learning of environmental science in higher education using multi-criteria group decision making," *Informatica*, vol. 47, no. 4, pp. 487–500, 2023. [Online]. Available: https://doi.org/10.31449/inf.v47i4.3401

[29] L. M. Regueras, M. J. Verdú, and J.-P. de Castro, "A rule-based expert system for teachers' certification in the use of learning management systems," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 7, pp. 75–81, 2022.

[30] Henderi, Q. Aini, A. D. Srenggini, and A. Khoirunisa, "Rule-based expert system for supporting assessment of learning outcomes," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1.2, 2020. [Online]. Available: https://doi.org/10.30534/ijatcse/2020/3991.22020

[31] S. Minn, Y. Yu, M. C. Desmarais, F. Zhu, and J.-J. Vie, "Deep knowledge tracing and dynamic student classification for knowledge tracing," in *Proc. 2018 IEEE Int. Conf. Data Mining (ICDM)*, Singapore, 2018, pp. 1182–1187. [Online]. Available: https://doi.org/10.1109/ICDM.2018.00156

[32] C. S. Teh and A. R. Kudus, "The application of fuzzy cognitive mapping in education: Trend and potential," *TEM Journal*, vol. 13, no. 2, pp. 976–991, May 2024. [Online]. Available: https://doi.org/10.18421/TEM132-13

[33] L. Martins, M. Fernández-Ferrer, and E. Puertas, "Analysing conversation pathways with a chatbot tutor to enhance self-regulation in higher education," *Education Sciences*, vol. 14, no. 6, p. 590, 2024. [Online]. Available: https://doi.org/10.3390/educsci14060590

[34] A. Hardy, A. Paranjape, and C. Manning, "Effective social chatbot strategies for increasing user initiative," in *Proc. 22nd Annu. Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, Singapore and Online, 2021, pp. 99–110.

[35] S. Alagarsamy and S. Mehrolia, "Exploring chatbot trust: Antecedents and behavioural outcomes," *Heliyon*, vol. 9, no. 5, p. e16074, 2023. [Online]. Available: https://doi.org/10.1016/j.heliyon.2023.e16074