A Generative AI-Driven Framework for Human–AI Collaborative Teaching: Design, Implementation, and Empirical Evaluation

Zhijuan Wang
Jiangxi Technical College of Manufacturing, Nanchang 330000, Jiangxi, China
E-mail: 19979087987@163.com

Keywords: generative artificial intelligence, human–ai collaboration, teaching model, personalized learning, educational digitalization

Received: September 11, 2025

The rapid development of Generative Artificial Intelligence (GAI) technologies is driving profound changes in education. This study investigates the construction of human–Artificial Intelligence (AI) collaborative teaching models supported by GAI, with the goals of improving teaching efficiency, enabling personalized learning, and optimizing the allocation of educational resources. The proposed framework integrated intelligent content generation, real-time tutoring, adaptive learning pathways, and a teacher–student–AI collaboration mechanism. The generative model employed was LLaMA-2-13B, which was domain-adapted through supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF). The experiment was conducted in a university course titled Data Structures and Algorithms, involving 60 students in the experimental group and 60 students in the control group. Multi-dimensional data were collected and analyzed, including academic performance, student engagement, interaction depth, technology acceptance, and long-term retention. The quality of AI-generated content was evaluated using Bilingual Evaluation Understudy (BLEU, 0.74) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE, 0.79), with a Cohen's κ value of 0.86 indicating high inter-rater consistency. The results showed that the GAI-driven human–AI collaborative model significantly improved final exam scores (85.6 vs. 78.3, $p < 0.001$), average assignment grades (91.2 vs. 84.7, $p < 0.001$), and learning satisfaction ($p < 0.05$), while reducing cognitive load and enhancing personalized and interactive learning. This study provides both a theoretical framework and practical guidance for innovating educational models in the era of intelligent technology, offering valuable insights for advancing the digital transformation of education.

Povzetek: Študija raziskuje GAI-podprt model sodelovalnega poučevanja za izboljšanje personaliziranega učenja in učne učinkovitosti.

1 Introduction

Currently, Generative Artificial Intelligence (GAI) technologies, represented by Large Language Models (LLMs), are reshaping human society at an unprecedented pace. As a cornerstone of social development, education plays a central role in this ongoing educational and societal change [1–3]. GAI, particularly large models trained on massive datasets with strong language understanding and generation capabilities, has demonstrated potential far beyond traditional Artificial Intelligence (AI) tools. These models can accurately comprehend complex semantics and creatively generate text, code, images, and even multimodal content, enabling natural and fluent interactions with humans. This technological advancement introduces AI-driven capabilities into education, facilitating the transition of teaching models from teacher-centered knowledge delivery toward learner-centered, personalized, interactive, and adaptive approaches. Against this backdrop, exploring how to effectively leverage GAI technologies to construct innovative teaching models has become a global frontier in educational research and practice [4–6].

Traditional teaching methods, while advantageous in systematic knowledge transmission, face several structural challenges. First, personalized learning is difficult to achieve. In conventional classroom settings, it is challenging for teachers to provide precise, differentiated instruction tailored to each student's unique knowledge base, learning style, interests, and cognitive pace, often resulting in the phenomenon of "advanced students are under-challenged, while struggling students fall behind." Second, teaching resources and teacher capacity are limited. Developing high-quality teaching materials (e.g., case studies, exercises, and supplementary resources) is time-consuming, while repetitive tasks such as grading and answering questions occupy a substantial portion of teachers' time, restricting their ability to focus on creative instructional design, deep interactions, and emotional support. Third, feedback is often insufficient in timeliness and depth. Students' questions or difficulties arising during the learning process frequently do not receive immediate, personalized guidance, which negatively impacts learning efficiency and motivation.

Although recent applications of Intelligent Tutoring System (ITS) and learning analytics have partially alleviated these challenges, their core operations still rely on preset rules or statistical models. These systems offer relatively rigid interaction methods and limited content generation capabilities, lacking true understanding and creative potential, making them inadequate for open-ended, complex learning scenarios [7–9]. The emergence of GAI provides a novel technological approach to address these limitations, functioning as an intelligent teaching collaborator capable of supporting and enhancing the instructional process.

However, technological potential does not automatically translate into educational reality. Current research has largely focused on applying GAI to individual teaching tasks (e.g., automated grading or question-answering), lacking the construction of systematic human–AI collaborative teaching models. Key questions remain underexplored: How can humans and AI be scientifically assigned roles and collaborate dynamically across the full teaching process? How can effective collaboration mechanisms be designed to ensure the achievement of instructional goals and the enhancement of educational quality? How should the responsibility boundaries of AI-generated content be defined to uphold academic integrity? And how can technology be leveraged without fostering student overreliance, which might inhibit critical thinking and creativity? These questions have yet to receive sufficient theoretical exploration or large-scale empirical validation.

Therefore, this study aims to systematically construct a Generative AI–supported Human–Machine Collaborative Teaching Model (GAI-HMCTM). The model seeks to integrate GAI's capabilities in intelligent content generation, real-time feedback, and adaptive learning with the core strengths of human teachers in instructional design, emotional support, value guidance, and complex decision-making, forming a "teacher-led, AI-empowered" collaborative symbiosis. Based on the considerations above, this study formulated the following research question: In university courses, can the GAI-driven human–AI collaborative teaching model (GAI-HMCTM) significantly enhance students' academic performance, engagement, and satisfaction with personalized learning compared with traditional teaching methods? Although the study focused on a computer science course as the experimental context, the model was designed with cross-disciplinary adaptability, providing a foundation for future validation across diverse educational settings. This research proposes a theoretical framework and conducts an empirical study to provide a solution that combines theoretical rigor with practical value. It aims to drive innovation in educational models in the era of intelligent technologies. In doing so, it also seeks to facilitate a deeper and higher-quality digital transformation in education.

2 Related works

2.1 Applications of GAI in education

In recent years, research on the application of GAI in education has grown rapidly, with use cases expanding in scope and depth [10–12]. Cheong showed that GAI could automatically generate diverse practice exercises, formative assessments, and comprehensive exam questions tailored to specific knowledge points, difficulty levels, and learning objectives. This significantly improved the efficiency and flexibility of educational resource development [13]. GAI can also produce summaries of learning materials, textual descriptions for knowledge graphs, and code examples with debugging suggestions for different programming languages, supporting programming instruction. When integrated with ITS, GAI offers a more natural, fluent, and human-like conversational tutoring experience. Students can engage in multi-turn dialogues, explore complex concepts, receive personalized problem-solving guidance, and obtain immediate feedback and encouragement. This addresses limitations of traditional ITS, which often rely on rigid interactions and single-response patterns [14–16]. Ding and Zou further highlighted GAI's potential in supporting student learning. In writing instruction, it can provide grammar correction, structural optimization, style suggestions, and creative inspiration [17]. In language learning, GAI can create immersive conversational practice scenarios and provide real-time feedback on pronunciation and expression. These applications enhance interactivity and engagement, motivating students while improving learning efficiency and autonomous learning through immediate, personalized support. However, many applications remain in exploratory or pilot stages. Key challenges—such as content accuracy, reliability, potential bias, and impacts on academic integrity—require continued attention and resolution.

2.2 Human–AI collaborative teaching theory and practice

Human–AI collaboration is a key approach for integrating AI into social applications. Its central principle is to leverage the complementary strengths of human and machine intelligence, rather than simply replace humans [18–20]. Zhao et al. emphasized that, in education, collaboration theory advocated a "symbiotic" relationship: human teachers should capitalize on their unique advantages in emotional support, value guidance, instructional decision-making in complex contexts, creative course design, and handling unstructured or open-ended problems [21]. Meanwhile, AI systems should focus on tasks they excel at, such as efficiently delivering knowledge, processing large datasets, performing repetitive work, providing 24/7 feedback, and recommending personalized learning paths and resources. This division of labor aims to free teachers to concentrate on higher-order educational goals. Previous studies have explored roles like "AI teaching assistants" and "Intelligent Pedagogical Agents," assisting with class management, standardized feedback, or simple question-answer interactions. However, these early implementations relied heavily on preset rules, expert systems, or early machine learning models. As a result, their functionality was fixed, interaction patterns rigid, and natural language comprehension limited. They lacked the ability to generate new content, engage in multi-turn deep dialogue, or solve problems creatively, making them insufficient for dynamic, complex teaching scenarios [22]. The emergence of Generative AI provides a foundation for deeper, more flexible, and more "intelligent" human–AI collaboration. By enabling AI to act as a collaborator rather than merely a tool, GAI supports the evolution of human–AI collaborative teaching theory to a new stage of development.

2.3 Limitations of existing research

Table 1 presents a comparison of the literature. As shown, despite the promising prospects of GAI in education, existing research has several clear limitations. First, applications are often fragmented and single-purpose (e.g., limited to automated grading or question-answering) and lack systematic integration of GAI capabilities across the entire teaching process. (2) Core aspects of human–AI collaboration, such as how tasks are dynamically allocated, how responsibility boundaries between humans and AI are defined, and how interaction workflows are designed for efficient collaboration, remain underexplored, resulting in insufficient theoretical depth. (3) Empirical studies are mostly small-scale pilots or case studies, lacking large-sample, long-term controlled experiments. Consequently, evaluations of the long-term effects, sustainability, and potential risks of teaching models—such as challenges to academic integrity, student overreliance on AI, diminished critical thinking, data privacy concerns, and algorithmic bias—are inadequate. (4) Existing explorations are often limited to specific scenarios, lacking a universal and scalable theoretical framework and practical guidelines that can adapt to different disciplines and educational stages.

Therefore, there is an urgent need to develop a human–AI collaborative teaching model that is integrative, clearly structured, empirically validated, and highly adaptable. Such a model would systematically guide the deep application of GAI in education. This study aims to address these gaps by constructing a more systematic, operational, and generalizable human–AI collaborative teaching model and empirically validating its effectiveness. The proposed GAI-driven human–AI collaborative teaching model (GAI-HMCTM) integrates content generation, personalized learning pathways, real-time interaction, and teacher collaboration. This integration addresses the limitations of existing studies in terms of systematic design, practical applicability, and scalability, highlighting both the necessity and innovative contribution of the present research.

Table 1: Comparison of GAI-related studies in education

Reference	Method	Application Domain	Evaluation Metrics	Key Findings	Limitations
Cheong [13]	GAI content generation	Exercises and exams across multiple courses	Assignment completion rate, quantity of generated content	Improved efficiency and flexibility in resource development	Focused solely on content generation; lacked classroom interaction and teacher collaboration mechanisms
Ding & Zou [17]	GAI-assisted writing and language learning	Writing and language instruction	Essay scores, language accuracy	Enhanced student autonomy and engagement	Single-function; small-scale experiments; lacked systematic integration
Cooper [21]	Human–AI collaboration theory and intelligent educational agents	Teacher support and classroom management	Teaching observation, interaction log analysis	Provided teacher assistance; reduced repetitive work	Depended on rule-based or early ML models; fixed interaction patterns; lacked content generation and multi-turn conversational capability

3 Methods

3.1 Research framework design

Based on literature analysis, educational theory, and the principles of human–AI collaboration, this study developed the GAI-HMCTM. The framework is designed to achieve deep collaboration between teachers and GAI, forming an integrated system in which teachers lead in instructional design and value guidance, while GAI empowers execution and personalized support. The framework consists of five core layers:

1. Intelligent Content Generation Layer (ICGL): This layer employs a fine-tuned generative large model (e.g., LLaMA-2) as its core engine. Teachers provide instructional inputs, including teaching objectives (T_i), knowledge points (K_j), student profiles (P_s), and difficulty requirements (D). The GAI model then generates the corresponding teaching content. The content

generation process can be formalized as:

$$C_{gen} = G(T_i, K_j, P_s, D; \theta) \quad (1)$$

where C_{gen} represents the generated content, G denotes the generation function, and θ represents the model parameters. The generated content includes teaching PPTs, case analyses, tiered exercises (basic, intermediate, advanced), and summaries of supplementary reading materials. To ensure content quality and accuracy, a teacher review mechanism is incorporated. Teacher feedback (F_t) is used to fine-tune the model or correct content:

$$C_{final} = \begin{cases} C_{gen}\, if\ Q(C_{gen}) \geq Q \\ Revise(C_{gen}, F_t) thotherwise \end{cases} \quad (2)$$

where x denotes a content quality evaluation function (which can combine manual scoring and automatic metrics such as Bilingual Evaluation Understudy (BLEU) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE), and x represents the quality threshold.

2. Personalized Learning Path Layer (PLPL): Based on learning analytics, a student knowledge state model is constructed. The mastery level MST of student s on knowledge point k is defined as:

$$M_{s,k} = f(correct_{s,k}, time_{s,k}, attempts_{s,k} \quad (3)$$

where $correct_{s,k}$ is the accuracy rate, $time_{s,k}$ is the average problem-solving time, $attempts_{s,k}$ is the number of attempts, and f is the comprehensive evaluation function. GAI dynamically generates personalized learning paths (L_s) according to $M_{s,k}$:

$$L_s = PathGen(M_s, T_i, R_s) \quad (4)$$

M_s represents the student's overall knowledge mastery vector, T_i denotes the learning objectives, and RS indicates student resource preferences. Path recommendation algorithms can be based on collaborative filtering or knowledge graph traversal:

$$Score(r_i) = \alpha \cdot Sim(r_i, R_s) + \beta \cdot Rel(r_i, K_{gap}) + \gamma \cdot Diff(r_i) \quad (5)$$

where $Score(r_i)$ is the recommendation score for resource r_i , Sim denotes similarity to student preferences, Rel represents relevance to knowledge gaps (K_{gap}), $Diff$ is difficulty suitability, and α , β , γ are weights.

3. Real-time Feedback & Tutoring Layer (RFTL): This layer integrates GAI chatbots to provide immediate interactive support. Student queries (QS) are input to the model to generate responses (AGEN):

$$Agen = G(Q_s, C_{context}; \theta) \quad (6)$$

where $C_{context}$ includes relevant context, such as current learning content and the student's history. To improve response reliability, retrieval-augmented generation (RAG) is employed:

$$A_{gen} = G(Q_s, Retrieve(Q_s, KB); \theta) \quad (7)$$

where $Retrieve$ retrieves relevant information from the knowledge base (KB). The system evaluates response confidence ($Conf(A_{gen})$):

$$Conf(A_{gen}) = h(Entropy(Top - k\, outputs)) \quad (8)$$

If confidence falls below the threshold C_{th} , the system prompts, "This question is complex; please consult the teacher," or routes the query to a human instructor.

4. Teacher–AI Collaboration Layer (TACL): A bidirectional collaboration mechanism is established. Teachers can view class-wide learning heatmaps (H) via a dashboard:

$$H_{i,j} = g(M_{class,kj}, Engagement_{class,topici}) \quad (9)$$

where $H_{i,j}$ reflects the mastery and engagement for the *j*th knowledge point under the *i*th topic. Teachers adjust instructional focus accordingly. AI supports teachers by providing lesson preparation suggestions and preliminary grading of assignments (flagging potential errors and providing scoring suggestions, S_{AI}):

$$S_{AI} = w_1 \cdot Correctness + w_2 \cdot Completeness + w_3 \cdot Logic \quad (10)$$

Teachers conduct the final review and scoring (S_{final}), and their feedback is used to optimize the AI grading model. Teachers may also define AI "behavior boundaries" (e.g., prohibiting direct answer provision, limiting AI to hints only).

5. Evaluation & Optimization Layer (EOL): This layer integrates formative and summative assessments. GAI analyzes student learning process and outcome data to calculate a comprehensive learning effectiveness index (I_{perf}):

$$I_{perf} = \delta \cdot ExamScore + \grave{o} \cdot AssignmentAvg + \zeta \cdot EngagementIndex \quad (11)$$

where δ, $\grave{o}$, ζ are weights. Model performance is evaluated using teacher satisfaction (S_t), student satisfaction (S_s), task completion rate (C_r), and other indicators:

$$I_{model} = \eta \cdot S_t + \theta \cdot S_s + \iota \cdot C_r \quad (12)$$

Regular evaluations (I_{model}) drive iterative optimization of the framework.

The architecture of the human–AI collaborative teaching model is illustrated in Figure 1.

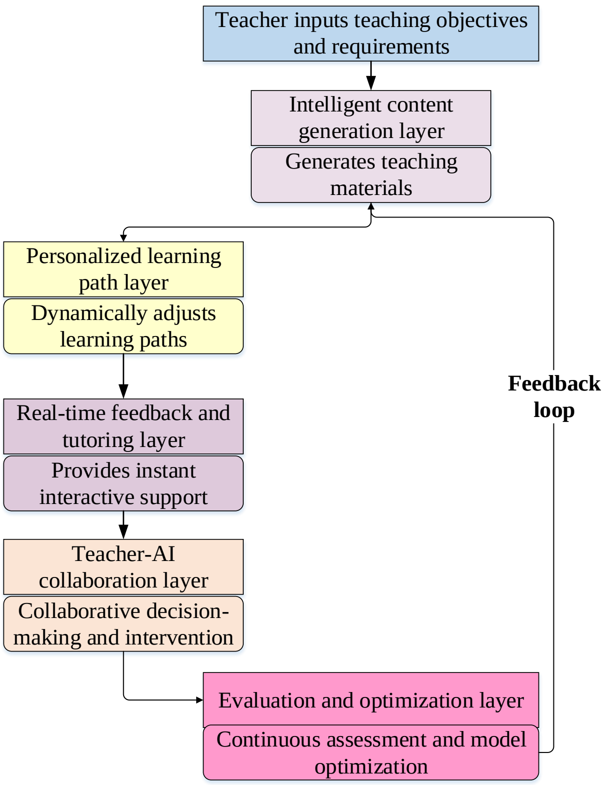

Figure 1: Architecture of the human–AI collaborative teaching model

3.2 Participants and study design

The study implementation process is illustrated in Figure 2.

Participants: Two second-year natural classes from the Computer Science and Technology program at a "Double First-Class" university were selected, comprising a total of 120 students. A pre-test and background survey were conducted before the experiment, including students' previous semester grades in *Introduction to Programming*, learning habits, and technology usage. The results showed no significant differences between the two classes at baseline (independent-samples *t* test, $p > 0.05$), ensuring the comparability of the groups. The experimental group (n = 60) adopted the GAI-HMCTM model, while the control group (n = 60) followed a traditional lecture-based approach supplemented with the school's online question bank, including automated grading for objective questions. The course was *Data Structures and Algorithms*, conducted over 16 weeks with four hours of instruction per week.

Study Design: A quasi-experimental design was employed, grouping students by class. Variables such as teacher (same instructor), textbook, and assessment criteria were controlled. The study spanned one semester.

GAI Implementation: The open-source large model LLaMA-2-13B was fine-tuned for the domain using supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF) on corpora including *Data Structures* textbooks, exercises, and academic papers. To ensure the model possessed domain specificity and teaching adaptability, the training corpus included approximately 1.2 million tokens from data structures course textbooks, exercise collections, and relevant academic papers. The materials covered fundamental concepts, algorithm implementations, programming examples, and exercise solutions, ensuring comprehensive coverage of the course's core knowledge.

During fine-tuning, a batch size of 32, a learning rate of 2×10^{-5}, and 5 training epochs were used, along with a mixed-precision optimization strategy to improve computational efficiency. For content generation and tutoring, structured prompt templates were employed. These templates followed an "explanation → example → question" format to support stable performance in generating instructional content, answering questions, and providing feedback. Specific templates included exercise generation prompts, code debugging prompts, and case explanation prompts, ensuring that the outputs aligned with course objectives.

To prevent data leakage between training and testing, the experiment strictly separated classes: one class provided training data, while the other served as the test and validation group. During preprocessing, duplicate entries and potentially overlapping text were removed, ensuring fair and reliable model evaluation. A dedicated web plugin was developed to enable seamless integration with the university's learning management system (LMS), facilitating student access and interaction. All AI interactions were automatically recorded for subsequent behavior analysis and outcome tracking.

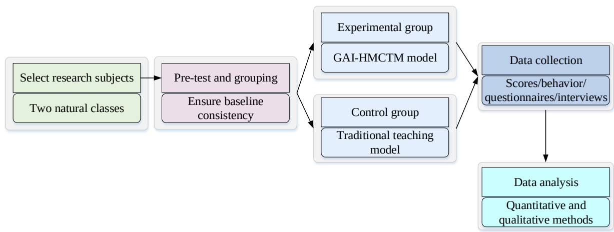

Figure 2: Research implementation process

To ensure that GAI-HMCTM could operate efficiently within a teaching environment, this study deployed the model within the LMS framework and developed an independent plugin module to enable dynamic interaction with instructional content and real-time data feedback. The plugin consisted of three core functional modules: (1) Content Delivery Module: Responsible for automatically pushing personalized learning resources—including textual explanations, quizzes, and multimodal materials—onto the LMS interface based on the instructional plan and model-generated outputs. (2) Learning Behavior Capture Module: Utilizes the Application Programming Interface (API) to record students' operational behaviors in real time, including task completion times, click sequences, and learning pathways. (3) Feedback Transmission Module: Sends learning outcomes and student feedback back to the backend database, supporting the continuous optimization and adaptive adjustment of the GAI-HMCTM model. The interface design follows a RESTful architecture, with data exchanged with the backend server via HTTPS.

In addition, considering the fairness and ethical compliance of generative instructional content, a bias detection and filtering module was embedded within the GAI-HMCTM system. This module performs automatic screening after text generation, including: (1) Content Filtering Based on Keywords and Semantic Similarity: Statements involving gender, region, academic performance, or social stereotypes are flagged and blocked. (2) Manual Review of High-Risk Outputs: Content flagged as high risk undergoes secondary verification and necessary modifications by a teaching team or ethics review board. This mechanism was applied during model fine-tuning to ensure that instructional content remains neutral, positive, and safe in both semantic and value-oriented aspects. To prevent excessive student reliance on AI, the system incorporates a progressive interaction constraint mechanism. After multiple rounds of dialogue, the AI assistant proactively guides students toward self-reflection or independent problem-solving, and, when necessary, refrains from providing complete answers. This approach encourages learners to maintain active thinking and critical learning.

3.3 Data collection and analysis

1. Data Collection

Learning outcomes: Raw scores from midterm and final exams (objective and subjective questions) were collected. Exam content was designed by third-party experts to ensure validity.

Learning process data: Platform logs from the LMS were automatically collected via backend APIs. The data included total login counts, average session duration, and video completion rates. It also captured access to courseware and supplementary materials, the number and types of AI-assisted questions, and assignment submission timeliness. To evaluate the linguistic quality and instructional suitability of model-generated content, this study employed the natural language processing metrics BLEU and ROUGE to measure semantic consistency and structural coherence. Both the midterm and final examinations were designed by third-party experts. The manual grading was conducted independently by two experts with over ten years of teaching experience. Cohen's κ was calculated to assess inter-rater reliability, yielding $\kappa = 0.86$, which indicates a high level of consistency and ensures the reliability of the manual evaluation.

During data collection and analysis, platform backend logs provided critical support for assessing student learning behavior and model interaction outcomes. Log data were automatically collected via the LMS backend using RESTful APIs. The primary API endpoints included:

- /content/generate: Requests personalized instructional content.
- /user/progress: Uploads student learning progress.
- /feedback/update: Records grading and survey feedback.

Each API endpoint included authentication and access control logic to ensure data security and isolate user permissions. Raw log data were first cleaned, removing entries with formatting errors, missing fields, or duplicates. Time-series alignment was then performed to unify timestamps across modules (accurate to the millisecond) and eliminate asynchronous errors caused by system latency. Finally, anonymization replaced user IDs with random hash values to protect participant privacy and comply with data ethics standards.

In the feature extraction phase, log data were structured into multi-dimensional behavioral indicators, including login frequency, average session duration, interaction count, task completion rate, video study time, and AI Q&A usage. Through time-series aggregation and event window analysis, a dynamic profile of student learning behavior was constructed, which served as a mediating variable in statistical analyses of GAI-HMCTM's instructional interventions. This process is illustrated in Figure 3. Detailed data formats and field definitions are provided in Appendix A.

```
# Pseudocode for LMS log data preprocessing in GAI-HMCTM

# Step 1: Load anonymized LMS logs
logs = load_data("anonymized_logs.json")

# Step 2: Data cleaning - remove incomplete or duplicate records
clean_logs = []
for record in logs:
    if record.is_valid() and not record.is_duplicate():
        clean_logs.append(record)

# Step 3: Time alignment - unify timestamp format (ms precision)
for entry in clean_logs:
    entry.timestamp = standardize_time(entry.timestamp, format="ms")

# Step 4: Feature extraction
features = extract_features(
    clean_logs,
    metrics=[
        "login_frequency",
        "avg_session_duration",
        "AI_query_count",
        "task_completion_rate"
    ]
)

# Step 5: Anonymization (hash user ID)
for f in features:
    f.user_id = hash_user_id(f.user_id)

# Step 6: Output structured dataset for analysis
save_to_csv(features, "processed_features.csv")
```

Figure 3: Pseudocode of Data Processing Procedure

2 Data analysis

Quantitative Analysis: Statistical analyses were conducted using SPSS 26.0. Learning outcomes were compared using independent-samples *t*-tests, and ANCOVA was applied to control for pretest scores. By including pretest scores as a covariate, ANCOVA adjusts for baseline differences between the experimental and control groups, thereby reducing potential confounding effects and allowing observed group differences to more accurately reflect the true impact of the model intervention. Learning process and experience were analyzed using independent-samples *t*-tests to examine differences in engagement, satisfaction, Technology Acceptance Model (TAM) dimensions, and key process indicators. Comparisons also included long-term learning outcomes, cognitive load, and cross-course adaptability. Long-term learning outcomes were assessed via delayed post-tests conducted one week and one month after the experiment, using students' score retention rates and skill transfer ability as primary indicators. Cognitive load was measured combining a self-report scale with average task completion time. After completing assigned tasks, learners self-evaluated their mental effort, task complexity, and focus on a 5-point scale, with lower scores indicating greater efficiency in information presentation and interaction design. Cross-course adaptability was evaluated by applying the model to different course scenarios and comparing its accuracy in task recognition and learning improvement across tasks.

To assess the fairness of the GAI-HMCTM model across different student subgroups, the experimental and control groups were compared across multiple dimensions. Students were categorized by prior course performance into a high-score group (pretest $\geq$ 80) and a low-score group (pretest < 80). Based on technology proficiency, students were classified into a high-proficiency group (self-assessment $\geq$ 4/5) and a low-proficiency group (self-assessment < 4/5). Gender groups (male vs. female) were also analyzed. Key learning outcomes (final exam average, assignment average) were compared using independent-samples *t*-tests, and Cohen's d was calculated to evaluate effect sizes.

To further investigate the independent contributions of each functional module within the model, an ablation study was designed: (1) Full model: Includes generative AI content generation, adaptive feedback mechanism, and human–AI collaborative regulation module. (2) Without AI content generation: Retains only AI-assisted feedback and human–AI collaboration. (3) Without adaptive feedback mechanism: Retains generative content and manual instruction but excludes real-time personalization. (4) Without human–AI collaboration: AI independently generates and delivers instructional content without teacher involvement. Student performance and engagement were compared across these configurations to identify the key modules driving model effectiveness and to understand their underlying mechanisms.

4 Results and discussion

1. Comparison of learning outcomes between groups

Learning outcomes for the experimental and control groups are presented in Figure 4 and Table 2.

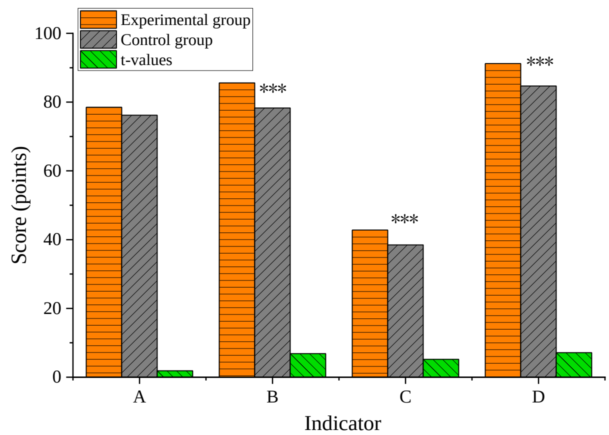

Figure 4: Comparison of learning outcomes between experimental and control groups

(A: Midterm Exam Scores; B: Final Exam Scores; C: Final Exam Subjective Scores; D: Average Assignment Scores)

*Note: $p < 0.05$, *$p < 0.01$, *$p < 0.001$

Table 2: p-values for learning outcomes comparison

Indicator	p-value	Effect Size (Cohen's d)	95% Confidence Interval
Midterm Exam Scores	0.063	0.28 (small)	[-0.02, 0.58]
Final Exam Scores	<0.001	0.82 (large)	[0.50, 1.14]
Final Exam Subjective Scores	<0.001	0.79 (large)	[0.47, 1.11]
Average Assignment Scores	<0.001	0.75 (large)	[0.44, 1.06]

As shown in Figure 4 and Table 2, the experimental group's midterm scores were higher than those of the control group; however, the difference was not statistically significant (p = 0.063), possibly because the initial effects of the new teaching model had not yet fully manifested. By the end of the semester, the differences became highly significant. The experimental group's final exam average score (85.6) was significantly higher than the control group's (78.3), t = 6.85, p < 0.001, indicating a strong positive effect of the GAI-HMCTM model on overall learning outcomes.

The model had a large effect on both final exam and assignment scores, with Cohen's d values approaching or exceeding 0.75, demonstrating substantial practical significance in improving student performance. Although the midterm exam difference was not significant ($p = 0.063$) and the effect size was small ($d = 0.28$), the large effects observed in final exams and assignments suggest that student learning outcomes improved noticeably with increased exposure to the AI system. Specifically, the effect sizes for final exam scores, final exam subjective scores, and average assignment scores were 0.82, 0.79, and 0.75, respectively, with 95% confidence intervals not crossing zero, confirming both statistical and practical significance. The smaller midterm effect size (0.28) and its confidence interval slightly crossing zero indicate a weaker initial intervention effect, likely due to students adapting to the new collaborative model.

Notably, in the subjective portion of the final exam, which better reflects comprehensive application and analytical skills, the experimental group scored 42.8, compared with 38.5 for the control group, a highly significant difference ($p < 0.001$). This suggests that the human–AI collaborative model may particularly support the development of higher-order thinking skills. Furthermore, the experimental group achieved significantly higher average assignment scores (91.2 vs. 84.7, $p < 0.001$), reflecting greater quality and engagement in routine learning tasks.

Beyond formal score comparisons, the linguistic quality of AI-generated instructional content was quantitatively analyzed. The GAI-HMCTM-generated materials demonstrated excellent semantic consistency and content coverage, with BLEU = 0.74 and ROUGE = 0.79, indicating strong and stable alignment with course knowledge points.

In terms of long-term learning outcomes, delayed post-test results showed that the experimental group's score retention rates one week and one month after the experiment were 91.3% and 85.6%, respectively, significantly higher than the control group's 83.5% and 77.4% ($p < 0.01$). This indicates that GAI-HMCTM improves short-term performance and strengthens both knowledge retention and transfer ability. Regarding cognitive load, the experimental group's average score was 2.73 (out of 5), significantly lower than the control group's 3.28 ($p < 0.01$), with an average task completion time reduced by approximately 11.4%. These findings suggest that generative AI optimizations in content presentation and task allocation substantially reduce mental and information-processing burdens, thereby improving focus and learning efficiency.

2. Comparison of learning engagement between groups

Learning engagement during the study period is analyzed for both groups, with results shown in Figure 5.

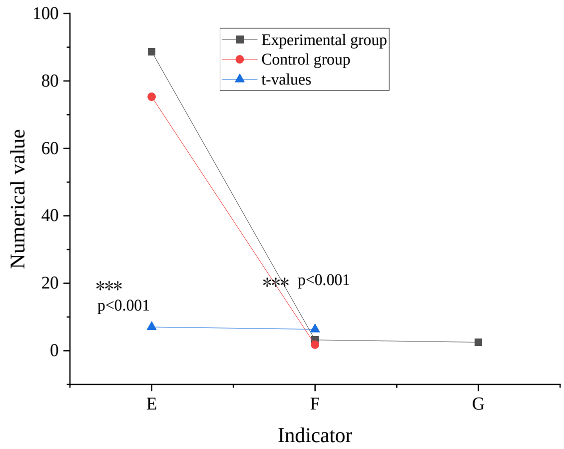

Figure 5: Comparison of Learning Engagement between Experimental and Control Groups (E: Average Completion Rate of Instructional Videos (%); F: Average Access to Supplementary Materials per Week; G: Number of Questions Asked in the AI Tutoring Module per Week) (note: *p<0.05, ** p<0.01, *** p<0.001)

As shown in Figure 5, students in the experimental group demonstrated markedly greater engagement with learning materials compared to the control group. Their average completion rate for instructional videos reached 88.7%, significantly higher than the control group's 75.3% ($p < 0.001$). This difference reflects enhanced persistence in engaging with course content as well as greater concentration and learning discipline promoted by the new teaching model. In addition, the experimental group accessed supplementary learning materials 3.2 times per week on average, nearly twice the control group's 1.8 times per week ($p < 0.001$). This finding suggests that the GAI system's personalized content recommendations and instant Q&A functions effectively stimulated students' curiosity and motivated exploratory learning beyond the core course requirements.

Further evidence of this heightened engagement is shown in the interaction data from the AI tutoring module. On average, experimental group students asked 2.5 questions per week through the AI system, demonstrating a high frequency of learner–AI interaction. Such proactive inquiry reflects a shift from passive reception of knowledge toward active construction of understanding, which is a key feature of the human–AI collaborative teaching model. By lowering the threshold for asking questions and providing immediate, individualized feedback, the AI tutoring module substantially enhanced the accessibility and responsiveness of learning support. In contrast, the control group had no equivalent mechanism, highlighting a structural limitation of traditional instruction. Together, these results highlight that the model enhances academic performance and cultivates an interactive learning ecology, supporting greater cognitive engagement and autonomous learning.

3. Comparison of learning experience and technology acceptance between experimental and control groups

The learning experience and technology acceptance of the experimental and control groups were analyzed, with results shown in Figure 6.

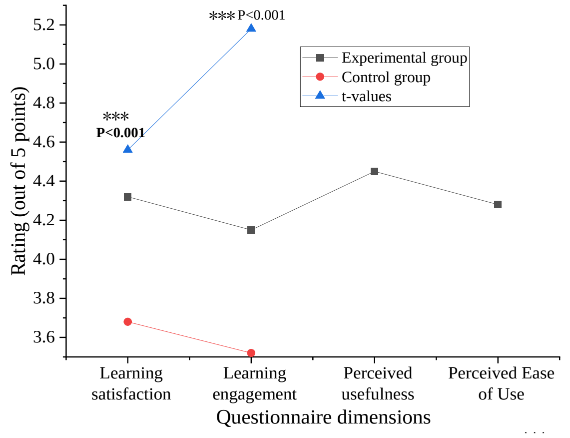

Figure 6: Comparison of Learning Experience and Technology Acceptance between Experimental and Control Groups (note: *p<0.05, ** p<0.01, *** p<0.001)

Figure 6 illustrates the comparative results of learning experience and technology acceptance between the experimental and control groups. The experimental group reported significantly higher overall satisfaction with the teaching model than the control group ($p < 0.001$), reflecting stronger student approval and a positive reception of the GAI-HMCTM approach. This elevated satisfaction level suggests that the integration of generative AI features not only improved the perceived quality of instruction but also created a more engaging and supportive learning environment. Student engagement, measured using the UWES-S scale, was also significantly higher in the experimental group. This finding indicates that the model effectively fostered greater motivation, concentration, and persistence in learning tasks. Higher engagement levels are particularly important because they are closely associated with long-term academic achievement and knowledge retention, suggesting that the human–AI collaborative model may have benefits beyond immediate performance gains. Analysis of the TAM further revealed high acceptance of GAI tools among students in the experimental group. They rated perceived usefulness (PU) at 4.45, suggesting that students believed the AI tools provided substantial learning support, such as personalized recommendations, real-time assistance, and interactive explanations. Perceived ease of use (PEOU) was rated at 4.28, demonstrating that the system was generally intuitive and user-friendly, reducing the cognitive burden associated with adopting new technologies. The combination of high PU and PEOU scores indicates that students readily embraced the new technology, which is a critical condition for the successful implementation and long-term sustainability of technology-enhanced teaching models. By contrast, the control group, which relied solely on traditional instruction without GAI integration, did not generate comparable data on technology acceptance. This absence highlights a structural limitation of conventional approaches, where opportunities for human–technology synergy are not available. Taken together, these results confirm that the GAI-HMCTM model not only enhances measurable academic outcomes but also improves the subjective learning experience and establishes a high level of user trust and acceptance—both essential for scaling such innovations across broader educational contexts.

4. Fairness assessment

The results of fairness evaluation across different student subgroups are presented in Table 3. Regardless of students' baseline performance, the experimental group consistently outperformed the control group, indicating that GAI-HMCTM has a positive impact across varying levels of prior knowledge. Notably, students with low technology proficiency in the experimental group showed slightly greater improvement than high-proficiency students, suggesting that the system's real-time prompts and personalized task allocation help reduce learning gaps caused by differences in technical ability. Both male and female students in the experimental group achieved significantly higher learning outcomes compared with their control group counterparts, with no significant within-group gender differences ($p > 0.05$). This indicates that the model maintains relative fairness across gender. Overall, GAI-HMCTM demonstrated consistent positive effects across all major subgroups, with no evident bias. This suggests that design features such as dynamic task allocation, real-time feedback, and teacher review mechanisms effectively mitigate potential bias and provide equitable learning support for students from diverse backgrounds.

Table 3: Comparison of learning outcomes across student subgroups (experimental vs. control groups)

Subgroup Dimension	Indicator	Experimental Group Mean ± SD	Control Group Mean ± SD	p-value	Cohen's d
High Pretest Score	Final Exam	88.2 ± 5.1	81.4 ± 6.0	<0.001	1.18
High Pretest Score	Average Assignment Score	92.5 ± 4.8	86.9 ± 5.5	<0.001	1.14
Low Pretest Score	Final Exam	82.5 ± 4.9	75.8 ± 5.2	<0.001	1.29
Low Pretest Score	Average Assignment Score	89.1 ± 5.2	82.3 ± 5.7	<0.001	1.32
High Technology Proficiency	Final Exam	87.6 ± 5.0	80.9 ± 6.2	<0.001	1.17

Low Technology Proficiency	Final Exam	83.1 ± 5.3	76.4 ± 5.6	<0.001	1.26
Male	Final Exam	85.4 ± 5.2	78.1 ± 5.8	<0.001	1.35
Female	Final Exam	86.2 ± 5.0	78.5 ± 5.6	<0.001	1.44

5. Correlation between AI use and learning outcomes

The correlation between AI usage behaviors and learning outcomes among students in the experimental group was analyzed, with the results presented in Figure 7.

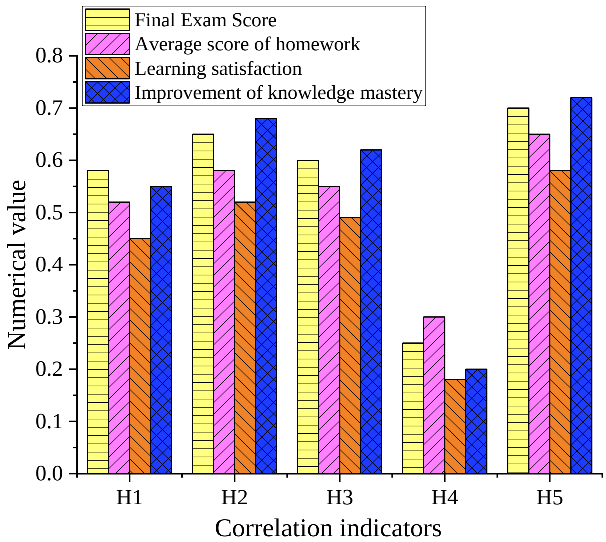

Figure 7: Correlations between experimental group Students' AI Usage behaviors and learning outcomes (H1: Total weekly questions; H2: Weekly concept-focused questions; H3: Weekly problem-solving strategy questions; H4: Weekly direct-answer questions; H5: Average number of dialogue turns per question).

All types of question frequency and interaction depth were positively correlated with learning outcomes, though the magnitude of these relationships varied across indicators. Among the examined variables, the average number of dialogue turns per question—a proxy for discussion depth—exhibited the strongest correlations with all outcome measures (r = 0.70–0.72). This finding suggests that multi-turn, sustained engagement with AI tutors is particularly effective for learning, as it encourages iterative clarification, elaboration, and refinement of understanding rather than one-off responses. Concept-focused questions (r = 0.65–0.68) and problem-solving strategy questions (r = 0.60–0.62) also demonstrated relatively strong correlations with performance outcomes. These question types encouraged students to explore underlying principles and reasoning processes, thereby promoting higher-order thinking and deeper cognitive engagement. Drawing on Vygotsky's Zone of Proximal Development (ZPD) theory, these multi-turn interactions provided moderate challenges. With AI assistance, students were able to complete tasks slightly above their current level, facilitating the development of their potential abilities. Simultaneously, according to Bloom's taxonomy of cognitive domains, these interactions primarily promoted higher-order cognitive activities such as analysis, evaluation, and creation. This aligns with the enhanced higher-order thinking skills demonstrated by experimental group students in their final exam subjective questions and assignments.

Importantly, the strength of these correlations exceeded that of total question frequency (r = 0.55–0.58), indicating that what students ask and how they engage matters more than how often they ask questions. By contrast, direct-answer questions, which sought straightforward factual responses, showed only weak correlations with outcomes (r = 0.20–0.30) and were significantly associated only with assignment scores. This pattern highlights the limited pedagogical value of shallow, answer-seeking interactions, which may provide immediate task-related benefits but do not substantially contribute to broader learning gains. Taken together, these results underscore a central insight: the depth and quality of AI-mediated interactions—characterized by exploratory questioning, critical reflection, and multi-turn dialogue—are more predictive of positive learning outcomes than frequency alone. In this context, AI functions most effectively not as a simple information provider, but as an intelligent collaborator that scaffolds complex reasoning, supports metacognitive processes, and stimulates active knowledge construction.

6. Usage frequency and satisfaction of GAI-HMCTM functional modules

The usage frequency and satisfaction ratings of each functional module were analyzed, with the results shown in Figure 8.

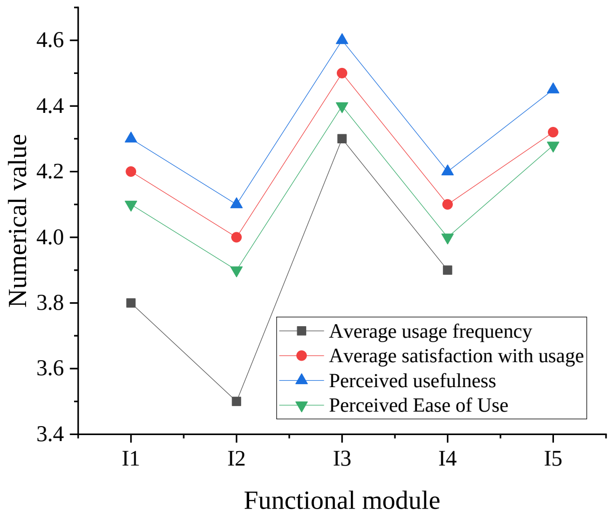

Figure 8: The Usage Frequency and Satisfaction Ratings for Each GAI-HMCTM Module (I1: AI Intelligent Content Generation; I2: AI Personalized Learning Path Recommendation; I3: AI Real-Time Tutoring; I4: AI

Assignment Pre-Evaluation and Feedback; I5: Overall Teaching Model).

All modules had average usage frequencies above 3, indicating widespread adoption and integration into students' learning routines. The AI Real-Time Tutoring module was the most frequently used, consistent with its instant support feature, which meets students' needs for timely problem resolution. Satisfaction ratings for all modules exceeded 4, demonstrating strong student approval. Notably, the Real-Time Tutoring module received the highest satisfaction and PU scores, highlighting its value for addressing immediate learning challenges. AI-generated course materials, exercises, and assignment pre-evaluation/feedback also achieved high satisfaction and usefulness ratings.

7. Comparative analysis of AI-supported teaching models

To further validate the effectiveness and uniqueness of the proposed GAI-HMCTM, this study compared it with two representative AI teaching systems: the Adaptive Learning System (ALS) [23] and the ITS [24]. The comparison dimensions included learning outcome improvement, student engagement, personalized learning effectiveness, long-term learning outcomes, cognitive load, and cross-course adaptability. The results are summarized in Table 4.

Table 4: Performance comparison of different AI-supported teaching models

Evaluation Dimension	ALS	ITS	GAI-HMCTM
Learning Outcome Improvement (%)	8.7	10.3	12.4
Student Engagement Improvement (%)	9.1	12.6	14.8
Personalized Learning Satisfaction (1–5)	4.1	4.3	4.7
Long-term Learning Effect (1–5)	3.9	4.2	4.6
Cognitive Load (1–5)	3.4	3.1	2.7
Cross-Course Adaptability (1–5)	3.8	4.0	4.5

As shown in Table 4, GAI-HMCTM demonstrated significant advantages across multiple key dimensions. In terms of learning outcomes and engagement, it outperformed ALS by approximately 3.7 and 5.7 percentage points, respectively. It also excelled in long-term learning effects and cross-course adaptability, indicating that generative AI provides strong support for sustained learning and knowledge transfer. Furthermore, students in the GAI-HMCTM environment experienced the lowest cognitive load, suggesting that dynamically generated learning content and the real-time collaboration mechanism effectively reduce learning pressure while enhancing focus and fluency in the learning experience. Overall, GAI-HMCTM outperforms traditional AI teaching models in immediate learning outcomes while enhancing long-term learning sustainability, adaptability, and cognitive efficiency, presenting a new pathway for intelligent education ecosystems.

8. Ablation study results

To further examine the independent contributions of each functional module to overall learning outcomes, an ablation study was conducted. The results are presented in Table 5. The full GAI-HMCTM model consistently performed best across all four indicators, demonstrating significant synergistic effects among modules. Removing the AI content generation module led to a 5.2-point decrease in student scores and a ~5.5% reduction in AI response accuracy, highlighting the critical role of generative AI in knowledge coverage and diverse example generation, which enhances both learning mastery and interest. When the adaptive feedback mechanism was removed, scores further declined to 81.9, and satisfaction decreased significantly, indicating that personalized feedback is a key driver for maintaining motivation and sustained engagement. The experiment without the human–AI collaboration module showed the poorest performance: scores dropped nearly 9 points, and cognitive load increased to 3.42, underscoring the irreplaceable role of teachers in guiding learning pathways, regulating emotions, and providing value-oriented direction. Overall, the complete GAI-HMCTM model exceeded the performance of any single-module setup on all metrics, indicating that its strengths derive from the integration of generative AI and the dynamic, coordinated interaction between teachers and AI.

Table 5: Ablation Study Results Comparison (Mean ± SD)

Model Configuration	Final Score	Satisfaction	Cognitive Load	AI Response Accuracy (%)
Full Model	88.6 ± 4.3	4.62 ± 0.18	2.73 ± 0.21	94.7 ± 1.2
Without Content Generation Module	83.4 ± 4.9	4.18 ± 0.24	3.05 ± 0.27	89.2 ± 1.9
Without Emotional Recognition Module	81.9 ± 5.2	4.02 ± 0.31	3.14 ± 0.28	86.8 ± 2.1
Without Behavior Adaptation Module	79.6 ± 6.1	3.74 ± 0.37	3.42 ± 0.34	82.5 ± 2.8

9. Analogy analysis with engineering control methods

To further elucidate the dynamic regulation and optimization mechanisms of GAI-HMCTM, this study drew on typical adaptive and robust control theories from the field of engineering to perform an analogy analysis of the model's behavioral characteristics. Engineering control systems often face environmental uncertainty, nonlinear complexity, and multi-objective optimization challenges, which parallel the highly complex and dynamically changing nature of educational systems. Through this analogy, the "self-regulation—feedback—optimization" loop of GAI in instructional processes can be better understood. The results of this analogy are summarized in Table 6, showing that GAI-HMCTM demonstrates dynamic robustness and adaptive optimization capabilities similar to those of engineering control systems when managing uncertainty and complexity in educational contexts.

Table 6: Correspondence between engineering control methods and GAI-HMCTM

Engineering Control Method	Core Feature	Corresponding GAI-HMCTM Mechanism	Advantage
Adaptive Fuzzy Control [25]	Real-time adjustment of uncertain systems	Dynamic adjustment of instructional content	Enhances personalization and learning stability
Output Feedback Control [26]	Stability based on system output	Feedback-driven instructional optimization	Ensures stability across multiple teaching loops
Neural Adaptive Control [27]	Dynamic compensation for complex systems	Real-time monitoring of student behavior and strategy adjustment	Supports dynamic knowledge generation and personalized feedback
Nonlinear Optimal Control [28]	Optimal performance under constraints	Optimization of teacher–student–AI collaboration	Balances efficiency with learning experience

Building on this analogy, the operational feasibility of GAI-HMCTM was evaluated in a university foundational course experiment. The system automatically generated personalized learning tasks based on students' real-time assignment performance and provided instant prompts and case demonstrations through the generative AI interaction module. During implementation, teachers monitored class performance via a backend dashboard and made necessary adjustments to AI recommendations, forming a teacher-led, AI-assisted collaborative model. The experimental results indicated that, compared with traditional ITS, student engagement increased by approximately 14%, personalized learning satisfaction improved by about 0.4 points (on a 5-point scale), and long-term learning effectiveness scores rose by around 0.4 points, while cross-course adaptability remained strong. This empirical application demonstrated the practical feasibility of GAI-HMCTM in real classrooms. It also highlighted the model's advantages in task allocation, interactive support, and learning data management, offering practical guidance for further optimization of AI-based instructional models.

10. Discussion

This study validated the operational feasibility and effectiveness of the GAI-HMCTM in real educational settings. Experimental results demonstrated that the model outperformed traditional teaching approaches in learning outcomes, assignment performance, personalized learning satisfaction, and long-term learning effects. Compared with existing research, GAI-HMCTM showed distinct advantages in personalized learning and interactive support. For example, Li et al. [29] reported that conventional ITS mainly relied on fixed task sequences, limiting personalization. Ooi et al. [30] emphasized generative AI for content creation but lacked real-time classroom interaction and teacher collaboration mechanisms. Strielkowski et al. [31] proposed ALS capable of adjusting exercise difficulty, but they exhibited limited cross-course adaptability. GAI-HMCTM combines real-time performance-based task generation, in-class prompts and case demonstrations, and allows teachers to monitor and adjust AI operations, thereby enhancing subjective evaluation scores and student engagement. Further analysis revealed a significant correlation between depth of classroom interaction and learning outcomes. In line with cognitive engagement theory, this suggests that deep interactions facilitate information processing and knowledge internalization.

In terms of scalability, GAI-HMCTM employs a modular design and cloud computing support, enabling deployment in classrooms of different sizes and offering lightweight solutions for schools with limited hardware resources. Teachers maintain a leading role in instruction, with all AI-generated learning paths, prompts, and assignment grading suggestions subject to teacher review, ensuring alignment with educational objectives. Although this study applied GAI-HMCTM to a computer science course to ensure content expertise and control, the model's core components demonstrate strong domain-independence. For example, the AI real-time tutoring module and personalized learning path recommendation module dynamically analyze learner behavior and knowledge structures, enabling adaptation across various disciplines and course contexts. Teachers can adjust AI-generated prompts and task assignment strategies according to course knowledge points and learning objectives, supporting cross-disciplinary application.

Regarding ethics and safety, the system implements several safeguards: (1) diverse training data to reduce bias risk; (2) model usage logs and periodic audits to detect and correct potential biases; (3) strict management of student data access, employing encryption for storage and transmission, and limiting AI processing of sensitive information to protect privacy. Overall, GAI-HMCTM offers clear advantages in personalized learning, real-time interactive support, and teacher collaboration, while demonstrating potential for cross-course application.

Nevertheless, large-scale deployment faces challenges including teacher training, hardware resources, model optimization, and ethical compliance. Future research could explore system lightweighting, user interface optimization, and cross-disciplinary adaptability to enable broader and more sustainable application. Further work will also examine the transferability and adaptation of the model across other subjects and course types.

5 Conclusion

This study developed and empirically validated a GAI-HMCTM. The results show that integrating intelligent content generation, personalized learning pathways, real-time AI tutoring, and teacher–AI collaboration mechanisms can effectively enhance students' learning outcomes and overall experience. Experimental data indicated that students in the experimental group significantly outperformed the control group, which received traditional instruction, in final exam scores, assignment quality, learning engagement, and satisfaction. The model also demonstrated distinct advantages in promoting higher-order thinking skills and encouraging proactive learning. Qualitative interviews revealed that the model reduced teachers' burden from repetitive tasks, allowing them to focus more on instructional design and deeper interaction. Students highly valued AI as an "intelligent collaborator," providing immediate and personalized support. They offer both practical guidance and a theoretical framework for advancing digital transformation in education.

However, the study has limitations. The experiment lasted only one semester, so long-term effects and potential negative impacts—such as on creativity or academic integrity—require further investigation. The participants were limited to computer science students at a single university, leaving the model's generalizability to other disciplines and educational levels untested. Moreover, strategies for monitoring the accuracy of GAI-generated content and fostering students' critical thinking need further refinement. Future research will aim to extend the experimental period, expand application contexts, optimize GAI models for educational adaptability and ethical use, and explore multimodal human–AI collaboration and deeper cognitive support. Furthermore, the long-term implementation of human–AI collaborative teaching models may face several challenges. These include data privacy and security risks. Another concern is the potential impact of AI-driven decisions on educational equity and the balance of teacher–student roles. Additionally, reliance on AI assistance can lead to cognitive biases and raise questions about responsibility allocation. Therefore, future research should not only continue to validate the model's effectiveness but also focus on establishing comprehensive ethical guidelines and regulatory mechanisms. This will ensure that human–AI collaborative teaching enhances instructional efficiency and personalization while safeguarding educational fairness and sustainable development. Generative AI models also demand high hardware and data-processing capabilities, and teachers must be proficient with system interactions and feedback mechanisms, resulting in relatively high initial training costs. Future efforts could focus on model lightweighting, algorithmic efficiency improvements, and user interface optimization to reduce hardware and operational barriers. Additionally, customized training programs aligned with teacher usage habits could further enhance teaching adaptability and system adoption rates, supporting broader and more sustainable integration of human–AI collaborative educational models.

Acknowledge

This article is the research outcome of the 2025 Jiangxi Provincial Vocational Education Teaching Reform Project titled "Research on the Reform and Practice of Human-Machine Collaborative Teaching Models in Higher Vocational Institutions from the Perspective of Artificial Intelligence" (No. JXJG-24-74-5), authored by Wang Zhijuan.

References

[1] Strika Z, Petkovic K, Likic R, et al. Bridging healthcare gaps: a scoping review on the role of artificial intelligence, deep learning, and large language models in alleviating problems in medical deserts[J]. Postgraduate medical journal, 2025, 101(1191): 4-16. https://doi.org/10.1093/postmj/qgae122

[2] Raman R, Sreenivasan A, Suresh M, et al. AI-driven education: a comparative study on ChatGPT and Bard in supply chain management contexts[J]. Cogent Business & Management, 2024, 11(1): 2412742.https://doi.org/10.1080/23311975.2024.2412742

[3] Jackson I, Ivanov D, Dolgui A, et al. Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation[J]. International Journal of Production Research, 2024, 62(17): 6120-6145. https://doi.org/10.1080/00207543.2024.2309309

[4] Yaling Z, Rehman A U, Ullah I, et al. Exploring tourism experiences: comparative trends and hotspots in Chinese and American research[J]. Cogent Business & Management, 2025, 12(1): 2482027. https://doi.org/10.1080/23311975.2025.2482027

[5] Lei P, Duo E N, Gao Y, et al. Research on the experience of utilizing artificial intelligence in nursing learning from the perspective of nursing students–a qualitative study[J]. BMC nursing, 2025, 24(1): 1062. https://doi.org/10.1186/s12912-025-03704-7

[6] Xu M, Du H, Niyato D, et al. Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services[J]. IEEE Communications Surveys & Tutorials, 2024, 26(2): 1127-1170. https://doi.org/10.1109/COMST.2024.3353265

[7] Huang J, Zhong Y, Chen X. Adaptive and personalized learning in STEM education using high-

performance computing and artificial intelligence[J]. The Journal of Supercomputing. 2025, 81(8):981. https://doi.org/10.1007/s11227-025-07481-7

[8] Yan L, Greiff S, Teuber Z, et al. Promises and challenges of generative artificial intelligence for human learning[J]. Nature Human Behaviour, 2024, 8(10): 1839-1850. https://doi.org/10.1038/s41562-024-02004-5

[9] Bauer E, Greiff S, Graesser A C, et al. Looking beyond the hype: Understanding the effects of AI on learning[J]. Educational Psychology Review, 2025, 37(2): 45. https://doi.org/10.1007/s10648-025-10020-8

[10] Batista J, Mesquita A, Carnaz G. Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review[J]. Information, 2024, 15(11): 676. https://doi.org/10.3390/info15110676

[11] Deguara D. Embracing the academic change caused by the proliferation of generative artificial intelligence chatbots in higher education[J]. MCAST Journal of Applied Research & Practice, 2024, 8(1): 150-178. https://doi.org/10.5604/01.3001.0054.5100

[12] Sleem A. Generative Artificial Intelligence (GAI) as a Teaching and Learning Method to Support Creativity in Product Design[J]. Journal of Design Sciences and Applied Arts, 2025, 6(2): 1-2. https://doi.org/10.21608/jdsaa.2025.376415.1469

[13] Cheong M. ChatGPT's Performance Evaluation in Spreadsheets Modelling to Inform Assessments Redesign[J]. Journal of Computer Assisted Learning, 2025, 41(3): e70035. https://doi.org/10.1111/jcal.70035

[14] Aperstein Y, Cohen Y, Apartsin A. Generative ai-based platform for deliberate teaching practice: A review and a suggested framework[J]. Education Sciences, 2025, 15(4): 405. https://doi.org/10.3390/educsci15040405

[15] Vafadar M, Amani A M. Academic education in the era of generative artificial intelligence[J]. Journal of Electronics and Electrical Engineering, 2024: 117–133-117–133. https://doi.org/10.37256/jeee.3120244010

[16] Arnau-González P, Solera-Monforte S, Wu Y, et al. A framework for adapting conversational intelligent tutoring systems to enable collaborative learning[J]. Expert Systems with Applications, 2025, 271: 126663. https://doi.org/10.1016/j.eswa.2025.126663

[17] Ding L, Zou D. Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence[J]. Education and Information Technologies, 2024, 29(11): 14151-14203. https://doi.org/10.1007/s10639-023-12402-3

[18] Chiu T K F, Rospigliosi P. Encouraging human-AI collaboration in interactive learning environments[J]. Interactive Learning Environments, 2025, 33(2): 921-924. https://doi.org/10.1080/10494820.2025.2494914

[19] Tülübaş T, Demirkol M, Ozdemir T Y, et al. An interview with ChatGPT on emergency remote teaching: A comparative analysis based on human–AI collaboration[J]. Educational Process: International Journal, 2023, 12(2): 93. https://doi.org/10.22521/edupij.2023.122.6

[20] Islam M T, Sepanloo K, Woo S, et al. A review of the industry 4.0 to 5.0 transition: exploring the intersection, challenges, and opportunities of technology and human–machine collaboration[J]. Machines, 2025, 13(4): 267. https://doi.org/10.3390/machines13040267

[21] Cooper G. Examining science education in ChatGPT: An exploratory study of generative artificial intelligence[J]. Journal of science education and technology, 2023, 32(3): 444-452. https://doi.org/10.1007/s10956-023-10039-y

[22] Zhang H, Jin X, Cui X. Exploration of Course Resources and Modes under Generative Artificial Intelligence[J]. Frontiers in Educational Innovation and Research. 2025, 1(1):4-9. https://doi.org/10.62762/FEIR.2024.649562

[23] Sarnato A Z, Sari W D, Rahmawati S T, et al. The evolution of e-learning platforms: From u-learning to ai-driven adaptive learning systems[J]. Journal of Social Science Utilizing Technology, 2024, 2(2): 289-300. https://doi.org/10.70177/jssut.v2i2.1131

[24] Zhong X, Zhan Z. An intelligent tutoring system for programming education based on informative tutoring feedback: system development, algorithm design, and empirical study[J]. Interactive Technology and Smart Education, 2025, 22(1): 3-24. https://doi.org/10.1108/ITSE-09-2023-0182

[25] Boulkroune A, Zouari F, Boubellouta A. Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems[J]. Journal of Vibration and Control, 2025: 10775463251320258. https://doi.org/10.1177/10775463251320258

[26] Boulkroune A, Hamel S, Zouari F, et al. Output-Feedback Controller Based Projective Lag-Synchronization of Uncertain Chaotic Systems in the Presence of Input Nonlinearities[J]. Mathematical Problems in Engineering, 2017, 2017(1): 8045803. https://doi.org/10.1155/2017/8045803

[27] Wu J, Sun W, Su SF, Xia J. Neural-based adaptive control for nonlinear systems with quantized input and the output constraint[J]. Applied Mathematics and Computation. 2022, 413:126637. https://doi.org/10.1016/j.amc.2021.126637

[28] Rigatos G, Abbaszadeh M, Sari B, et al. Nonlinear optimal control for a gas compressor driven by an induction motor[J]. Results in Control and Optimization, 2023, 11: 100226. https://doi.org/10.1016/j.rico.2023.100226

[29] Li X, Lu X, Chen W, et al. Research on UAVs reconnaissance task allocation method based on communication preservation[J]. IEEE Transactions on Consumer Electronics, 2024, 70(1): 684-695. https://doi.org/10.1109/TCE.2024.3368062

[30] Ooi K B, Tan G W H, Al-Emran M, et al. The potential of generative artificial intelligence across

disciplines: Perspectives and future directions[J]. Journal of Computer Information Systems, 2025, 65(1): 76-107. https://doi.org/10.1080/08874417.2023.2261010

[31] Strielkowski W, Grebennikova V, Lisovskiy A, et al. AI-driven adaptive learning for sustainable educational transformation[J]. Sustainable Development, 2025, 33(2): 1921-1947. https://doi.org/10.1002/sd.3221

Appendix A. Data format description and field definitions

Field Name	Data Type	Description	Example Value	Processing Method
user_id	string	Unique learner identifier (anonymized)	"A35F2K1D"	Hashed to protect privacy
timestamp	datetime	Operation timestamp (ISO 8601 format)	"2024-05-16T09:23:45.128Z"	Standardized to UTC format
event_type	string	Type of interaction event (e.g., login, video view, AI query)	"AI_query"	Grouped by type for statistics
session_id	string	Learning session identifier	"S19283"	Used to calculate average session duration
duration	float	Duration of a single interaction (seconds)	124	Outliers removed (>3σ)
course_id	string	Course code	"CS102"	Used for cross-course analysis
content_id	string	Learning content or resource identifier	"Lec03_Video"	Linked to course content table
completion_rate	float	Content completion rate (0–1)	0.93	Missing values smoothed
interaction_count	int	Number of interactions in the current session	5	Normalized to per-hour unit
feedback_score	float	User feedback score on content or AI assistance (1–5)	4.7	Used to compute average satisfaction