DESIGN AND APPLICATION OF HUMAN-COMPUTER COLLABORATIVE TEACHING INTERACTIVE SYSTEM DRIVEN BY ARTIFICIAL INTELLIGENCE

Zhijuan WANG[1, *]

It is challenging for teachers to quickly and accurately assess each student's learning status due to the ongoing growth in classroom size and the complexity of the curriculum in contemporary educational settings. This limits classroom interaction and the efficacy of individualized instruction. This study uses a Transformer-based multi-modal fusion model (Multi-Modal Transformer, MMT) to build and implement a human-computer collaborative teaching interaction system to address this problem. By fusing multi-modal data such as classroom Q&A, student behavioral actions, and grades and participation, this study realizes real-time recognition of students' learning status and personalized teaching recommendations. Combined with teacher collaboration strategies, it dynamically adapts to classroom changes through online incremental learning. This study conducts experiments in three university classes with a total of 180 students. According to the experimental data, the experimental group's average score, 79.8 points, compared to 72.1 points for the control group, is noticeably higher than that of the students in the control group. Classroom interactions are about doubled in frequency and engagement. With an average of roughly 4.7 points, teachers in the experimental group scored significantly higher than those in the control group on the following categories: classroom interaction improvement, teaching assistance effect, ease of use of the system, and recognition of personalized suggestions. This demonstrates that the MMT approach offers notable benefits in enhancing classroom engagement, student learning outcomes, and teachers' teaching experiences. It also serves as a workable, useful guide for the intelligent and individualized development of AI-driven education.

Keywords: Artificial intelligence; Multi-modal fusion; Transformer; Human-computer collaborative teaching; Individualized teaching.

1. Introduction

In modern educational environments, traditional classroom teaching faces numerous challenges [1]. With the continuous expansion of class sizes and the increasing complexity of teaching content, it is difficult for teachers to conduct timely and accurate assessment and guidance on each student’s learning situation [2]. This limitation not only results in insufficient classroom interactivity but also makes it hard to meet students’ personalized learning needs, thereby affecting

[1] Jiangxi Technical College of Manufacturing, Nanchang 330000, China, e-mail: 19979087987@163.com

learning efficiency and overall participation [3]. Furthermore, teachers find it difficult to fully understand each student's comprehension and learning status throughout class. This problem is especially noticeable when teaching in big courses and also limits the flexibility to modify and optimize teaching tactics [4, 5].

The quick advancement of artificial intelligence (AI) technology in the realm of education in recent years has opened up new avenues for resolving issues with conventional teaching [6]. The application of intelligent tutoring system, educational robots, virtual teachers, and learning behavior prediction technology can assist teachers in making teaching decisions and managing students to a certain extent, thereby realizing personalized teaching and real-time feedback [7, 8]. However, most current educational systems still have obvious shortcomings in real-time interaction, personalized recommendation, and teacher-assisted decision-making. At the same time, they lack the ability to systematically analyze and deeply integrate multi-modal data (such as classroom Q&A, students' behavioral movements, facial expressions, grades, and participation) [9, 10]. This restriction limits the use of AI technology in traditional classroom instruction as well as its ability to enhance student learning and teaching effectiveness [11-13].

To address this issue, this study proposes and designs an AI-based human-computer collaborative teaching interaction system. Through the cooperation of AI technology and educators, it seeks to provide real-time feedback and individualized instruction, improving learning outcomes and classroom engagement. The core innovation of the system lies in the adoption of a self-attention mechanism for multi-modal data fusion. The system designs a feedback mechanism for collaboration between teachers and AI: AI generates personalized teaching suggestions, and teachers can select and adjust them according to the actual classroom situation. The feedback is then used to update the model online, enabling the system to continuously adapt to the needs of different classroom environments and student groups.

2. Literature Review

AI's use in education has drawn a lot of attention in recent years. Vujinović et al. [14] pointed out that intelligent tutoring systems could model students' learning processes to achieve personalized tutoring and learning path recommendation, thereby improving learning effects and teaching efficiency. According to Hwang et al. [15], adaptive learning systems had the ability to dynamically modify the content and level of difficulty of instruction based on students' learning outcomes, hence improving the targeting of the learning process. In addition, educational robots and virtual teachers had shown certain advantages in classroom interaction, knowledge transfer, and learning motivation stimulation, especially in online education and blended teaching environments, they can enhance students' learning experience and sense of classroom participation [16]. However, most studies still focus on a single teaching function or single-modal data processing (e.g., only analyzing students' answer records or behavioral data), lacking the comprehensive utilization of multi-modal information [17].

In research on human-computer collaborative teaching, Dai et al. [18] indicated that AI could assist teachers in making teaching decisions or automatically generating teaching suggestions. For instance, some studies have attempted to predict academic performance through student behavior monitoring or learning data [19]. As Fitria [20] pointed out, such systems generally had limitations in practical classrooms. In addition, existing studies involve insufficient systematic evaluation in practical applications and lack empirical analysis on the improvement of classroom interaction effects and student participation [21].

To sum up, although existing studies have achieved certain results in intelligent tutoring, behavior prediction, and human-computer collaborative teaching, there are still obvious research gaps. As noted by Chango et al. [22] and Wang et al. [23], the multi-modal information fusion capability was insufficient, failing to systematically integrate data such as classroom text, behavior, facial expressions, and academic performance. Meanwhile, the collaborative strategy between AI and teachers had not yet formed a closed-loop mechanism, lacking real-time adaptation and feedback update [24]. To achieve effective teacher-AI collaboration, this study attempts to build a multi-modal data fusion model based on the self-attention mechanism. It also verifies the model's role in improving teaching interactivity, student participation, and learning effects in real classroom environments, thereby filling the gaps existing in current research.

3. Research Methodology

3.1 System Design

The proposed overall architecture of the human-computer collaborative teaching interaction system (as shown in Fig. 1) consists of a teacher terminal, a student terminal, an AI module, and a data analysis & interaction module. The teacher terminal is mainly used to receive personalized teaching suggestions generated by AI and select and adjust them according to the actual classroom situation. The student terminal focuses on collecting students' behavioral performance and learning feedback in class, which serves as an important basis for AI decision-making. The primary functions of data processing and intelligent decision-making are handled by the AI module. It offers teachers and students real-time assistance by fusing and modeling multi-modal data. The data analysis & interaction module is responsible for integrating and visually displaying the collected data (such as classroom Q&A, behavioral features, academic performance, and participation), thereby forming an operable teaching feedback loop. Through this architecture, the system can realize comprehensive collection of classroom interactions, real-time push of instant feedback, and dynamic generation of personalized teaching recommendations.

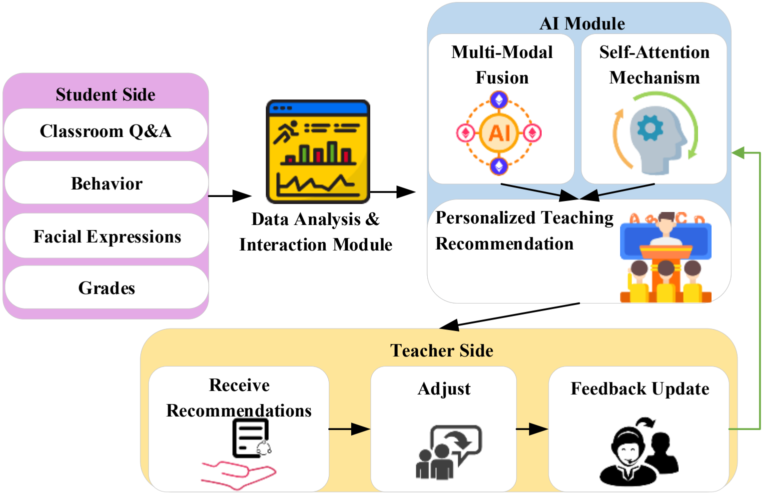


Fig. 1. Schematic diagram of system architecture

3.2 AI-driven Human-Computer Collaboration Mechanism

In the human-computer collaborative teaching interaction system of this study, the core of the AI module is a Transformer-based multi-modal fusion model (Multi-Modal Transformer, MMT). The system operation mechanism mainly includes three key links: multi-modal fusion, model optimization, and teacher collaboration.

First, in terms of multi-modal data fusion, the system collects three types of data sources: text modality, behavior modality, and academic performance & classroom participation modality. The text modality includes classroom Q&A, discussion records, and teacher comments. The system captures the semantic information and expression logic of students' response material by encoding each text passage using a lightweight Bidirectional Encoder Representations from Transformers (BERT) model to produce semantic vectors with defined dimensions. Nonverbal cues including students' body language, gestures, and changes in sitting position during class are all covered by the behavior modality. The system performs frame-by-frame feature extraction on video data through Convolutional Neural Network (CNN) to obtain a sequence of behavioral feature vectors, which are used to characterize students’ attention status and emotional changes. The academic performance & classroom participation modality includes indicators such as homework scores, test marks, and frequency of classroom questions, answers, and group discussion participation. These data are standardized into numerical vectors to facilitate unified processing with text and behavioral features. During the fusion process, the feature vectors of the three modalities extract potential patterns through their respective encoders, providing input for the subsequent self-attention mechanism and realizing comprehensive modeling and correlation analysis of cross-modal information.

Subsequently, the pre-processed text, behavioral, and performance feature vectors are mapped to a unified dimension and fed into the MMT model. This study adopts the standard multi-head self-attention mechanism proposed by Belagur et al. [25] to process these multimodal sequences. Under the action of this mechanism, the model can calculate the interdependencies between different modal features, thereby dynamically capturing cross-modal semantic associations. For instance, when a student's behavioral features (e.g., confused facial expressions) and text features (e.g., incorrect answers) appear simultaneously, the attention mechanism automatically assigns higher weights to these associated features, enabling more accurate characterization of the student's real learning state. Through such in-depth feature interaction and fusion, the high-dimensional fused vector output by the system provides context-rich decision support for subsequent personalized teaching recommendations.

In this way, the model can adaptively assign weights to different modalities. For example, when a student answers a question correctly but shows anxious facial expressions, the model will increase attention to the behavior modality, thereby capturing the student's real learning status more accurately.

Secondly, in the aspect of model optimization strategy, online incremental learning mechanism is introduced in this study, so that the model parameters can be continuously updated with the classroom process. For each batch of newly input data "(" "x" _"t" "," "y" _"t" ")" , the model parameters are updated according to the gradient descent rule:

$$\theta_{t+1}=\theta_t-\eta\nabla L(f(x_t;\theta_t),y_t) \quad (1)$$

η is the learning rate. L(·) is the loss function. f(·) is the prediction function of MMT. To meet the individualized demand, the individualized weight α_i of student groups is introduced into the loss function, which is used to distinguish the learning characteristics of different student groups and make the optimization process more personalized:

$$L=\sum_i \alpha_i \cdot l(f(x_i;\theta),y_i) \quad (2)$$

ℓ(·) is the basic loss function, and α_i is used to amplify or weaken the influence of certain students' learning behavior in model optimization.

Finally, in the teacher collaboration phase, the personalized teaching suggestions generated by the AI module are first presented to the teacher. The teacher then screens and adjusts these suggestions based on teaching objectives and actual classroom conditions. The teacher's selections and feedback are recorded in real time and sent back to the MMT model as part of the training signals, which in turn influences the model's subsequent inference. The overall operational logic of the system is illustrated in Fig. 2 (schematic diagram of human-computer collaboration mechanism).

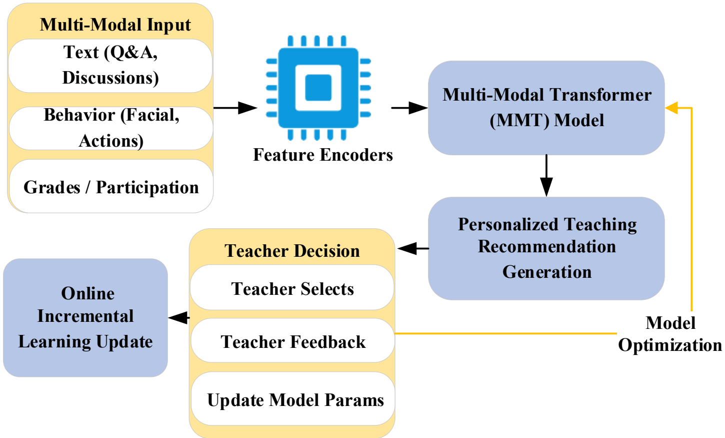


Fig. 2. Schematic diagram of human-computer collaboration mechanism

3.3 Experimental Design

To verify the effectiveness of the MMT model in real classroom environments, this study conducts experiments in three university classes with a total of 180 students, covering different grades and majors to ensure the broad applicability of the results. A within-class random grouping method is adopted to divide students in each class into an experimental group and a control group. The specific operations are as follows: First, students are numbered according to the class list. Then, random drawing or random number generation is used to evenly assign students to the experimental group and the control group to ensure that the two groups are as balanced as possible in terms of gender ratio, professional background, and basic academic performance. Taking the three university classes as examples: Class A has 60 students, among whom 30 are randomly assigned to the experimental group (using the AI system) and 30 to the control group (adopting traditional classroom teaching methods). Class B has 56 students, with 28 in the experimental group and 28 in the control group. Class C has 64 students, including 32 in the experimental group and 32 in the control group. This random grouping method can minimize the impact of individual differences on experimental results, ensure the comparability of the experimental group and the control group in terms of initial conditions, and thus scientifically evaluate the effects of the AI-driven human-computer collaborative system on classroom interaction, student performance, participation, and other aspects.

The experimental group used the AI-driven human-computer collaborative teaching system, while the control group adopted traditional classroom teaching methods. During the experiment, the system collects multi-source data from the student terminal, including classroom interaction records, student behavior logs, and questionnaires. Classroom interaction records are used to measure students' participation in asking questions, answering questions, and discussions. Behavior logs capture students' movement and facial expression features via cameras or sensors to characterize their learning status and emotional changes. Questionnaires are used to obtain students' learning experience and sense of participation.

The questionnaire survey covers three dimensions: students' learning experience, sense of classroom participation, and satisfaction with system use. The teacher questionnaire focuses on the system's ease of use, auxiliary effect on classroom teaching, recognition of personalized suggestions, and observational evaluation of improvements in classroom interaction. All data are collected with the informed consent of students and teachers and anonymized to ensure that data use complies with the norms of educational institutions.

In terms of experimental evaluation, this study comprehensively adopts four types of indicators to systematically measure the effectiveness of the AI-driven human-computer collaborative teaching interaction system: Improvement of students' academic performance, frequency of classroom interaction, student participation, and teacher satisfaction. The overall experimental process is illustrated in Fig. 3 (schematic diagram of experimental design and data processing flow).

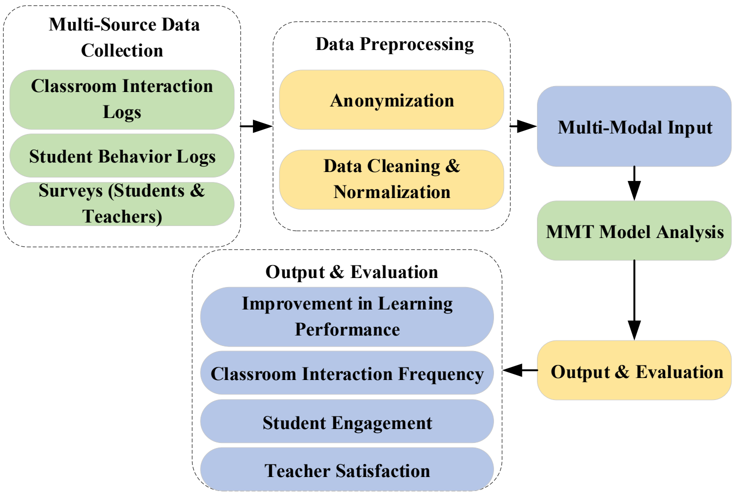


Fig. 3. Schematic diagram of experimental design and data processing flow

To ensure this study complies with academic ethics and fully protects the privacy rights of participants, the research team strictly implement an ethical review process and established a comprehensive data protection mechanism prior to the study initiation. All participating students and teachers are fully informed of the research objectives, scope of data collection, and usage methods before the experiment, and signed informed consent forms. These forms explicitly authorized the system to collect their classroom behaviors and facial expression data for educational analysis, while preserving their right to withdraw from the experiment at any time. Addressing the reviewers' concerns regarding potential psychological discomfort caused by facial video recording and its possible impact on learning progress, the study adopted a non-intrusive and unobtrusive data collection strategy. Instead of installing additional prominent devices, existing classroom cameras are utilized for edge computing processing. This effectively avoids interference from the Hawthorne effect, allowing students to maintain natural classroom performances. Furthermore, all collected facial image data undergo real-time feature extraction and anonymization desensitization on local servers. The system only stores and uploads digitized feature vectors that cannot be reversed to specific facial images, fundamentally eliminating the risk of privacy leakage. Feedback from actual classroom observations indicated that, thanks to this transparent and secure ethical protection mechanism, students eliminated anxiety about being monitored after a short adaptation period. No distractions or resistance due to technical intervention are observed, thus ensuring that normal learning progress and classroom participation remained unaffected.

To ensure the reproducibility of experimental results, this study implements the MMT model based on the PyTorch deep learning framework and conducts training on a server equipped with an NVIDIA RTX 3090 GPU. During the data preprocessing phase, all input sequences are padded or truncated to a fixed length. The model training adopts the Adam optimizer, combines with a cosine annealing strategy for dynamic adjustment of the learning rate. To prevent overfitting, Dropout technology is applied after both the multi-head attention layers and feed-forward network layers. Detailed configurations of model hyperparameters are presented in Table 1.

Table 1

Key hyperparameter setting for MMT model training

Parameter name	Value	Explanation
Batch Size	32	Training batch size
Learning Rate	1e-4	Initial learning rate
Epochs	100	Total training rounds
Attention Heads	8	The number of heads in the multi-head attention mechanism
Encoder Layers	4	Transformer encoder layer number
Hidden Dimension	256	Hidden layer dimension of feature vector
Dropout Rate	0.2	Rejection rate of preventing over-fitting
Optimizer	Adam	Optimization algorithm ($\beta_1 = 0.9, \beta_2 = 0.999$)

4. Result and Discussion

4.1 The Performance Analysis of Teaching System

To evaluate the application effect of the Transformer-based multi-modal fusion model in the AI-driven human-computer collaborative teaching system, this study analyzes the model performance by comparing the experimental group (using the AI collaborative system) with the control group (adopting traditional teaching methods). The results are shown in Fig. 4. Regarding the performance of the multi-modal fusion model under different modal combinations: The accuracy of single-modal models (only text, behavior, or performance/participation) ranges from 0.70 to 0.78, and their F1-scores are slightly lower than the accuracy. This indicates that comprehensively considering classroom Q&A, students' behavioral movements, as well as performance and participation information can more accurately capture students' learning status and potential interaction patterns, providing a reliable basis for personalized teaching.

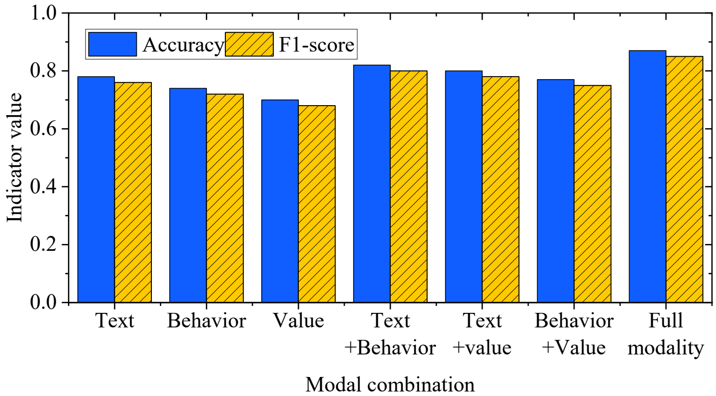


Fig. 4. Performance analysis of teaching system

To further verify the rationality of selecting the Transformer architecture, this study conducts comparative experiments between the MMT model and three mainstream temporal processing models under the condition of full-modal input (text + behavior + performance/participation). The comparative models include: 1) Long Short-Term Memory (LSTM): a classic RNN variant proficient in processing sequential data; 2) Gated Recurrent Unit (GRU): a Recurrent Neural Network (RNN) with a simpler structure than LSTM; 3) CNN-LSTM: a hybrid model combining CNN for feature extraction and LSTM for temporal processing. The comparative results are presented in Table 2.

Table 2

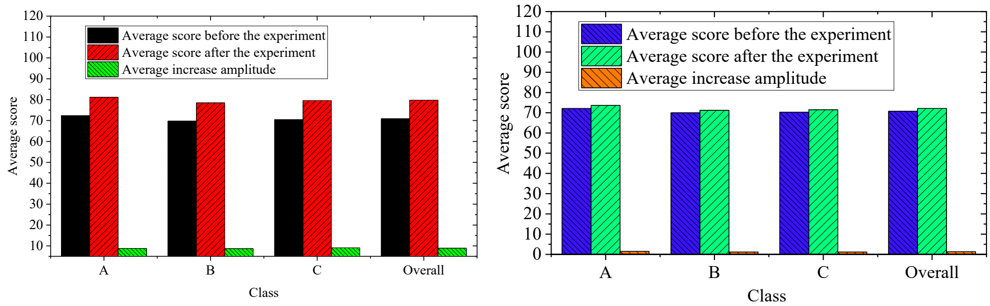
Performance comparison of different model architectures on multimodal teaching data

Model architecture	Accuracy	Precision	Recall rate	F1 score
LSTM	0.76	0.74	0.75	0.74
GRU	0.75	0.73	0.74	0.73
CNN-LSTM	0.81	0.79	0.80	0.79
MMT (Ours)	0.87	0.86	0.85	0.85

Analysis of Table 2 reveals that the simple LSTM and GRU models significantly underperform the proposed MMT model in terms of accuracy and F1-score (LSTM achieves an accuracy of 0.76, compared to 0.87 for MMT). This result indicates that while LSTM is effective in processing short sequences, its ability to capture long-range dependencies is limited when confronted with 45-minute complex classroom interaction data. Consequently, it struggles to establish effective correlations between students' early-class states and late-class behavioral patterns. In contrast, CNN-LSTM improves feature extraction capability by introducing convolutional layers, leading to enhanced performance (accuracy of 0.81), but still lags behind MMT. The superiority of the MMT model is primarily attributed to its self-attention mechanism. Unlike the sequential processing of RNN-based models, MMT can process multimodal sequences in parallel and dynamically calculate global correlation weights among text, behavioral, and performance features. This enables the model not only to identify local abnormal behaviors but also to conduct in-depth reasoning by integrating contextual information. Thus, experiments demonstrate that adopting the Transformer architecture is not redundant in complex human-machine collaborative teaching scenarios, but rather a necessary choice for achieving high-precision personalized recommendations.

4.2 Students' Academic Performance

Fig. 5 displays the academic performance comparison results between the experimental and control groups. The experimental results show that students in the experimental group significantly improve their scores after utilizing the AI system, whereas the control group's score gain is little.



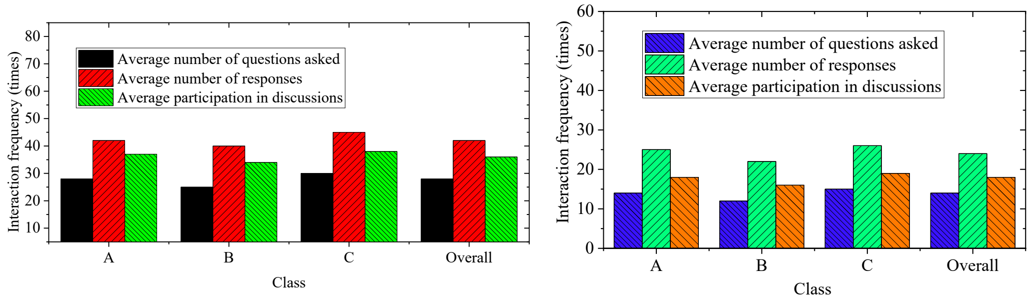
(a) Experimental group (b) Control group

Fig. 5. Comparison of students' academic performance between the control group and the experimental group (score/percentage system)

This demonstrates the system's usefulness in supporting learning. For instance, the control group's average performance in the same class only improves by roughly 1.5 points, from 72.1 to 73.6. While the control group's average score only rises by roughly 1.3 points, the experimental group's average score typically rises by about 8.9 points. This shows that the personalized teaching suggestions generated by the AI system, combined with teachers' collaborative adjustments, can significantly enhance students' knowledge mastery and learning effects. The improvement is even more obvious, especially for students with lower initial learning levels or lower participation.

4.3 Classroom Interaction Frequency

The number of students' questions, answers, and discussions is counted based on classroom interaction records, as shown in Fig. 6. The experimental group engages in a much greater number of interactions than the control group, particularly when participating actively in group discussions. Analysis of classroom interactions reveals that students in the experimental group participated in discussions, asked a lot more questions, and provided responses than students in the control group. In the experimental group, for instance, there are typically around 28 questions per class hour. There are roughly forty-two responses. About 36 people have participated in the debate. In contrast, the comparable numbers in the control group are roughly 14, 24, and 18, respectively. The experimental group's students participate more actively in class discussions and inquiries. This indicates that the AI system can effectively guide students to participate in classroom interactions. At the same time, teachers adjust their strategies based on system feedback, making classroom discussions more targeted and livelier, thereby improving the overall teaching interaction effect.

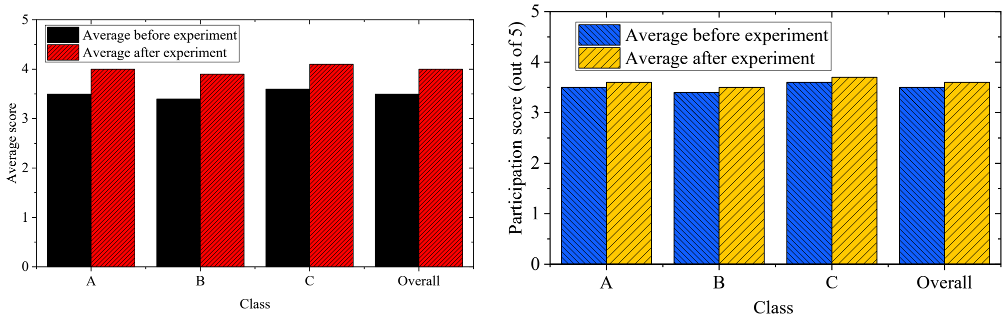


(a) Experimental group (b) Control group

Fig. 6. Comparison of classroom interaction frequency

4.4 Student Participation

Student participation is evaluated by integrating behavior logs and questionnaire scores, which are represented on a 1-5 point Likert scale, as shown in Fig. 7. The experimental group's total participation is substantially higher than the control group's, according to a thorough assessment of student participation that combines behavior logs and questionnaire results. While the control group's classroom engagement only increased from 3.5 to 3.6 points, the experimental group's increased from an average of 3.5 points before to the trial to 4.0 points. In particular, students with originally low participation shows a significant increase in engagement in the experimental group.



(a) Experimental group (b) Control group

Fig. 7. Students' participation in class

This indicates that the multi-modal fusion model can recognize students' status in real time, generate personalized learning guidance, and effectively improve students' classroom engagement and active learning behaviors.

4.5 Teacher Satisfaction

Teachers rated the system in terms of ease of use, teaching assistance effect, recognition of personalized suggestions, and evaluation of classroom interaction improvement, as shown in Tables 3-4.

Table 3

Teachers' satisfaction score of experimental group (1-5)

Teacher number	System usability	Teaching auxiliary effect	Recognition of personalized suggestions	Evaluation of classroom interaction improvement
T1	4.5	4.7	4.6	4.8
T2	4.3	4.6	4.5	4.7
T3	4.4	4.8	4.6	4.8
T4	4.5	4.7	4.7	4.9
Average	4.43	4.7	4.6	4.8

Table 4

Teachers' satisfaction score of control group (1-5)

Teacher number	System usability	Teaching auxiliary effect	Recognition of personalized suggestions	Evaluation of classroom interaction improvement
T1	3.8	3.9	3.7	3.8
T2	3.9	4.0	3.8	3.9
T3	3.7	3.8	3.6	3.7
T4	3.8	3.9	3.7	3.8
Average	3.8	3.9	3.7	3.8

The results indicate that teachers in the experimental group gave scores ranging from 4.3 to 4.9 (with an average of approximately 4.7) for the system's ease of use, teaching assistance effect, recognition of personalized suggestions, and improvement of classroom interaction. Teachers generally agreed that the AI system is easy to operate, could generate personalized teaching suggestions based on each student's learning status, and effectively assisted in classroom management, enabling more accurate adjustments to teaching strategies. The system's real-time data and feedback enable teachers to differentiate instruction based on the various needs of their students, increasing classroom productivity and interaction quality. Overall, the human-computer collaborative system not only optimizes teaching decision-making but also enhances teachers' confidence in classroom management and teaching effects, making the teaching experience more efficient and satisfying.

In conclusion, a comprehensive analysis of the model performance and experimental results reveals the following: First, the full-modal Transformer model achieves the best performance in capturing multi-source student information, predicting learning status, and generating personalized suggestions. Second, in terms of student involvement, classroom interaction, and academic performance, the experimental group performs noticeably better than the control group; students who initially participated poorly showed the biggest improvement. Third, the increase in teacher satisfaction indicates that the human-computer collaboration strategy can fully leverage the personalized suggestions generated by AI while maintaining teachers' dominant role, thereby optimizing teaching decision-making and classroom experience. Overall, the results verify the effectiveness and operability of the human-computer collaborative teaching system based on the MMT in university classrooms.

4.6 Cost-Benefit Analysis of the System

Addressing the economic viability and sustainability of AI systems in practical applications, this study focuses on improving learning outcomes and conducts a quantitative evaluation of the system's maintenance costs and comprehensive benefits. While the average score improvement of students in the experimental group (approximately 7.7 points) is statistically significant, to verify its rationality relative to the input costs, the resource consumption and outputs between the traditional teaching model and the AI human-machine collaborative model are compared. The results are presented in Table 5. Although the introduction of the AI system increases a small amount of hardware depreciation and server operation costs (classified as "low" maintenance costs due to the system's online incremental learning capability, which eliminates the need for frequent manual intervention and retraining), the benefits it brings are multi-dimensional.

Table 5

Cost-benefit comparison between traditional teaching and AI collaborative teaching

Evaluation index	Traditional teaching mode	AI collaborative teaching mode	Amplitude of variation
Average score improvement	+1.5 points	+7.7 points	Increase by 413%

Teachers' weekly feedback time	8.5 hours	3.2 hours	Reduce by 62%
System maintenance cost	0 (manpower only)	Low (server +electricity fee)	Slight increase
Teaching output per unit time	0.17 minutes/hour	2.40 minutes/hour	Significantly improved efficiency

In terms of direct gains, the learning improvement of the experimental group (+7.7 points) is significantly higher than that of the control group, confirming the system's effectiveness in enhancing teaching quality. Additionally, regarding implicit costs, the AI system takes over tedious data analysis and preliminary feedback tasks, reducing teachers' weekly time spent on academic performance feedback from 8.5 hours to 3.2 hours. This means teachers can devote more energy to high-value personalized tutoring. Finally, by calculating the "teaching output per unit time" (average score improvement divided by teachers' input time), the efficiency of the AI collaborative model is more than 14 times that of the traditional model. Thus, despite the system's basic operation and maintenance costs, the additional investment is fully reasonable and cost-effective from an educational economics perspective, considering its significant reduction in labor costs and substantial improvement in teaching efficiency.

5. Conclusion

This study designs and implements a human-computer collaborative teaching interaction system based on a Transformer-based multi-modal fusion model. By fusing multi-modal data such as classroom Q&A, students' behavioral movements, academic performance, and participation, the system realizes real-time identification of students' learning status and personalized teaching recommendations. Through experiments conducted in university classrooms, the system's usefulness and operability in real-world teaching settings are confirmed by the experimental results, which demonstrate that it may greatly increase students' learning effects and classroom involvement while also improving teachers' teaching experiences and satisfaction. However, this study still has certain limitations. The experimental subjects are mainly university students, the classroom types are concentrated on theoretical courses, and the sample size is limited. Thus, the universality of the results needs further verification. Meanwhile, this study mainly collects data on classroom Q&A, behavioral movements, academic performance, and participation, without fully considering factors such as students' psychological status, emotions, or family environment. The experimental cycle is short, and the stability and sustainability of the system in long-term teaching has not yet been examined.

Future research could further expand the sample size to cover different grades, disciplines, and educational stages, aiming to verify the system's applicability in a broader range of teaching scenarios. Additionally, incorporating multi-modal data related to psychology, emotions, and physiology could enhance the model's ability to recognize individual differences among students. Exploring model optimization strategies based on reinforcement learning may also enable a higher level of intelligent human-computer collaborative teaching. These efforts will contribute to the deep integration and long-term development of AI in the field of education.

REFERENCES

[1] Martin-Alguacil N, Avedillo L, Mota-Blanco R, et al. Student-centered learning: some issues and recommendations for its implementation in a traditional curriculum setting in health sciences. Education Sciences, 2024, 14(11): 1179.

[2] Purike E, Aslan A. A comparison of the effectiveness of digital and traditional learning in developing countries. Indonesian Journal of Education, 2025, 5(1): 179-186.

[3] Mulenga R, Shilongo H. Hybrid and blended learning models: Innovations, challenges, and future directions in education. Acta Pedagogia Asiana, 2025, 4(1): 1-13.

[4] Yeganeh L N, Fenty N S, Chen Y, et al. The future of education: A multi-layered metaverse classroom model for immersive and inclusive learning. Future Internet, 2025, 17(2): 63.

[5] Mesuwini J, Mokoena S. Exploring online teaching and learning challenges for the technical and vocational education and training lecturer. Journal of Education and E-Learning Research, 2024, 11(1): 193-202.

[6] Tan X, Cheng G, Ling M H. Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence, 2025, 8: 100355.

[7] Ateş H. Integrating augmented reality into intelligent tutoring systems to enhance science education outcomes. Education and Information Technologies, 2025, 30(4): 4435-4470.

[8] Zhong X, Zhan Z. An intelligent tutoring system for programming education based on informative tutoring feedback: system development, algorithm design, and empirical study. Interactive Technology and Smart Education, 2025, 22(1): 3-24.

[9] Sperling K, Stenberg C J, McGrath C, et al. In search of artificial intelligence (AI) literacy in teacher education: A scoping review. Computers and Education Open, 2024, 6: 100169.

[10] Díaz B, Nussbaum M. Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. Computers & Education, 2024, 217: 105071.

[11] Yim I.H.Y., A critical review of teaching and learning artificial intelligence (AI) literacy: Developing an intelligence-based AI literacy framework for primary school education. Computers and Education: Artificial Intelligence, 2024, 7: 100319.

[12] Almasri F. Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. Research in Science Education, 2024, 54(5): 977-997.

[13] Bower M., Torrington J, Lai J.W.M., et al. How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. Education and Information Technologies, 2024, 29(12): 15403-15439.

[14] Vujinović A, Luburić N, Slivka J, et al. Using ChatGPT to annotate a dataset: A case study in intelligent tutoring systems. Machine Learning with Applications, 2024, 16: 100557.

[15] Hwang G J, Sung H Y, Chang S C, et al. A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. Computers and Education: Artificial Intelligence, 2020, 1: 100003.
[16] Ying Z. Experience of intelligent speech robot in music online classroom based on deep learning and virtual reality. Entertainment Computing, 2025, 52: 100795.
[17] Ahumada-Newhart V, Eccles J S. Robots for learning: An exploration of teacher roles, perceptions, and challenges in robot-mediated learning. Frontiers in Robotics and AI, 2025, 12: 1441382.
[18] Dai R, Thomas M K E, Rawolle S. The roles of AI and educational leaders in AI-assisted administrative decision-making: a proposed framework for symbiotic collaboration. The Australian Educational Researcher, 2025, 52(2): 1471-1487.
[19] Kim J. Leading teachers' perspective on teacher-AI collaboration in education. Education and Information Technologies, 2024, 29(7): 8693-8724.
[20] Fitria T N. The use of artificial intelligence in education: Can AI replace the teacher's role?. Epigram, 2023, 20(2): 165-187.
[21] Tammets K, Ley T. Integrating AI tools in teacher professional learning: a conceptual model and illustrative case. Frontiers in artificial intelligence, 2023, 6: 1255089.
[22] Chango W, Lara J A, Cerezo R, et al. A review on data fusion in multimodal learning analytics and educational data mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2022, 12(4): e1458.
[23] Wang Z, Deng Y, Zhou S, et al. Achieving sustainable development goal 9: A study of enterprise resource optimization based on artificial intelligence algorithms. Resources Policy, 2023, 80: 103212.
[24] Dong Q, He J, Li N, et al. Exploring the cognitive reconstruction mechanism of generative AI in outcome-based design education: A study on load optimization and performance impact based on dual-path teaching. Buildings, 2025, 15(16): 2864.
[25] Belagur H, Reddy N S, Krishna P R, et al. Cross-modal multi-headed attention for long multimodal conversations[J]. Multimedia Tools and Applications, 2023, 82(29): 45679-45697.