International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6566-6575

https://doi.org/10.70517/ijhsa463564

A study of the moderating role of emotion perception on learners' behavioral patterns in artificial intelligence-assisted education based on an intelligent sentiment analysis model

Siyu Li[1], Wendi Duan[1], Lifeng Yang[2], Zhenyi Li[3] and Huan Liao[4,*]
[1] Education School, University of Glasgow, Glasgow, G128QQ, United Kingdom
[2] Hongyun Honghe Group Kunming Cigarette Factory, Kunming, Yunnan, 650231, China
[3] Education school, University of Nakhon Phanom, Nakhon Phanom, 48000, Thailand
[4] School of Foreign Languages, Wuyi University, Jiangmen, Guangdong, 529020, China
Corresponding authors: (e-mail: alisali1227@163.com).

Abstract With the in-depth application of artificial intelligence technology in the field of education, intelligent teaching systems have evolved from simple computer-assisted learning to complex systems capable of understanding and responding to learners' emotional states. This paper explores the moderating role of emotion perception on learners' behavioral patterns in AI-assisted education based on an intelligent sentiment analysis model. The study designs a bimodal sentiment analysis model GTA-BERT that retains text-sentence dependency analysis information and fuses text-speech masked attention, which consists of four parts: text feature extraction, speech feature extraction, DEGCN text-sentiment enhancement module and masked attention fusion. Through comparison experiments with mainstream sentiment analysis models and questionnaire surveys, the study verifies the effectiveness of the model in sentiment recognition and learning behavior regulation. The results show that the GTA-BERT model performs well in multimodal sentiment analysis, with Acc-2, F1, Acc2-weak, and Corr values of 93.24, 83.96, 75.01, and 72.67, respectively, which are the highest values among all the compared models. The empirical study confirmed the direct impact of AI-assisted instruction on learners' behavioral patterns, while emotional perception and mood played an important mediating role in it. The conclusion of the study shows that the intelligent sentiment analysis model can effectively identify learning emotions, while the application of emotion perception in the teaching process helps to regulate learners' behavioral patterns and improve the effectiveness of AI-assisted education.

Index Terms Artificial intelligence-assisted education, Emotion perception, Learner behavior pattern, Sentiment analysis model, GTA-BERT, Mediating effect

I. Introduction

With the continuous progress of big data, cloud computing, machine learning and other technologies, the application of artificial intelligence in the field of education not only brings unprecedented convenience and efficiency to education, but also changes the nature and mode of education on a deep level [1]. The traditional education model often focuses on teachers' one-way teaching and students' passive acceptance, while AI-assisted education enables students to learn independently according to their own learning characteristics and needs through intelligent and personalized teaching methods, thus stimulating students' interest and potential in learning, and improving students' satisfaction and learning results [2]-[4]. At the same time, AI-assisted education can break the limitations of geography, resources, etc., and can provide teachers and students with scientific and accurate learning assistance, help teachers better understand the learning situation of students, and develop more reasonable learning plans and strategies [5]-[7].

Emotion perception is an important direction for exploring student learning. And artificial intelligence in AI-assisted education can realize the emotional perception of students by analyzing data such as voice, text and facial expression. Through technologies such as deep learning and machine learning, AI can analyze the input data for sentiment analysis and emotion classification [8], [9]. For example, AI can determine whether a person is in an emotional state such as pleasure, sadness, or anger based on changes in his or her voice pitch and expression. Through deep learning generative models, AI can generate textual content with emotional colors according to the input text or context, such as emotional news headlines, literary works, etc. [10]. In addition, AI can also simulate human emotional expression and communication, for example, through chatbots to communicate with users emotionally, help users to relieve emotional stress and provide psychological support [11], [12]. And emotion as an important factor affecting students' behavioral change, through the perception of emotion to regulate their own behavior, as a way to seek a better learning mode and improve the learning effect, such as the secretion of dopamine can extend the students' learning time to a certain extent, and a certain degree of emotional ups and downs can cause knowledge transfer changes [13]-[15]. These studies have opened new ideas for emotion-based learning and educational technology, and by studying the moderating effect of emotion perception on students' behavioral patterns, AI-assisted education can be further optimized.

The application of artificial intelligence (AI) technology in education has evolved from early computer-assisted learning to today's intelligent tutoring systems and personalized learning platforms. In the 1950s, AI as an independent discipline began to focus on the creation of intelligent machines capable of problem solving, reasoning, and learning, but at that time, the integration with education was relatively limited. With the popularity of mobile devices and advances in computing technology, the application of AI in the field of teaching and learning has continued to expand, and a variety of intelligent computer-assisted teaching technologies such as natural language processing, user modeling, expert systems, and intelligent tutoring systems have gradually formed an independent research field. In recent years, the emergence of generative AI tools such as ChatGPT has further pushed the change of education technology, and AI-assisted teaching systems have transformed from simple knowledge transfer to deeply personalized learning experiences. This personalized learning experience not only requires the intelligent system to accurately assess the learner's knowledge state, but also requires the system to understand and respond to the learner's emotional state, because emotion, as an important factor in the learning process, directly affects the learner's cognitive activities and learning effects. Positive emotions can promote learners' cognitive activities and creative thinking, while negative emotions may lead to decreased learning efficiency and classroom engagement. However, traditional AI-assisted teaching systems often lack an effective perception and response mechanism for learners' emotions, resulting in overly mechanized interactions between the system and the learners, making it difficult to establish an effective emotional connection. Therefore, it is of great significance to construct intelligent models that can accurately perceive and analyze learners' emotional states, and to study the moderating effect of emotion perception on learning behavior, in order to improve the effectiveness of AI-assisted education.

In this study, a bimodal sentiment analysis model GTA-BERT, which retains the dependency analysis information of text sentences and incorporates text-speech masked attention, is designed to address the above problems, and the model realizes high-precision recognition of learners' emotional states by combining the information of two modalities, text and speech. The study first conducted unimodal and multimodal sentiment analysis experiments on the model to verify the effectiveness of the model in sentiment recognition. Then, we collected sample data from 300 AI-assisted teaching users through questionnaires, constructed a research model of “AI-assisted teaching-emotion perception-emotion-learners' behavioral patterns”, and used structural equation modeling to analyze the mediating roles of emotion perception and emotion in the process of AI-assisted teaching influencing learners' behavioral patterns. The results verify the application value of emotion perception technology in AI-assisted education, and provide theoretical and practical basis for improving the emotional interaction capability of intelligent teaching systems. This study not only enriches the theoretical research in the field of AI-assisted education, but also provides a useful reference for the development of more intelligent and humanized educational technology systems.

II. Intelligent Sentiment Analysis Model in Artificial Intelligence-Assisted Education

II. A. Artificial intelligence-assisted education

In the field of educational technology, the application of AI has gradually moved from simple computer-assisted learning to complex intelligent systems and algorithms. The 1950s marked the emergence of AI as a distinct discipline focused on the creation of intelligent machines capable of problem solving, reasoning, and learning from observation. Early AI technologies were mainly at the level of theoretical exploration, with more limited integration with education. The popularization of mobile devices has created conditions for the further application of AI technology, and the application of AI in teaching has gradually expanded. In order to better carry out personalized learning, various technologies of intelligent computer-assisted teaching have become a new field, mainly applying the concepts and technologies of artificial intelligence, such as natural language processing, user modeling, expert systems and intelligent tutoring systems. And in recent years, the rise of generative AI technology tools such as ChatGPT has further advanced the field of educational technology [16]. Artificial intelligence has gradually evolved into intelligent tutoring systems and personalized learning platforms that can analyze learner behavior and provide them with personalized feedback. Currently, the application of AI in teaching is gradually deepening, from simulated communication scenarios to personalized learning, and more innovative applications are likely to emerge in the future.

Artificial intelligence-assisted education provides students with a customized learning experience through smart technology, helping learners better master knowledge and skills. It helps students improve relevant skills by providing instant feedback and simulating real-life conversational situations. In addition, embedding learner-specific emotion recognition and analytics models in AI-assisted education systems helps to instantly understand the learner's learning status, so as to instantly adjust the teaching approach and content.

II. B.Learning emotions

Emotions, as the external expression of human inner feelings, reflect people's attitudes towards external things. In psychology, emotion is regarded as a psychological phenomenon reflecting the connection between object things and people's need relations and expectations, as well as people's own cognitive evaluation of these two connections. Positive emotional reactions are usually triggered when objective events coincide with the individual's desires or exceed expectations. And when these events fail to meet personal needs or fall short of expectations, negative emotional experiences are often triggered.

Emotions in learning encompasses a range of emotional responses that students encounter during educational activities, which may be related to a variety of aspects such as academic performance, classroom interactions, daily learning experiences, and testing environments. In-depth analyses of these emotions are extremely valuable to teachers because they can reveal learners' states of learning, motivating factors, points of interest, and intensity of concentration.

Learning emotions have a significant impact on students' cognitive activities in the classroom, with positive emotions facilitating cognitive activities and negative emotions hindering cognitive activities, decreasing learning efficiency and classroom engagement, and even disengaging from teaching and learning activities. Emotions are important to the cultivation of students' creative thinking, and happy emotional experience can promote the development of students' creative thinking, which is mainly manifested in the fluency and flexibility of students' divergent thinking. Anxious emotional experience, on the other hand, can inhibit the performance of students' creative thinking.

II. C.Bimodal Sentiment Analysis Model Based on Deep Learning

The general flow of the text-speech bimodal sentiment classification problem is shown in Figure 1, which can be divided into three parts: text word embedding and feature selection and extraction, speech quantization and feature selection and extraction, and fusion of the two features and training classification.

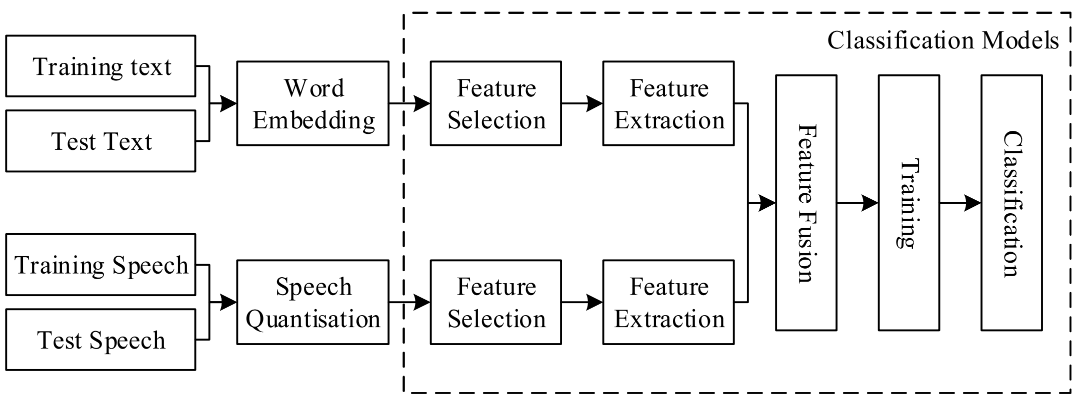

Figure 1: Text emotional classification general process

II. C. 1) General structure of the GTA-BERT model

In this paper, we propose a bimodal sentiment analysis model GTA-BERT that retains the dependency analysis information of text sentences and fuses text-speech masked attention.Specifically, the GTA-BERT model is divided into four parts: text feature extraction, speech feature extraction, DEGCN text sentiment enhancement module, and masked attention fusion.

II. C. 2) Feature extraction

(1) RGL-net text feature extractor

In this paper, we use the RoBERTa model for word embedding and combine it with a customized feature extractor for extracting text sentiment features.The RoBERTa model represents a sentence as a matrix, and its input matrix

$E=\{E_1,E_2,\cdots,E_n\}$ corresponds to the words converted into a vector by words or characters converted into vectors, which can also be understood as word vectors. Unlike other language models, RoBERTa's input matrix E contains a special vector $[CLS]$, which can be directly applied to a special vector in the recognition problem, which recognizes complete utterances and serves for utterance recognition. In the downstream task, the input matrix of RoBERTa is $E=\{[CLS],E_1,E_2,\cdots,E_n\}$, and the output matrix is $T=\{\mathrm{CLS},T_1,T_2,\cdots,T_n\}$ .The RoBERTa model employs a pre-training approach on a language, which leads to learning a universal expression of a language, which can be used for several natural language processing tasks, and is widely used in tasks such as sentiment analysis, text categorization, and so on, due to its excellent performance.

The text data is processed using RoBERTa Segmenter [17], adding [CLS] markers to indicate the start bit boundaries of the sentence and restricting the maximum length of the sentence to 75 words, if said length is not enough then the complement 0 operation is performed, for the segmentation coding of the sentence, all are set to 1 and attention masking coding is added where the complement 0 is 0 in the position of the complement 0 and 1 in the other positions.Segmentation process is performed to get the input quantization values and attention masking vectors. After RoBERTa feature extraction, the representation feature vector of the text words can be obtained, and the sequence of this feature vector can be expressed as:

$$X_t=[CLS,T_1,T_2,\cdots,T_n] \tag{1}$$

The output series after the feature extractor is:

$$X_t'=[CLS,T_1',T_2',\cdots,T_n'] \tag{2}$$

(2) Speech feature extraction module

For audio data, the Hubert model has completed the pre-training at 16kHz sampling frequency [18], after Hubert feature extraction, the representation feature vector of speech can be obtained, which can be expressed as a sequence of feature vectors:

$$X_a=[CLS,A_1,A_2,\cdots,A_n] \tag{3}$$

Aiming at the problem of inconsistency in the dimensions of the feature vectors extracted by Hubert and text feature extraction methods, a method based on one-dimensional time-domain convolution is proposed to compress the dimensions of the two feature vectors to the same dimension. Specifically, the following formula can be used for processing:

$$X_a'=\mathrm{conv1d}\left(X_a,k_{\{t,a\}}\right) \tag{4}$$

where $k_{\{t,a\}}$ represents the size of the features extracted by the text modality, which needs to be convolved with a one-dimensional convolution kernel matching the dimension of the features extracted by the pre-training model of speech. This is because the X_a' value of X_a generated by the pre-training model of speech after one-dimensional time-domain convolution is much larger than the size of the value of X_t' extracted by the feature extractor of the text. In order to avoid the product of the dot product results being too large, measures need to be taken to prevent a situation where the gradient is extremely small before entering the Softmax function. So the feature X_t' is scaled to $\hat{X}_{t'}$, and the feature X_a' is scaled to $\hat{X}'_a$, and the scaling process is expressed as:

$$\hat{X}_t'=\frac{X_t'}{\sqrt{\|X_t'\|_2}} \tag{5}$$

$$\hat{X}_a'=\frac{X_a'}{\sqrt{\|X_a'\|_2}} \tag{6}$$

(3) DEGCN Text Sentiment Enhancement Module

Syntactic dependency analysis is an important method of syntactic analysis [19], which can effectively reflect the dependencies and collocations between words in a sentence, which are closely related to the meaning of the whole sentence.

In this paper, we design a text sentiment enhancement module DEGCN, which constructs a graph for all the words of each sentence, uses graph convolutional networks to learn unstructured information features such as the dependency analysis information of the sentence to strengthen the sentiment strength of the text, and preserves the global information of the text to enhance the ability of the sentiment analysis model to understand the semantics.

The syntactic dependency graph $G=(V,E)$ is an undirected graph constructed by the syntactic dependency structure of sentences, where V is the set of nodes in the graph, i.e., the set of words in a sentence. E is the set of edges and edge relations, i.e., the dependencies between words in the sentence. In order to utilize graph convolutional network for feature extraction of the undirected graph G, this paper adopts a two-layer graph convolutional network convolution operation, which firstly makes $\tilde{A}=\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$, and $\tilde{D}$ is the degree matrix of $\tilde{A}$, and $\tilde{D}=\sum_i \tilde{A}_{ij}$. Finally, the convolved feature matrix G_m is obtained, and the output G_m of the graph convolutional network is defined as:

$$G_m = \tilde{A}\,\mathrm{Re}LU\left(\tilde{A}XW^{(0)}\right)W^{(1)} \tag{7}$$

In this case, the ReLU function is used for the activation function, the weight matrix $W^{(0)}$ denotes the connection weight between the hidden layer and the input layer, and the weight matrix $W^{(1)}$ denotes the connection weight between the hidden layer and the output layer.

II. C. 3) Masking Attention Fusion

The masked attention fusion module is the core component of the GTA-BERT model, whose main purpose is to fuse the pre-trained features extracted by the textual feature extractor and the pre-trained model Hubert, while dynamically adjusting the weights of the different modal features through the attention mechanism.

The module first evaluates the importance of words in text features and pre-training features separately, using the text feature query vector Q_t and the keyword vector K_t and setting them equal $\hat{X}'_t$, $\hat{X}'_t$ being the scaled text features. Then, using the speech feature query vector Q_a and the keyword vector K_a and setting them equal $\hat{X}'_a$, $\hat{X}'_a$ are the representation vectors generated by the pre-trained model of speech after the scaling process. Therefore, the attention matrices A_t and B_a for text features and speech features are computed as shown in (8) and (9), respectively. Therefore, the masked attention fusion model can effectively utilize the features of multiple channels, thus improving the accuracy of emotion recognition:

$$A_t = \mathrm{Re}LU\left(Q_t K_t^T\right) \tag{8}$$

$$B_a = \mathrm{Re}LU\left(Q_a K_a^T\right) \tag{9}$$

The attention matrices A_t and B_a are computed for text and audio modalities, respectively, and then they are summed to obtain the attention matrix W_m.

After obtaining the multimodal attention matrix, W_m is multiplied by the feature matrix G_m of the graph convolutional network to obtain the output of the masked fusion attention matrix X_{GTA}:

$$X_{GTA} = W_m G_m \tag{10}$$

Finally, the masked fusion attention matrix X_{GT} and the output matrix of graph convolution G_m are summed through the residual network to obtain the joint feature matrix, and finally the sentiment classification is performed through the fully connected layer and Softmax.

III. Analysis of emotional perceptual properties and regulatory effects

III. A. Emotion perception analysis

III. A. 1) Single-mode experimental results

To test the intelligent sentiment analysis model in AI-assisted education in this paper, this paper first conducts unimodal sentiment analysis experiments for each unimodal modality, which is used to validate the usability of each unimodal annotation, in order to compare the difference between independent single labels for each modality and multimodal labels, the model performance for the two types of unimodal sentiment analysis, the unimodal label-supervised and multimodal label-supervised, is shown in Table 1, specifically, for the Model performance with unimodal annotations is shown to the left of "/". The left side of "/" is supervised evaluation using unimodal annotations, and the right side of "/" is supervised evaluation using multimodal annotations.

From the analysis of the results in Table 1, the first inter-modal comparisons are made, and it can be found that the textual modality is more likely to accurately predict the emotion, while the non-textual modality is difficult to predict accurately, and the acoustic modality is more obvious, and the binary Acc2 prediction accuracy is less than 75%, which suggests that it is difficult to tap the emotional cues in the audio information, and it is found that the acoustic features of the same samples in the experiments are obviously longer than the textual features and the visual features. In the sentiment analysis of text, acoustic A and acoustic B, the GTA-BERT model in this paper gives the best results. A side-by-side comparison within modalities reveals that supervision and testing of independent labels for each modality gives better performance, suggesting that single-label annotations fit the unimodal information better, however, using uniform multimodal labels to supervise each unimodal modality suffers from affective differences, and the model has difficulty in capturing affective cues from the modalities. In addition to this, the supplemental acoustic B experiment achieved better performance, for which it was able to demonstrate that the use of pre-trained acoustic modalities better aids the model in representation learning.

To summarize, the poor performance of the current mainstream models in the acoustic unimodal task reveals the challenges faced by acoustic feature extraction of emotional features in future research. In addition, the difference in performance under unimodal and multimodal annotations validates that multimodal annotations may mislead the unimodal representation learning process.

Table 1: Single modal experiment results

Modal	Model	Supervisory label	Acc2	Acc2_weak	Corr	MAE
Text	Bi-GRU	T training T predict	80.58	68.83	48.95	0.419
	Wav2vec2	M training M predict	74.73	68.43	52.09	0.261
	VggFace	A training A predict	78.55	66.63	50.16	0.306
	GTA-BERT	M training M predict	92.82	79.99	74.01	0.237
Acoustics (A)	Bi-GRU	A training A predict	78.84	58.49	49.87	0.339
	Wav2vec2	M training M predict	87.13	63.57	59.29	0.403
	VggFace	A training A predict	78.53	57.94	55.14	0.407
	GTA-BERT	M training M predict	91.25	73.98	72.68	0.302
Acoustics (B)	Bi-GRU	V training V predict	74.13	64.65	53.65	0.366
	Wav2vec2	M training M predict	83.87	63.46	63.31	0.369
	VggFace	V training V predict	85.86	55.28	65.49	0.353
	GTA-BERT	M training M predict	92.77	74.07	70.67	0.309

III. A. 2) Multimodal experimental results

For the above single-label experiments proved that single-label annotation is necessary, CH-SIMS v2.0 dataset is a multi-label dataset, however, the mainstream baseline model only performs multimodal single-label supervision, in order to further validate that single-modal labeling of text and audio can help the performance of multimodal sentiment analysis models, this paper conducts multi-task multi-label multimodal sentiment analysis experiments.

The traditional MSA models using single multimodal supervised supervision include EF-LSTM, MFN, MISA, MAG-BERT, Self-MM, LF-DNN, TFN, and LMF. The methods that introduce multilabel supervision (multimodal, textual, and acoustic three-modal labeling) to guide the learning of MSA representations include MTFN, MLF-DNN, and MLMF models. The experimental results of each baseline model on this dataset are shown in Table 2.

From Table 2, it can be found that firstly, each model performs poorly in Acc2-weak metrics, which indicates that the models perform poorly in predicting weak emotion samples close to the real situation, in addition to this, the regression metrics such as Corr, R-square, and MAE perform poorly as a whole, which suggests that although the existing models can predict emotion polarity more accurately, they can't accurately determine the emotion intensity degree, the existing model is still insufficient in the prediction ability of fine-grained emotion.

On the other hand, the Acc-2, F1, Acc2-weak, Corr, and R-square values of the GTA-BERT sentiment analysis model constructed in this paper are 93.24, 83.96, 75.01, 72.67, and 46.63, respectively, which are the largest values among all the compared models, and the MAE value of the GTA-BERT model is 0.303, the the smallest value among all models. It can be seen that the GTA-BERT sentiment analysis model in this paper has the best sentiment analysis effect.

Table 2: Multimodal experiment results

Model	Acc-2	F1	Acc2-weak	Corr	R-square	MAE
EF-LSTM	83.47	80.03	68.13	64.33	33.13	0.349
MFN	81.44	76.07	69.75	65.08	37.19	0.347
MAG-BERT	74.23	75.14	67.73	53.14	35.44	0.339
Self-MM	78.32	75.63	66.72	63.38	35.24	0.322
LF-DNN	80.88	79.26	71.06	68.48	38.11	0.338
TFN	83.93	79.11	68.97	64.17	36.73	0.363
LMF	80.69	81.03	69.36	58.84	35.88	0.338
MLF-DNN	82.19	80.41	70.37	69.88	38.57	0.368
MTFN	78.45	77.93	71.87	56.37	33.21	0.326
MLMF	84.06	76.58	72.14	61.35	33.22	0.372
GTA-BERT	93.24	83.96	75.01	72.67	46.63	0.303

III. B. Analysis of regulatory effects

III. B. 1) Research hypotheses

In this paper, the following hypotheses are made to investigate the moderating effect of emotional perception on learners' behavioral patterns in AI-assisted education:

H1: Artificial intelligence-assisted teaching can significantly influence learners' behavioral patterns.

H2: Artificial intelligence-assisted teaching can significantly affect emotional arousal.

H3: AI-assisted teaching can significantly affect emotion perception.

H4: Emotion perception can significantly influence emotion arousal.

H5: Emotion perception can significantly influence learners' behavioral patterns.

H6: Emotion can significantly influence learners' behavioral patterns.

H7: Emotion perception plays a significant mediating role in the influence of artificial intelligence-assisted teaching on learners' behavioral patterns.

H8: Emotion plays a significant mediating role in the process of the influence of artificial intelligence-assisted teaching on learners' behavior patterns.

H9: Emotion perception and emotion play a significant dual mediating role in the process of the influence of artificial intelligence-assisted teaching on learners' behavioral patterns.

III. B. 2) Reliability analysis

(1) Reliability test

The results of the reliability test showed that the Clonbach's alpha coefficients for the scale as a whole and for AI-assisted teaching, emotion perception, emotion, and learner's behavioral patterns were 0.826, 0.834, 0.796, 0.893, and 0.862, respectively, which were greater than 0.7, indicating that the internal consistency of the survey data was good and the reliability was high.

(2) Structural validity

First, the initial validity factor analysis and the correlation coefficient test between different dimensions were done for the four variables of AI-assisted teaching, emotion perception, emotion and learners' behavioral patterns. The results of the model fit indices are shown in Table 3. The results showed that $\chi^2/df=1.085$, CFI=0.976, TLI=0.946, SRMR=0.037, and RMSEA=0.021, which indicated that the first-order model fit was better and the standardized correlation coefficients between the dimensions were greater than 0.7, so it could be proceeded to the second-order validity factor analysis. Second, the results of the second-order validation factor analysis showed that $\chi^2/df=1.077$, CFI=0.983, TLI=0.965, SRMR=0.039, and RMSEA=0.017, with each fitting index slightly improved over the first-order model, and the standardized correlation coefficients between the dimensions were all greater than 0.7, which indicated better structural validity between the variables.

Table 3: Model fit index results

Measuring model	First-order model	Second-order model
$\chi 2$	320.156	322.048
df	295	299
$\chi 2$/df	1.085	1.077
RMSEA	0.021	0.017
SRMR	0.037	0.039
CFI	0.976	0.983
TLI	0.946	0.965

(3) Distinctive validity

The results of the discriminant validity test are shown in Table 4, where AI, EP, E, and BM denote artificial intelligence-assisted teaching, emotion perception, emotion, and learner behavior patterns, respectively. As can be seen from the results, the square root of all AVE values is greater than the correlation coefficients between all latent variables, indicating that the internal correlation between all dimensions is greater than the external correlation, which suggests that the discriminant validity between the latent variables passed the test.

Table 4: Results of discriminant validity test

Dimension	AI	EP	E	BM
AI	0.768			
EP	0.635***	0.834		
E	0.612***	0.493***	0.841	
BM	-0.385***	-0.423***	-0.382***	0.806

III. B. 3) Direct effect tests

The author collected sample data from 300 users of AI-assisted teaching by means of questionnaires, and analyzed and processed the formal survey data by using SPSS 26.0 and Amos 24.0 software. In order to improve the rigor of the study, the reliability and validity of the formal survey data were tested by Cronbach's alpha coefficient and Bartlett's sphere. After the test, the data in this paper has good reliability and validity.

The results of the direct effect test are shown in Table 5. The results in Table 5 show that learners' behavioral patterns, emotions and emotion perception are directly affected by AI-assisted teaching, with standardized effect values of 0.526, 0.543 and 0.387, respectively, which are all significant at the 1% level, and the hypotheses H1, H2, and H3 are verified. Emotion perception has a significant effect on emotion and learners' behavioral patterns with standardized effect values of 0.168 and 0.175 respectively, both significant at 1% level and hypotheses H4 and H5 were verified. Emotion has a significant effect on learners' behavioral patterns, with a standardized effect value of 0.173, which is significant at the 5% level, and hypothesis H6 is verified. In summary, artificial intelligence-assisted teaching, emotion perception and emotion all significantly affect learner behavior patterns. Artificial intelligence-assisted teaching significantly affects emotion perception and significantly affects emotion and learner behavior patterns. Emotion perception significantly affects emotion and learner behavior patterns, and emotion significantly affects learner behavior patterns.

Table 5: Results of direct effect test

Hypothesis	Path	Standardized effect value	SE	CR	P	Result
H1	AI→BM	0.526	0.163	5.784	0.000	Pass
H2	AI→E	0.543	0.127	6.859	0.000	Pass
H3	AI→EP	0.387	0.129	5.748	0.000	Pass
H4	EP→E	0.168	0.063	2.823	0.000	Pass
H5	EP→BM	0.175	0.074	3.556	0.000	Pass
H6	E→BM	0.173	0.089	2.286	0.012	Pass

III. B. 4) Mediated effects test

The results of the mediation effect test are shown in Table 6. In terms of direct effect, the standardized effect value of AI-assisted teaching → learners' behavioral pattern path is 0.526, P=0.000, and the proportion of direct effect to total effect is 76.34%. In terms of mediating effect, the standardized effect value of AI-assisted teaching → emotion perception → learner behavior pattern path is 0.075, P=0.003, 95% CI: 0.023 to 0.153, and the mediating effect accounts for 10.89% of the total effect, and the hypothesis H7 is verified. The standardized effect value of the path of artificial intelligence-assisted teaching → emotion → learner behavior pattern was 0.075, P=0.019, 95% CI: 0.016 to 0.184, the mediating effect accounted for 10.89%, and hypothesis H8 was verified. The standardized effect value of the path of artificial intelligence-assisted teaching→emotion perception→emotion→learner's behavioral pattern was 0.013, P=0.032, 95% CI: 0.000 to 0.032, the mediating effect accounted for 1.88%, and hypothesis H9 was verified.

In summary, the standardized effect value of the path of AI-assisted teaching → emotion perception → emotion → learner behavior pattern is much smaller than the standardized effect value of the path of AI-assisted teaching → emotion perception → learner behavior pattern and AI-assisted teaching → emotion → learner behavior pattern. And the standardized effect value of AI-assisted teaching→emotion perception→learner behavior and AI-assisted teaching→emotion→learner behavior pattern paths did not differ much, thus proving that both emotion perception and emotion play an important moderating role in the process of AI-assisted teaching's influence on learners' behavior patterns.

Table 6: Mediation effect test results

Type	Path	Standardized effect value	SE	95%CI		P	Effect proportion
				Lower	Upper		
Direct effect	AI→BM	0.526	0.115	0.245	0.672	0.000	76.34%
Mediation effect	AI→EP→E→BM	0.013	0.006	0.000	0.032	0.035	1.88%
	AI→EP→BM	0.075	0.034	0.023	0.153	0.003	10.89%
	AI→E→BM	0.075	0.037	0.016	0.184	0.019	10.89%
Total mediation effect	-	0.163	0.056	0.054	0.267	0.006	23.66%
Total effect	-	0.689	0.081	0.472	0.783	0.000	100%

IV. Conclusion

In this study, we constructed a bimodal-based intelligent sentiment analysis model, GTA-BERT, and explored the moderating effect of emotion perception on learners' behavioral patterns. The study shows that the proposed GTA-BERT model performs well in the sentiment analysis task with an MAE value of 0.303, which is the smallest among all the compared models, indicating that the model is able to predict the intensity of emotions more accurately. The empirical study verified the importance of emotion perception in AI-assisted education, and the results showed that AI-assisted teaching had a significant direct effect on learners' behavioral patterns, with a standardized effect value of 0.526 (P<0.001), and the direct effect accounted for 76.34% of the total effect. Meanwhile, emotion perception and emotion play an important mediating role in this process, and the standardized effect values of the two paths of AI-assisted teaching → emotion perception → learner behavioral pattern and AI-assisted teaching → emotion → learner behavioral pattern are both 0.075, each accounting for 10.89% of the total effect. The study also found that the standardized effect value of the three-variable dual mediation effect AI-assisted teaching → emotion perception → emotion → learner behavior mode path was 0.013, accounting for a relatively small percentage (1.88%), indicating that emotion perception and emotion prefer to play a mediating role alone.

The above findings indicate that integrating emotion perception technology into AI-assisted education system has an important value, which can effectively regulate the learning behavior and improve the emotional interaction ability and teaching effect of intelligent education system by timely recognizing and responding to the emotional state of learners.

References

[1] Strunin, D. A. (2023). Artificial intelligence in the field of education. Young scientist, (6), 453.
[2] Hu, S. (2024). The effect of artificial intelligence-assisted personalized learning on student learning outcomes: A meta-analysis based on 31 empirical research papers. Science Insights Education Frontiers, 24(1), 3873-3894.
[3] Zhang, W., Xiong, Y., Zhou, D., Liu, C., Gu, Y., & Yang, H. (2025). Balancing human and AI instruction: insights from secondary student satisfaction with AI-assisted learning. Interactive Learning Environments, 1-16.
[4] Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: A comparative study. Journal of Educational Computing Research, 62(8), 1961-1994.
[5] Haderer, B., & Ciolacu, M. (2022). Education 4.0: Artificial intelligence assisted task-and time planning system. Procedia computer science, 200, 1328-1337.

[6] Yao, W., & Li, N. (2022). Construction of artificial intelligence-assisted English learning resource query system. Frontiers in Psychology, 13, 970497.

[7] Kasztelnik, K. (2024). Artificial Intelligence-Assisted Curriculum Development: Innovations in Designing Educational Content for the 21st Century Learner. Journal of Higher Education Theory and Practice, 24(11), 51-59.

[8] Abdullah, T., & Ahmet, A. (2022). Deep learning in sentiment analysis: Recent architectures. ACM Computing Surveys, 55(8), 1-37.

[9] Zangeneh Soroush, M., Maghooli, K., Setarehdan, S. K., & Motie Nasrabadi, A. (2018). Emotion classification through nonlinear EEG analysis using machine learning methods. Int. Clin. Neurosci. J, 5(4), 135-149.

[10] Firdaus, M., Thangavelu, N., Ekbal, A., & Bhattacharyya, P. (2022). I enjoy writing and playing, do you?: a personalized and emotion grounded dialogue agent using generative adversarial network. IEEE Transactions on Affective Computing, 14(3), 2127-2138.

[11] Wang, Y., & Liu, W. (2023). Emotional Simulation of Artificial Intelligence and Its Ethical Reflection. Academic Journal of Humanities & Social Sciences, 6(5), 11-15.

[12] Chin, H., Song, H., Baek, G., Shin, M., Jung, C., Cha, M., ... & Cha, C. (2023). The potential of chatbots for emotional support and promoting mental well-being in different cultures: mixed methods study. Journal of Medical Internet Research, 25, e51712.

[13] Järvenoja, H., Järvelä, S., Törmänen, T., Näykki, P., Malmberg, J., Kurki, K., ... & Isohätälä, J. (2018). Capturing motivation and emotion regulation during a learning process. Frontline Learning Research, 6(3), 85-104.

[14] Yagishita, S. (2020). Transient and sustained effects of dopamine and serotonin signaling in motivation‐related behavior. Psychiatry and clinical neurosciences, 74(2), 91-98.

[15] Ba, S., Stein, D., Liu, Q., Long, T., Xie, K., & Wu, L. (2021). Examining the effects of a pedagogical agent with dual-channel emotional cues on learner emotions, cognitive load, and knowledge transfer performance. Journal of Educational Computing Research, 59(6), 1114-1134.

[16] Yaser Hasan Al Mamary & Aliyu Alhaji Abubakar. (2025). Empowering ChatGPT adoption in higher education: A comprehensive analysis of university students' intention to adopt artificial intelligence using self-determination and technology-to-performance chain theories. The Internet and Higher Education,66,101015-101015.

[17] Sander Puts,Catharina M L Zegers,Andre Dekker & Iñigo Bermejo. (2025). Developing an ICD-10 Coding Assistant: Pilot Study Using RoBERTa and GPT-4 for Term Extraction and Description-Based Code Selection. JMIR formative research,9,e60095.

[18] Moung Ho Yi,Keun Chang Kwak & Ju Hyun Shin. (2024). KoHMT: A Multimodal Emotion Recognition Model Integrating KoELECTRA, HuBERT with Multimodal Transformer. Electronics,13(23),4674-4674.

[19] Zhang Fan,Zheng Wenbin & Yang Yujie. (2024). Graph Convolutional Network with Syntactic Dependency for Aspect-Based Sentiment Analysis. International Journal of Computational Intelligence Systems,17(1).