Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.34671

Research and Analysis on the Impact of Adaptive Learning Systems Based on Large Language Models and VAD Emotional Models on Cognitive and Non-Cognitive Performance

Chao Huang[1*], Zibei Shi[2], Sijin Wang[3]

[1]*Chuxiong Normal University, Chuxiong, China*
[2]*Renmin University Affiliated Middle School, Beijing, China*
[3]*University of Macau, Macau, China*
**Corresponding Author. Email: huangchao954161954@gmail.com*

Abstract. Traditional emotion recognition methods typically rely on unimodal signals or coarse-grained emotion categories, which lack the sensitivity and interpretability required to capture subtle emotional fluctuations during the learning process. Build an adaptive learning system that uses dual-driven emotion recognition based on large language models and VAD emotion dimension models, investigate the effect of this system on learners' cognitive and non-cognitive performance, and verify the accuracy and interpretability of the emotion recognition module. Constructing the VAD system architecture, setting up experimental and control groups, and applying statistical analysis to the data, such as the SAM scale, it has been found that the research data are normally distributed, and the results show that the VAD model performs better than the ordinary model in learning disability identification and learning persistence. However, this study has some limitations; namely, it uses a relatively small sample size and is only based on text emotion recognition; in the future, multi-modal emotion perception will be introduced to enhance recognition accuracy and intervention effectiveness.

Keywords: Adaptive Learning System, Emotion Recognition, Large Language Models, Artificial Intelligence

1. Introduction

In recent years, instructional support systems have gradually moved away from the traditional "one-size-fits-all" approach to personalised adaptive learning. Adaptation to learners' individual interests and abilities is used to adjust the content in these adaptive systems, as shown in paper [1] and paper [2], which can help improve learning outcomes. Recently, large language models (LLMs) have shown strong text-processing and domain-specific knowledge abilities, thus bringing substantial new possibilities to adaptive learning [3], and AI-driven systems have further promoted positive effects on educational results [4].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.34671

Studies show that emotional factors are related to how well students learn [5]. At present, many new teaching systems aim to promote the growth of students' emotions through artificial intelligence. Some studies have found that by observing changes in students' emotions and providing timely feedback, both their academic results and mental health have been promoted [6-8]. However, the AI-based personalised feedback system is still a "black box" and stochastic system [9], and thus faces trust issues in real-world applications. Build an objective and reliable system for measuring emotions in a scientific way to help AI recognize the emotional state of learners during their study.

2. Principles and methodology

This paper proposes an instructional support system with a dual-drive architecture that integrates LLMs and emotion lexicons based on the VAD (Valence-Arousal-Dominance) model.

The VAD continuous dimension model [10] is used as the discriminant standard by the system to parameterize affective states as vectors in a three-dimensional space, and the scores of each dimension are mapped to the [0, 1] interval.

$$\boldsymbol{E} = [\boldsymbol{V}, \boldsymbol{A}, \boldsymbol{D}] \hat{} \boldsymbol{T} \tag{1}$$

To solve the problem of interpretability for generative models, the system integrates traditional rule-based and lexicon-based emotion recognition methods (ANEW, SentiWordNet and VADER) [11] and applies them in Python. Dictionary-based methods are relatively deterministic, easy to explain, and have good academic support compared with LLMs. Establish an explicit audit trail for each affective score (V, A, D) to prevent the "black-box" problem and set up deterministic anchors for LLM-based emotional reasoning.

Lexicon-based methods are relatively deterministic and explainable compared with large language models (LLMs), and have more academic support. A clear audit trail can thus be established to prevent the problem of a "black box" in deep learning and provide some deterministic, symbolic supports for the emotional reasoning of LLMs [12]. However, the above methods are inherently limited in their capacity to deal with complex semantic contexts and long-term dependencies, and they cannot remember what a learner has learned previously for comparison.

Therefore, the paper will also use the GPT-5.4 model as a secondary core engine for emotion recognition simultaneously. Leveraging the strong zero-shot reasoning ability and large parameter scale of GPT-5.4, it is good at detecting subtle affective cues in fragmented or euphemistic expressions of learners. Given that emotional development is not constant, GPT-5.4's large-context window will be used to maintain the change in emotion over an extended period of teaching and learning. Therefore, historical feedback can be incorporated to improve the accuracy of current VAD scoring and reduce the impact of affective misjudgment due to the limitations of isolated sentence analysis.

The whole architecture of the educational support system is shown in Figure 1 below. The system will employ a "plug-and-play" mode and, through Dify, build an unstructured pedagogical interaction environment with structured affective audit algorithms. At the bottom of the above integration is a real-time feedback loop that, after getting student text input, activates both the LLM reasoning stream and the Python-scripted lexicon auditing stream concurrently. Therefore, the generation of pedagogical content must be in harmony with the learner's emotional state at all times.

Dify is an open-source LLM application development platform that will be used as the workflow orchestration engine in this study. Build a large number of data pipelines in an easy-to-operate way.

DOI: 10.54254/2755-2721/2026.34671

By using Dify's code nodes, this study has integrated lexicon-based auditing logic to perform real-time calibration of LLM outputs and simultaneously generated an interactive web interface.

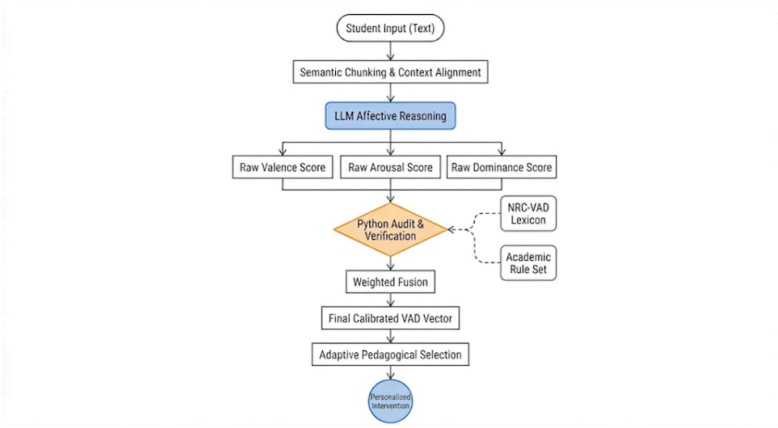


Figure 1. The overall architecture of the education support system (picture credit: original)

First, the student-generated text is sent to the Semantic Perception Layer for the first classification of intent, feature recognition and preprocessing. The data is then passed to the Dual-Stream Affective Recognition Layer, and the two emotional recognition engines operate independently at the same time. These engines map the affective dimensions to the [0, 1] range and then obtain the final VAD vector by a weighted-average method.

Then the resulting vector is fed into the Adaptive Intervention Layer. Using conversational history to create personalised learner profiles and mapping emotional fluctuations, this level will then determine how to change learning strategies for both cognition and affect. After introducing the scaffolding teaching strategy [13], the system will generate and present an empathetic and guided response powered by GPT-5.4.

3. Experiments and analysis

3.1. Experimental design

Independent variable is the use of Emotion-aware Adaptive Learning Assistant (Experimental Group vs. Control Group) Control Group, dependent variables include academic emotions (positive/negative), test anxiety, learning efficacy (quiz scores, assignment quality) and accuracy of emotion recognition (as a process variable)

3.2. Evaluation metrics

Affective metrics include Self-Assessment Manikin (SAM: A visual scale to measure each participant's Valence, Arousal, and Dominance) and Westside Test Anxiety Scale (Used to assess participants' anxiety levels regarding examinations).

Learning efficacy metrics include Course unit test scores (Pre-test and Post-test), Completion rates and accuracy of learning tasks, Platform log data, including learning duration and interaction frequency.

DOI: 10.54254/2755-2721/2026.34671

3.3. Experimental procedure

Recruitment of participants and signing of informed consent forms to ensure ethical compliance (privacy protection, data de-identification).

All participants completed the SAM scale, Westside Test Anxiety Scale, and a baseline knowledge test.

(1) Experimental Group. Used the emotion-aware platform for assigned weekly tasks (30–60 mins per session). The system recorded real-time affective data and adaptive feedback logs.

(2) Control Group. Used a standard learning platform for identical tasks without emotion recognition or feedback.

(3) Weekly online. SAM questionnaires were collected from both groups for process monitoring.

(4) Post-test (Day 7). All participants re-completed the SAM scale, Westside Test Anxiety Scale, and an equivalent post-test knowledge assessment.

4. Experimental results

4.1. Summary of group statistics

Table 1. Descriptive statistics table of valence, arousal, dominance, and accuracy by experimental and control groups

Metric	Group	Mean	Std.Dev(SD)	Sample Size(N)
Valence	Experimental	0.53	0.24	14
Arousal	Experimental	0.57	0.26	14
Dominance	Experimental	0.69	0.22	14
Accuracy(%)	Experimental	51.43%	14.46%	14
Valence	Control	0.65	0.21	14
Arousal	Control	0.73	0.21	14
Dominance	Control	0.76	0.23	14
Accuracy(%)	Control	61.21%	15.01%	14

As shown in Table 1-3, it compares the means, standard deviations, and sample sizes of the experimental and control groups across four indicators which are valence, arousal, dominance, and accuracy, providing a clear visual representation of the distribution differences between the two groups on these variables.

Table 2. Experimental group psychological state (efficacy/arousal/control) and task accuracy data

experimental group				
code	Potency	arousal	sense of control	accuracy
1	0.68	0.49	0.72	58
2	0.52	0.38	0.61	42
3	0.79	0.65	0.85	65
4	0.63	0.52	0.68	54
5	0.48	0.29	0.55	39
6	0.82	0.78	0.81	68

DOI: 10.54254/2755-2721/2026.34671

Table 2. (continued)

7	0.66	0.55	0.7	55
8	0.59	0.45	0.65	49
9	0.75	0.69	0.79	62
10	0.57	0.44	0.64	48
11	0.3	0.7	0.7	50
12	0.2	0.8	0.7	20
13	0.3	0.7	0.8	50
14	0.2	0.6	0.5	50

Table 3. control group psychological state (efficacy/arousal/control) and task accuracy data

control group				
code	Potency	arousal	sense of control	accuracy
1	0.63	0.72	0.75	62
2	0.48	0.58	0.62	48
3	0.71	0.85	0.88	75
4	0.59	0.69	0.72	59
5	0.45	0.55	0.6	45
6	0.74	0.82	0.85	72
7	0.6	0.7	0.73	60
8	0.54	0.63	0.67	53
9	0.69	0.79	0.81	68
10	0.57	0.67	0.69	56
11	0.6	0.9	0.9	67
12	0.8	0.8	0.8	53
13	0.9	0.8	0.8	67
14	0.8	0.7	0.8	70

4.2. Normality test summary

Table 4. Normality test and descriptive statistics results table for multiple indicators under experimental and control groups

indices	grouping	mean	standard deviation	skewness	skewness description	kurtosis	Kurtosis description (based on the normal kurtosis value of 3)	W statistic	P value	Normality conclusion
Potency	experimental	0.562	0.217	-0.384	Mild skewness (skewness < 0.5)	4.002	Leptokurtic (> 3)	0.939	0.411	normally distributed

DOI: 10.54254/2755-2721/2026.34671

Table 4. (continued)

arousal	experimental	0.447	0.229	0.599	moderate skewness (0.5 ≤ skewness < 1)	3.139	Moderate kurtosis (close to 3)	0.957	0.668	normally distributed
sense of control	experimental	0.642	0.221	0.25	Slight skewness (skewness < 0.5)	3.103	Moderate kurtosis (close to 3)	0.982	0.983	normally distributed
accuracy	experimental	49.775	12.807	-0.345	Mild skewness (skewness < 0.5)	2.394	Platykurtic (< 3)	0.947	0.517	normally distributed
Potency	control	0.65	0.173	0.155	Mild skewness (skewness < 0.5)	1.818	Platykurtic (< 3)	0.953	0.614	normally distributed
arousal	control	0.739	0.255	-0.864	Moderate skewness (0.5 ≤ skewness < 1)	3.814	Leptokurtic (> 3)	0.904	0.131	normally distributed
sense of control	control	0.738	0.16	-0.18	Mild skewness (skewness < 0.5)	2.586	Platykurtic (< 3)	0.965	0.809	normally distributed
accuracy	control	58.348	13.177	0.276	Mild skewness (skewness < 0.5)	5.361	High leptokurtosis (much greater than 3)	0.916	0.194	normally distributed

Table 4 presents the normality test and descriptive statistical results of multiple indicators for the experimental and control groups.

4.3. Analysis of normality

Shapiro-Wilk normality tests were performed on all 8 datasets (4 metrics across 2 groups). Results indicate that all p-values exceed 0.05 (ranging from 0.131 to 0.983), achieving a 100% pass rate for the Shapiro-Wilk test. Combined with Q-Q plot analysis, all groups follow a normal distribution, satisfying the prerequisites for subsequent parametric tests (e.g., Independent Samples t-test).

For Skewness, 75% of the groups (6/8) exhibited slight skewness (| Skewness | <0.5), and 25% exhibited moderate skewness (0.5≤ | Skewness | <1). No severe skewness was observed.

For Kurtosis, Kurtosis values ranged from 1.818 to 5.361. Three groups were platykurtic (<3), three were moderate (≈3), and two were leptokurtic (>3).

Despite variations in kurtosis, the successful Shapiro-Wilk tests confirm that the data conforms to the normal distribution hypothesis.

For Shapiro-Wilk Test, p>0.05 indicates a normal distribution. | Skewness | <0.5 (Slight), 0.5≤ | Skewness | <1 (Moderate), | Skewness | ≥1 (Severe). Reference value for Kurtosis of a normal distribution is 3 (<3 for platykurtic, >3 for leptokurtic).

DOI: 10.54254/2755-2721/2026.34671

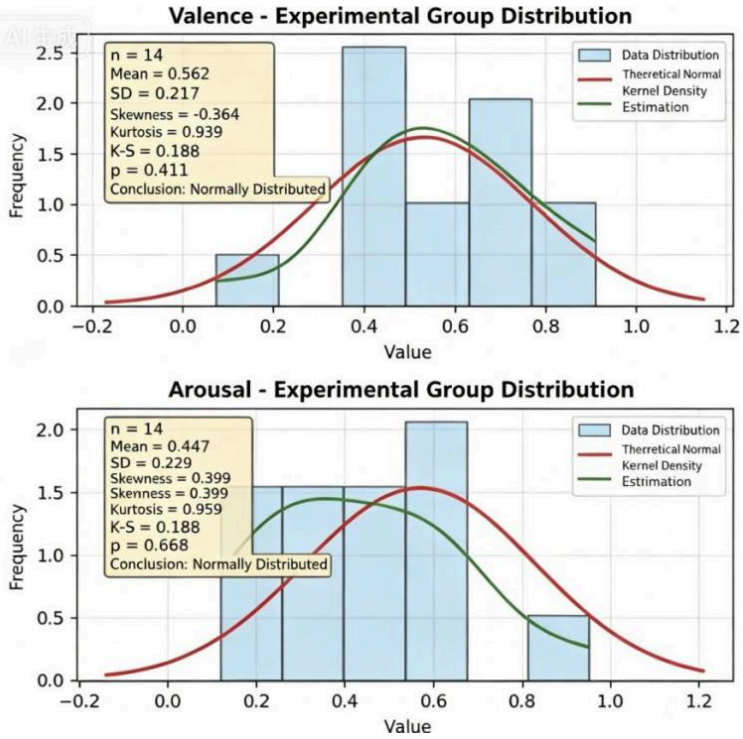


Figure 2. Histograms, kernel density plots and normal curve-based distributions of valence/arousal for the experimental group and normality tests (picture credit: original)

Based on the above statistical tests, it is found that all of the variables in both the experimental and control groups meet the requirement of normality, as shown in the figure 2. Therefore, this paper reports that the affective-aware model, in conjunction with the VAD framework, is superior to the traditional teaching model in detecting problems in students' learning and boosting their learning motivation.

5. Conclusion

This study has some deficiencies and potential stochasticity due to a small sample size and its sole use of single-modality text recognition. It should be pointed out that the addition of features to the VAD model is not a new feature but rather a continuous, interpretable, and psychologically consistent three-dimensional space for affective computing. Future research will explore developing a multimodal model for people-object interaction based on images and sound, etc. The above progress will likely enhance the accuracy of emotion recognition, deepen the understanding of learners' emotional states, and improve the effectiveness of teaching interventions.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Oxman, S., & Wong, W. (2014). White paper: Adaptive learning systems. Integrated Education Solutions.
[2] Na, I., Hegelheimer, V. (2025). AdaptiVe learning systems. In: McCallum, L., Tafazoli, D. (eds) The palgrave encyclopedia of computer-assisted language learning. Palgrave Macmillan, Cham. https: //doi.org/10.1007/978-3-031-51447-0_1-1

DOI: 10.54254/2755-2721/2026.34671

[3] Wölfel M.Shirzad M. B.ReichA.AndererK. (2024). Knowledge-based and generative-AI-driven pedagogical conversational agents: a comparative study of Grice's cooperative principles and trust. Big Data Cogn. Comput. 8, 1–2010.3390/bdcc8010002

[4] Chen Y-N, Wei L-N. What is the Influence of AI-driven adaptive learning on students' personalized learning outcomes?—A meta-analysis of 29 experimental and quasi-experimental research. Modern Educational Technology, 2025, 35(12): 55-65. https: //doi.org/10.3969/j.issn.1009-8097.2025.12.006

[5] Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. Frontiers in psychology, 8, 1454. https: //doi.org/10.3389/fpsyg.2017.01454

[6] Gutierrez R, Villegas-Ch W and Govea J (2025) Development of adaptive and emotionally intelligent educational assistants based on conversational AI. Front. Comput. Sci. 7: 1628104. doi: 10.3389/fcomp.2025.1628104

[7] MaazN.MounsefJ.MaaloufN. (2025). CARE: towards customized assistive robot-based education. Front. Robot. AI 12: 1474741. 10.3389/frobt.2025.1474741

[8] A. Darejeh, T. S. Moghadam, M. Delaramifar and S. Mashayekh, "A framework for AI-powered decision making in developing adaptive e-learning systems to impact learners' emotional responses, " 2024 11th International and the 17th National Conference on E-Learning and E-Teaching (ICeLeT), Isfahan, Iran, Islamic Republic of, 2024, pp. 1-6, doi: 10.1109/ICeLeT62507.2024.10493103.

[9] Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society Jan-Jun 2016, 1–12

[10] Mehrabian A, Russell J A. An approach to environmental psychology [M]. MIT, 1974.

[11] Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225. https: //doi.org/10.1609/icwsm.v8i1.14550

[12] Bestgen, Y. (1994). Can emotional valence in stories be determined from words? Cognition and Emotion, 8(1), 21–36. https: //doi.org/10.1080/02699939408408926

[13] van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. Educational Psychologist, 38(1), 5–13. https: //doi.org/10.1207/S15326985EP3801_2