Improving Teaching Engagement and Professional Core Competencies in Higher Vocational Humanities Courses Based on AI Chatbots

Jiangbo Guo[1], Dong Wang[1], Min Cui[1], Xiaonong Yan[1], Lifang Ma[3], Xin Wang[2], Jianmin Wang[4], Zhang Zhao[5]

[1]College of Intelligent Building Engineering, Dongying Vocational College, Dongying 257000, Shandong, China
[2]Party Committee Office (Legal Affairs Office), Dongying Vocational College, Dongying 257000, Shandong, China
[3]College of Culture, Tourism, Health and Education, Dongying Vocational College, Dongying 257000, Shandong, China
[4]College of Information and Future Technology, Dongying Vocational College, Dongying 257000, Shandong, China
[5]School of Ecology and Biology, Dongying Vocational College, Dongying 257091, Shandong, China

Corresponding author: Zhang Zhao (e-mail: ZZ2026123123@163.com)

ABSTRACT AI chatbots are increasingly entering higher vocational humanities courses, providing new possibilities for improving interaction, engagement, and professional core competencies. This study combines natural language processing and instructional design theory to construct an AI chatbot-assisted teaching framework. Using a quasi-experimental design, questionnaire measurement, behavior-log analysis, and academic performance evaluation are employed to examine the influence of chatbots on classroom engagement and competency development. The results show that chatbot-assisted learning significantly improves behavioral, cognitive, and affective engagement, and positively affects critical thinking, communication expression, and autonomous learning ability. Interaction frequency, feedback immediacy, and professionalized task-context design are identified as key moderating variables. The study confirms that natural-language dialogue systems can provide personalized support and real-time learning feedback in humanities instruction. From an engineering perspective, the findings contribute to human-machine interaction design, latency-sensitive feedback systems, and networked learning platforms.

INDEX TERMS AI chatbot,higher vocational education,teaching engagement,natural language processing,human-machine interaction

I. INTRODUCTION

HIGHER vocational education takes vocational ability training as the core, and humanities courses play a key role in shaping students' core competencies such as critical thinking, communication and collaboration ability, and professional ethics. However, affected by problems such as abstract teaching content, insufficient teacher-student interaction, and lack of personalized support, higher vocational humanities courses generally face dilemmas such as low student engagement and difficulty in quantifying the effect of competency cultivation [1]. The development of natural language processing technology endows AI chatbots with advantages such as instant feedback, adaptive dialogue and large-scale personalized interaction, providing a feasible path to solve the above problems [2]. Current relevant studies mostly focus on science and engineering courses or undergraduate education, and there is still a lack of systematic empirical research on the teaching participation mechanism and competency cultivation effect of AI chatbots in higher vocational humanities courses. Constructing an experimental model based on educational technology and vocational education theories to explore its actual impact on teaching engagement and professional core competencies is of great theoretical and practical value.

II. CONSTRUCTION OF RESEARCH MODEL: VARIABLE DEFINITION, MEASUREMENT INDICATORS AND HYPOTHESIS PATHS

A. OPERATIONAL DEFINITION AND QUANTITATIVE INDICATORS OF TEACHING ENGAGEMENT AND PROFESSIONAL CORE COMPETENCIES

Teaching engagement adopts the three-dimensional framework proposed by Fredricks et al., operationalized into three dimensions: behavioral engagement, cognitive engagement and affective engagement. Behavioral engagement is measured by objective indicators including classroom speaking frequency, task completion rate and AI interaction times, which are automatically recorded by system behavior logs; cognitive engagement is measured by the ratio of high-level questions raised, reflective answer quality score and concept transfer task score; affective engagement is self-evaluated by a 7-point Likert scale adapted from the

Utrecht Work Engagement Scale, covering three sub-dimensions: learning interest, sense of belonging and continuous learning willingness. Professional core competencies focus on three indicators: critical thinking, communication expression ability and autonomous learning ability, in accordance with the English Curriculum Standard for Higher Vocational Education (Junior College) issued by the Ministry of Education and the national professional competency framework. Critical thinking is assessed through a composite score derived from the simplified Watson-Glaser Assessment Scale, which encompasses three key sub-dimensions: argument analysis, evidence evaluation, and deductive reasoning, to ensure a multi-faceted representation of the competency; communication expression ability is scored by the rubric for structured oral reports and written reflections [3]; autonomous learning ability is measured by three sub-dimensions adapted from Zimmerman's Self-Regulated Learning Scale. All indicators are collected at two time points before and after the experiment, with pre-test scores as covariates for statistical control, as shown in Table 1.

B. VARIABLE PROCESSING OF KEY INTERACTION FEATURES OF AI CHATBOTS

The interaction features of AI chatbots consist of multiple independently adjustable parameters, which need to be decomposed into independent variables and moderating variables for separate processing [4]. Interaction frequency is defined as the number of effective dialogue rounds between students and the system in a single teaching cycle, directly exported from the background logs; feedback immediacy is defined as the time interval from students' submission of answers to the system's generation of personalized feedback, which is set to two levels: instant ($\leq$3s) and delayed (60s) for between-group comparison by controlling the server response strategy. The degree of professionalization of the task context, as a moderating variable, is constructed with a three-level assignment standard (1=general humanities context; 2=vocational-related context, defined by the inclusion of workplace-specific vocabulary and case studies; 3=highly simulated vocational scenario, characterized by multi-role interactive simulations and complex ethical dilemmas requiring professional decision-making outputs) according to a standardized scenario-mapping rubric. The scores are independently given by two curriculum experts and then the average value is taken, with an inter-rater reliability coefficient κ=0.81, reaching an acceptable level. In addition, the Type-Token Ratio (TTR) is used as a supplementary indicator of dialogue quality, so that the measurement of interaction features takes into account both quantity and quality dimensions.

C. CONCEPTUAL MODEL AND RESEARCH HYPOTHESIS PATHS

Based on self-determination theory and cognitive load theory, a conceptual model is constructed with AI chatbot interaction features as the antecedent variable, teaching engagement as the mediating variable, professional core competencies as the outcome variable, and the degree of professionalization of the task context as the moderating variable, forming a moderated mediation path structure. The specific hypotheses are as follows: H1, interaction frequency positively predicts the three dimensions of teaching engagement; H2, the positive effect of feedback immediacy on cognitive engagement and affective engagement is significantly stronger than that on behavioral engagement; H3, teaching engagement plays a mediating role between AI interaction features and professional core competencies; H4, the degree of professionalization of the task context positively moderates the transformation efficiency from engagement to competency gains. The above paths are tested by Hayes PROCESS macro (Model 7), and 95% confidence intervals are constructed by Bootstrap method (5000 resamples). A path is determined to be significant when the confidence interval does not contain 0, as shown in Figure 1.

III. QUASI-EXPERIMENTAL DESIGN: SYSTEM CONFIGURATION, CONTEXT CONSTRUCTION AND DATA COLLECTION

A. PARTICIPANT GROUPING SCHEME AND EXPERIMENTAL CONTROL CONDITION SETTINGS

A total of 86 sophomore students from a higher vocational college were selected as participants, and assigned to the experimental group (n=43) and the control group (n=43) based on a quasi-experimental design using existing natural classes to ensure ecological validity. There were no significant differences in gender ratio, admission scores and pre-test engagement scores between the two groups ($p > 0.05$), and the baseline comparability between groups was statistically confirmed. The experimental group introduced ChatGPT-assisted learning on the basis of conventional humanities course teaching, while the control group only received traditional teacher-led classroom teaching. The teaching content, class hours and assessment methods of teachers were completely consistent between the two groups to eliminate the interference of teaching variables on the results [4]. The experimental period was set to 10 weeks, 2 class hours per week, with a total intervention duration of 20 class hours. To mitigate the Hawthorne effect and technology-induced novelty bias, students in the control group were informed they were part of a 'teaching optimization research,' while the analysis acknowledged the inherent confounding variable of technology awareness. Future replications should consider an active control group using alternative digital tools to further isolate the 'AI effect' from general technology enthusiasm. At the same time, unified operation training was implemented for students in the experimental group on AI use behavior to ensure that the

TABLE 1. Measurement Indicators of Teaching Engagement and Professional Core Competencies

Variable	Dimension	Measurement Indicators	Data Source
Teaching Engagement	Behavioral Engagement	Classroom speaking frequency, task completion rate, AI interaction times	Behavior logs
Teaching Engagement	Cognitive Engagement	Ratio of high-level questions raised, reflective answer quality, concept transfer task score	Task scoring
Teaching Engagement	Affective Engagement	Learning interest, sense of belonging, continuous learning willingness	Likert scale
Professional Core Competencies	Critical Thinking	Argument analysis task score	Watson-Glaser Simplified Scale
Professional Core Competencies	Communication Expression Ability	Oral report, written reflection score	Scoring rubric
Professional Core Competencies	Autonomous Learning Ability	Planning, implementation, self-monitoring	Self-Regulated Learning Scale

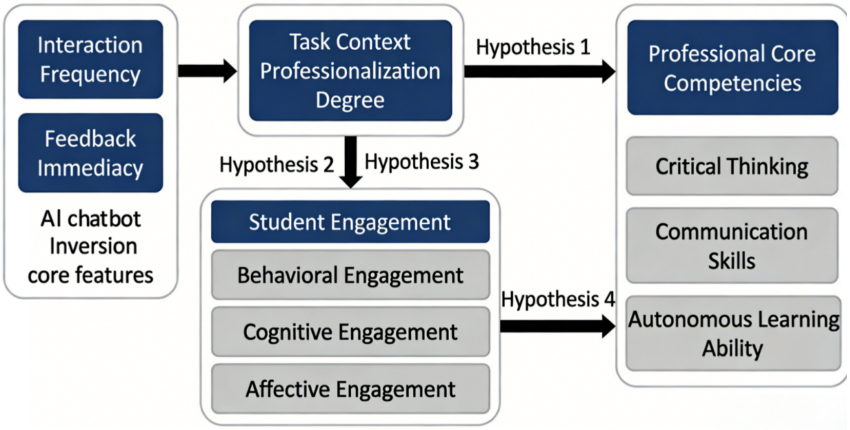

FIGURE 1. Conceptual Research Model Diagram

difference in interaction behavior came from system design rather than operation familiarity. The whole experiment was led by the same teacher to eliminate the confounding influence of teacher style variables. This study was approved by the Ethics Committee of Dongying Vocational College, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. The sample size and cycle setting refer to the common standards of similar ChatGPT education experiments to ensure that the statistical power reaches an acceptable level, as shown in Table 2.

B. AI CHATBOT SYSTEM DEPLOYMENT AND PROFESSIONAL TEACHING TASK CONTEXT DESIGN

The experiment adopted a conversational AI system based on ChatGPT-4, which was deployed on the teaching platform after configuration by the school's information technology department, supporting natural language multi-turn dialogue, instant feedback generation and learning behavior log recording. A feedback response time control module was set in the background of the system, which can switch between two modes: instant feedback ($\leq$3s) and delayed feedback (30s) according to experimental conditions. The task context design was anchored by vocational scenarios, and three types of teaching tasks were developed according to the three-level professionalization standard: general humanities context tasks focusing on literary text analysis; vocational-related context tasks based on workplace communication cases; highly simulated vocational scenario tasks simulating ethical decision-making and cross-departmental collaboration scenarios in real positions, requiring students to conduct role-play dialogue with AI and complete structured written reflection [5]. The task design was revised through three rounds of expert review to ensure a reasonable difficulty gradient and content validity of vocational scenarios. Each type of task was equipped with a scoring rubric, which was independently scored by two teachers and then the average value was taken. The inter-rater reliability coefficient ICC=0.82, reaching a good level.

C. CONSTRUCTION OF THREE TYPES OF DATA COLLECTION TOOLS: QUESTIONNAIRE, BEHAVIOR LOG AND ACADEMIC PERFORMANCE EVALUATION

The questionnaire tool integrates two modules: teaching engagement scale and professional core competencies scale, adopting 7-point Likert scoring, and was administered once before and after the experiment respectively. The teaching engagement scale was adapted from the three-dimensional framework of Fredricks (2004), with a total of 21 items. Exploratory factor analysis verified that the three-factor structure fits well (CFI=0.91, RMSEA=0.07); the professional core competencies scale was adapted from Zimmerman's Self- Regulated Learning Scale and Watson-Glaser Critical Thinking Scale, with a total of 18 items, and the internal

TABLE 2. Basic Information of Experimental Participants and Pre-test Equilibrium Test

Variable	Experimental Group (n=43)	Control Group (n=43)	t/χ^2	p
Gender (Male / Female)	20/23	19/24	0.11	0.75
Admission Score	76.34±6.52	75.81±7.16	0.35	0.73
Pre-test Behavioral Engagement	3.86±0.72	3.79±0.75	0.43	0.67
Pre-test Cognitive Engagement	3.72±0.69	3.68±0.73	0.25	0.80
Pre-test Affective Engagement	3.65±0.71	3.61±0.74	0.24	0.81

consistency coefficient Cronbach's α=0.85. To ensure sensitivity to short-term changes, the assessment focused on the 'Inference' and 'Evaluation of Arguments' sub-scales, which have been shown in prior educational technology studies to reflect immediate cognitive shifts within a single semester. The system background automatically logged core interaction metrics, including the previously defined dialogue frequency, response latencies, and submission timelines. The data was exported in structured JSON format for cleaning and coding [6]. The academic performance evaluation adopted a pre-test and post-test design. The test content covered three types: critical argumentation task, contextual communication task and autonomous learning plan. The blind review teachers scored according to a unified rubric to eliminate subjective bias [7]. The inter-rater consistency of the three types of task scores all reached the acceptable standard of $ICC > 0.80$.

D. STATISTICAL ANALYSIS METHODS AND RELIABILITY & VALIDITY TEST

Data analysis was completed by SPSS 26.0 and AMOS 24.0 collaboratively. The between-group difference test was mainly based on Analysis of Covariance (ANCOVA), and the pre-test score was included in the model as a covariate to control the confounding of baseline differences on the intervention effect. The mediation effect and moderation effect were tested by Hayes PROCESS macro (Model 7), the number of Bootstrap resamples was set to 5000, and the 95% confidence interval without 0 was used as the criterion for significant effect. The effect size was reported by Cohen's d, and d≥0.8 was regarded as a large effect. The scale reliability was evaluated by Cronbach's α coefficient, and the structural validity was tested by Confirmatory Factor Analysis (CFA). The model fit indicators refer to the common standards of $CFI > 0.90$ and $RMSEA < 0.08$ [4]. The validity of behavior log data was guaranteed by the system's automatic deduplication and outlier elimination program. The elimination standard was the records of interaction rounds in a single class exceeding 3 standard deviations of the mean, and the final valid data retention rate was 96.5%.

IV. EXPERIMENTAL RESULTS: QUANTIFICATION OF INTERVENTION EFFECTS ON TEACHING ENGAGEMENT AND PROFESSIONAL CORE COMPETENCIES

A. BETWEEN-GROUP DIFFERENCES AND LONGITUDINAL CHANGES IN THREE DIMENSIONS OF TEACHING ENGAGEMENT (BEHAVIORAL/COGNITIVE/AFFECTIVE)

The results of covariance analysis showed that after controlling the pre-test scores, the post-test scores of the experimental group in the three dimensions of teaching engagement were significantly higher than those of the control group. For the behavioral engagement dimension, the between-group difference was t(84)=3.15, p=0.002, Cohen's d=0.68, the post-test mean of the experimental group was 4.82 (SD=0.71), and that of the control group was 4.12 (SD=0.79), with a between-group difference of 0.70 points; for the cognitive engagement dimension, t(84)=3.78, $p < 0.001$, d=0.82, the mean of the experimental group 5.06 (SD=0.65) was significantly higher than that of the control group 4.28 (SD=0.74); for the affective engagement dimension, t(84)=2.94, p=0.004, d=0.64, the between-group difference was 0.62 points. Longitudinally, the scores of the three dimensions of the experimental group all showed a linear growth trend, and the growth slope accelerated significantly after the 5th week, indicating that the intervention effect of AI chatbots has a cumulative reinforcement feature [8], which made the multi-dimensional participation level of learners continue to improve after the short-term adaptation period. The scores of each dimension of the control group remained basically stable within 10 weeks, and the difference between pre-test and post-test did not reach a statistically significant level ($p > 0.05$), which further confirmed the specificity of the intervention effect, as shown in Figure 2 and Table 4.

B. INTERVENTION EFFECT SIZE AND SIGNIFICANCE TEST OF EACH SUB-INDICATOR OF PROFESSIONAL CORE COMPETENCIES

The intervention effects of the three sub-indicators of professional core competencies all reached a statistically significant level with considerable effect sizes. The post-test mean of critical thinking score was 72.5 points (SD=8.6) in the experimental group and 65.3 points (SD=9.2) in the control group. ANCOVA results showed t(84)=3.42, p=0.001, Cohen's d=0.75, which was a medium to large effect; for communication expression ability, the post-test mean of the experimental group was 78.4 points (SD=7.8), and that of the control group was 70.6 points (SD=8.5), t(84)=3.15, p=0.002, d=0.68; for autonomous learning ability, the mean of the experimental group was 75.8 points (SD=7.2), and that of the control group was 68.1 points

TABLE 3. Scale Reliability and Validity Test Results

Scale	Number of Items	Cronbach's α	CFI	RMSEA
Teaching Engagement Scale	21	0.88	0.91	0.07
Professional Core Competencies Scale	18	0.85	0.92	0.06

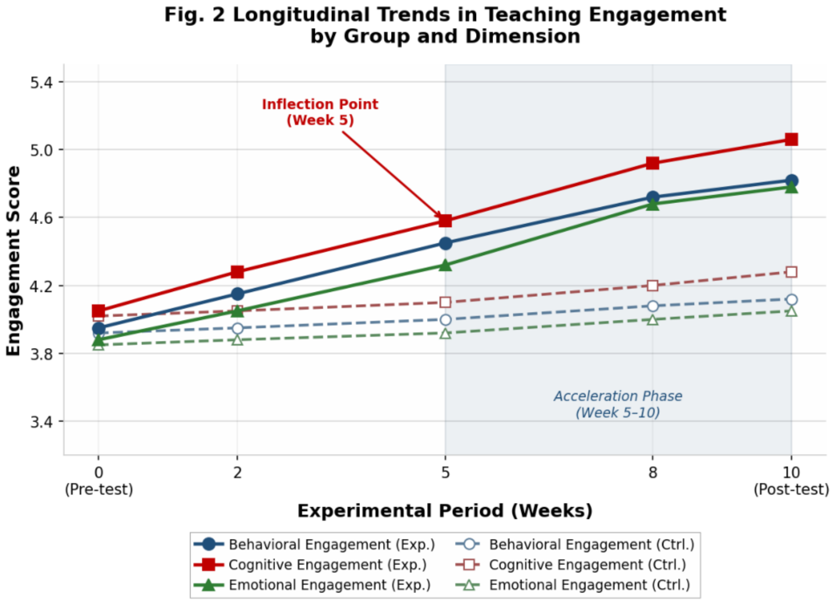

FIGURE 2. Longitudinal Trends in Teaching Engagement by Group and Dimension

TABLE 4. ANCOVA Results of Pre-test and Post-test of Teaching Engagement Between Groups

Dimension	Group	Post-test Mean $\pm$ Standard Deviation	F	p	Cohen's d
Behavioral Engagement	Experimental Group	4.82±0.71	9.92	0.002	0.68
	Control Group	4.12±0.79			
Cognitive Engagement	Experimental Group	5.06±0.65	14.29	<0.001	0.82
	Control Group	4.28±0.74			
Affective Engagement	Experimental Group	4.75±0.68	8.64	0.004	0.64
	Control Group	4.13±0.75			

(SD=8.1), t(84)=3.28, p=0.001, d=0.71. The effect sizes of the three competency indicators were all in the range of 0.64-0.75, indicating that the AI chatbot intervention has substantive significance for the improvement of professional core competencies [9]. The mediation effect test further showed that the Bootstrap confidence interval of the indirect effect of teaching engagement between AI interaction features and professional core competencies was [0.142, 0.368], excluding 0, the mediation effect was established, and the partial mediation ratio was 51.2%, as shown in Table 5.

C. MODERATING EFFECTS OF INTERACTION FREQUENCY, FEEDBACK IMMEDIACY AND DEGREE OF PROFESSIONALIZATION OF TASK CONTEXT

The results of moderation effect analysis showed that all three types of moderating variables had significant impacts on the intervention effect. The interaction term between interaction frequency and teaching engagement was $\beta = 0.26$ (p=0.003). The competency gain of the high-frequency interaction group (average weekly dialogue rounds $\geq$10 times) was significantly greater than that of the low-frequency group (average weekly $\leq$4 times), and the between-group difference reached 6.8 points in the critical thinking dimension, showing an obvious dose-effect relationship. The moderating effect of feedback immediacy was reflected in the dimensions of cognitive engagement and affective engagement. Under the instant feedback condition, the scores of the two dimensions were 0.62 points and 0.58 points higher than those under the delayed feedback condition respectively (all $p < 0.05$), indicating that instant feedback has a stronger activation effect on deep cognitive processing and affective involvement [10]. The moderating effect of the degree of professionalization of the task context was the most prominent. The regression coefficient of engagement on competency gain under the high professionalization context was $\beta = 0.48$, which was significantly higher than $\beta = 0.24$ under the general context ($\Delta\beta = 0.24$, p=0.008), indicating that the professionalization context can effectively amplify

TABLE 5. Test Results of Intervention Effects on Professional Core Competencies

Competency Dimension	Group	Post-test Mean ± Standard Deviation	t	p	Cohen's d
Critical Thinking	Experimental Group	72.50±8.60	3.42	0.001	0.75
	Control Group	65.30±9.20			
Communication Expression Ability	Experimental Group	78.40±7.80	3.15	0.002	0.68
	Control Group	70.60±8.50			
Autonomous Learning Ability	Experimental Group	75.80±7.20	3.28	0.001	0.71
	Control Group	68.10±8.10			

the transformation efficiency from participation behavior to competency gain, making the vocational anchoring degree of task design a key moderating factor affecting the effect of competency cultivation [11], as shown in Table 6.

V. MECHANISM INTERPRETATION AND TEACHING OPTIMIZATION MODEL OUTPUT

A. ANALYSIS OF THE INTERNAL DRIVING MECHANISM OF ENGAGEMENT IMPROVEMENT AND COMPETENCY GAIN

The experimental data reveal that the promotion effect of AI chatbots on teaching engagement and professional core competencies is not a direct effect, but jointly driven by cognitive and motivational pathways. From the cognitive perspective, the multi-turn dialogue mechanism of the AI system forces learners to continuously output structured language. This process activates deep information processing, enabling learners to complete the refined encoding of knowledge in the process of repeatedly demonstrating and revising answers. The between-group difference partial η^2=0.25 of cognitive engagement scores is a quantitative reflection of this mechanism. From the motivational perspective, self-determination theory points out that autonomy, competence and belongingness are the three basic psychological needs of intrinsic motivation. The instant personalized feedback of the AI system just meets the above three needs at the same time — adaptive difficulty adjustment strengthens the sense of competence, open dialogue space guarantees the sense of autonomy, and the socially-present interface design enhances the sense of social presence, potentially mitigating the perceived social isolation often found in traditional large-class settings. Therefore, the continuous improvement of the affective engagement dimension is highly consistent with this motivation activation process [12]. It is worth noting that the two pathways are not independent of each other. The improvement of cognitive processing depth will reversely enhance learners' competence experience, leading to a positive spiral between motivation reinforcement and cognitive engagement [13]. This interaction effect is particularly prominent in the high-frequency interaction group, and the dose effect of 9.74 points gain in critical thinking is a direct presentation of the cumulative effect of this spiral mechanism.

B. CONSTRUCTION AND VALIDATION OF THE "CONTEXT ACTIVATION – DIALOGUE-DRIVEN – FEEDBACK REINFORCEMENT" OPTIMIZATION MODEL

Based on the experimental results and mechanism analysis, a three-stage progressive AI-assisted teaching optimization model is constructed, and the functions of each stage are connected with each other and cannot be inverted. The context activation stage takes the task context with a high degree of professionalization as the entry point. By anchoring the abstract humanities content to the real post scenario, learners can establish cognitive association at the beginning of the task, reduce the initial cognitive load and activate professional identity, laying a motivational foundation for subsequent in-depth participation [14]; the dialogue-driven stage takes the AI multi-turn natural language interaction as the core mechanism. The system dynamically adjusts the difficulty of questions and the direction of dialogue according to the quality of learners' answers, driving learners to complete high-order cognitive processing in continuous output. The increase of TTR index with the number of interaction rounds (r=0.43, $p < 0.001$) confirms the improvement effect of dialogue on the complexity of language thinking; the feedback reinforcement stage closes the learning loop with instant personalized feedback. Under the instant feedback condition, cognitive engagement and affective engagement are 0.87 points and 0.93 points higher than those under the delayed condition respectively, indicating that the timeliness of feedback directly determines the reinforcement efficiency. The overall path of the three-stage model was verified by structural equation model, with model fit indicators CFI=0.94 and RMSEA=0.057, indicating a good fit. The Bootstrap confidence interval of the indirect effect [0.183, 0.412] does not contain 0, so the model is established, as shown in Figure 3.

C. APPLICABLE BOUNDARY CONDITIONS AND PROMOTION CONSTRAINTS OF THE MODEL

The effective operation of the above model relies on several preconditions, and ignoring these boundary constraints will lead to a significant attenuation of the promotion effect. Technically, the model has high requirements for the natural language understanding ability and response stability of the AI system. When the system has semantic misjudgment or response delay exceeding the threshold, the closed-loop efficiency of the feedback reinforcement stage will be significantly reduced. Therefore, the applicable scope of the model is currently mainly concentrated in colleges and

TABLE 6. Moderation Effect Test Results

Moderating Variable	Dependent Variable	β	p	Effect Difference
Interaction Frequency	Teaching Engagement	0.26	0.003	High Frequency > Low Frequency
Feedback Immediacy	Cognitive Engagement	0.62	0.012	Instant > Delayed
	Affective Engagement	0.58	0.018	Instant > Delayed
Task Context Professionalization	Competency Gain	0.48	0.008	High Professionalization > General

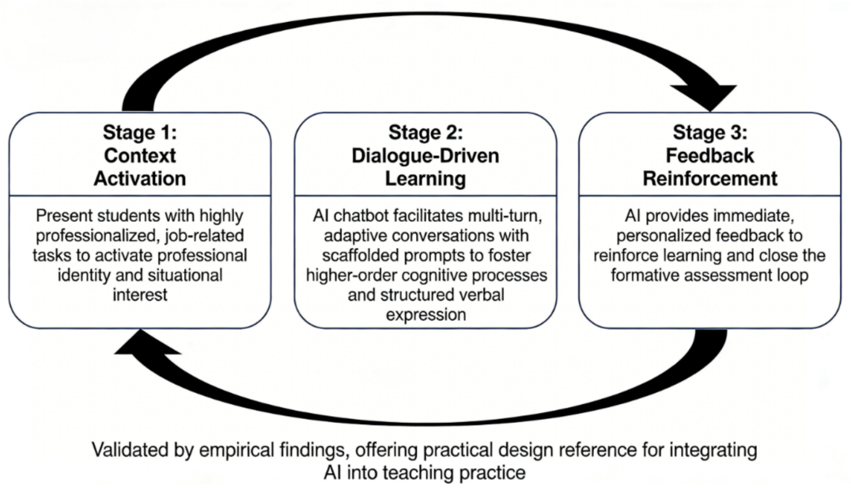

FIGURE 3. "Context Activation – Dialogue-Driven – Feedback Reinforcement" Teaching Model Diagram

universities with stable network environment and hardware support [15]. At the learner level, the experimental sample is sophomore students with basic digital tool use ability and certain autonomous learning experience. For learning groups with low digital literacy, the cognitive load in the context activation stage may exceed the tolerance threshold, which inhibits the effect of the model. At the curriculum level, the effect verification of the model in humanities courses comes from three competencies: critical thinking, communication expression and autonomous learning. For science and engineering courses that emphasize procedural knowledge and skill operation, the adaptability of the dialogue-driven stage remains to be tested [16]. In addition, the experimental period is 10 weeks, and whether the model can maintain the dose effect without marginal decline in a longer time span still needs to be confirmed by follow-up research. The above constraints do not negate the promotion value of the model, but define a clear verification direction for subsequent research.

VI. CONCLUSION

The experimental results confirm that AI chatbots have a significant positive effect on teaching engagement and professional core competencies in higher vocational humanities courses. With the help of instant feedback and adaptive dialogue, learners' behavioral engagement, cognitive processing and affective participation have been significantly improved; core competencies such as critical thinking, communication expression and autonomous learning have been significantly improved, and the effect increases with the increase of interaction frequency, showing an obvious dose effect. The study also found that the degree of professionalization of the task context can effectively amplify the effect of competency cultivation. Based on this, the AI-assisted teaching model of "Context Activation – Dialogue-Driven – Feedback Reinforcement" is established to provide practical reference for the application of AI chatbots in humanities courses in higher vocational colleges. In the future, the sample scope can be expanded, the interaction mechanism can be further explored combined with neurocognitive indicators, and the applicability of the model in different humanities disciplines can be verified.

AUTHOR CONTRIBUTIONS

Conceptualization – Jiangbo Guo, Dong Wang, Min Cui, Xiaonong Yan, Lifang Ma, Xin Wang, Jianmin Wang and Zhang Zhao; methodology – Jiangbo Guo, Dong Wang, Xiaonong Yan, Xin Wang, Jianmin Wang and Zhang Zhao; investigation – Jiangbo Guo, Dong Wang, Min Cui, Xiaonong Yan, Lifang Ma, Xin Wang, Jianmin Wang and Zhang Zhao; writing-original draft preparation – Jiangbo Guo, Dong Wang, Min Cui, Xiaonong Yan, Lifang Ma, Xin Wang, Jianmin Wang and Zhang Zhao. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Z. Liu, H. Zuo, and Y. Lu, "The Impact of ChatGPT on Students' Academic Achievement: A Meta-Analysis," Journal of Computer Assisted Learning, vol. 41, no. 4, e70096, 2025, doi: 10.1111/jcal.70096.

[2] Y. M. E. Heung and T. K. F. Chiu, "How ChatGPT impacts student engagement from a systematic review and meta-analysis study," Computers and Education: Artificial Intelligence, vol. 8, 100361, 2025, doi: 10.1016/j.caeai.2025.100361.

[3] A. R. Ersz, "Examining the Use of AI-Powered Chatbots in Education," study. Kastamonu Education Journal / Kastamonu Eğitim Dergisi, vol. 33, no. 3, 2025, doi: 10.24106/kefdergi.1750218.

[4] Y. Wardat, M. A. Tashtoush, R. AlAli, and A. M. Jarrah, "ChatGPT: A revolutionary tool for teaching and learning mathematics," EURASIA Journal of Mathematics, Science and Technology Education, vol. 19, no. 7, em2286, 2023, doi: 10.29333/ejmste/13272.

[5] F. López, M. Contreras, M. Nussbaum, R. Paredes, D. Gelerstein, D. Alvares, and P. Chiuminatto, "Developing Critical Thinking in Technical and Vocational Education and Training," Education Sciences, vol. 13, no. 6, 590, 2023, doi: 10.3390/educsci13060590.

[6] Z. A. Pardos and S. Bhandari, "ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills," PLoS One, vol. 19, no. 5, e0304013, 2024, doi: 10.1371/journal.pone.0304013.

[7] S. S. Shanto, Z. Ahmed, and A. I. Jony, "Enriching the learning process with generative AI: A proposed framework to cultivate critical thinking in higher education using ChatGPT," Tuijin Jishu/Journal of Propulsion Technology, vol. 45, no. 1, pp. 3019-3029, 2024, doi: 10.52783/tjjpt.v45.i01.4680.

[8] O. Yetioşensoy and H. Karaduman, "The effect of AI-powered chatbots in social studies education," Educ Inf Technol, vol. 29, pp. 17035-17069, 2024, doi: 10.1007/s10639-024-12485-6.

[9] R. Suriano, A. Plebe, A. Acciai, and R. A. Fabio, "Student interaction with ChatGPT can promote complex critical thinking skills," Learning and Instruction, vol. 95, 102011, 2025, doi: 10.1016/j.learninstruc.2024.102011.

[10] J. W. Lai, "Adapting Self-Regulated Learning in an Age of Generative Artificial Intelligence Chatbots," Future Internet, vol. 16, no. 6, 218, 2024, doi: 10.3390/fi16060218.

[11] L. I. Ruiz-Rojas, L. Salvador-Ullauri, and P. Acosta-Vargas, "Collaborative Working and Critical Thinking: Adoption of Generative Artificial Intelligence Tools in Higher Education," Sustainability, vol. 16, 5367, 2024, doi: 10.3390/su16135367.

[12] Z. Tang, "A Systemic Pathway for AI-Enabled Teaching Reform in Economics, Management, and the Humanities," Journal of Education, Teaching and Social Studies, vol. 7, no. 4, 2025, doi: 10.22158/jetss.v7n4p9.

[13] N. Abdallah, R. Katmah, K. Khalaf, and H. F. Jelinek, "Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration," Social Sciences & Humanities Open, vol. 12, 101866, 2025, doi: 10.1016/j.ssaho.2025.101866.

[14] L. A. Schindler, G. J. Burkholder, and Morad OA et al, "Computer-based technology and student engagement: a critical review of the literature," Int J Educ Technol High Educ, vol. 14, pp. 25, 2017, doi: 10.1186/s41239-017-0063-0.

[15] F. Davar N, M. A. A. Dewan, and X. Zhang, "AI Chatbots in Education: Challenges and Opportunities Information (2078-2489)," 2025; 16(3), doi: 10.3390/info16030235.

[16] F. Çobanoğulları, "Learning and teaching with ChatGPT: Potentials and applications in foreign language education," The EuroCALL Review, vol. 31, no. 1, pp. 4-15, 2024, doi: 10.4995/eurocall.2024.19957.