A Study on the Automatic Generation and Optimization of Chinese Language Teaching Content in Higher Education Based on Artificial Intelligence and Natural Language Processing

Jiangbo Guo[1], Dong Wang[1], Lihua Sui[1], Min Cui[1], Xin Wang[1], Xiaonong Yan[1], Huayan Shang[1], Jianmin Wang[1,*]
[1]Dongying Vocational College, Dongying 257000, Shandong, China
Corresponding author: Jianmin Wang (e-mail: dywjm1023@126.com).

ABSTRACT The automatic construction and dynamic optimization of Chinese language teaching resources in higher education have new technical ideas based on artificial intelligence and natural language processing technology. This study introduces an automatic content generation model which combines semantic encoding, knowledge enhancement, conditional control of learning objectives, and multi-granularity generation to solve the problems of knowledge bias in the generated content, lack of fit with learning objectives and lack of structural hierarchy. Moreover, a closed-loop optimization mechanism including quality evaluation, problem diagnosis and strategy refinement is set up. Experimental results show that the proposed model achieved BLEU-4, ROUGE-L, and BERTScore scores of 38.12%, 50.36%, and 0.887, respectively, representing improvements of 2.71, 2.74 percentage points, and 0.018 compared to T5; the knowledge accuracy rate and instructional objective alignment rate reached 91.8% and 90.7%, respectively, validating the model's effectiveness in terms of generation quality and instructional suitability.

INDEX TERMS Artificial intelligence; Natural language processing; Instructional content generation; Knowledge enhancement; Instructional adaptation

I. INTRODUCTION

Advances in artificial intelligence and natural language processing technologies have driven the shift in instructional resource development from manual authoring to data-driven generation, providing new technical pathways for the large-scale production, dynamic updating, and personalized adaptation of Chinese language course content in higher education. Yu [1] applied adaptive learning systems to university Chinese language education, validating the practical value of intelligent technologies in learning support and content matching; Shen [2] analyzed the effectiveness of natural language processing in the automatic generation of instructional resources, emphasizing the impact of semantic modeling on content quality. Huang et al. [3] reviewed the technological evolution and application challenges in the field of language education, pointing out that adaptation to teaching scenarios and content credibility remain key challenges. Law [4] summarized the application forms of generative technologies from a language teaching perspective, arguing that content generation must balance instructional objectives, learner diversity, and assessment feedback. Lu et al. [5] and Yang et al. [6] further discuss the resource expansion capabilities of generative content in higher education; however, research on knowledge constraints, hierarchical organization, and closed-loop optimization remains insufficient. Liu et al. [7] focus on application consensus in school education, while Liu and Xiao [8] reveal differences in teachers' technological engagement across different teaching stages, illustrating that generative content must be embedded within specific teaching workflows.

This study takes the teaching textbooks of Chinese as the research object, establishes the semantic encoding, knowledge enhancement, objective control, multi-granularity generation and feedback optimization technology framework. It's a framework that includes textbook explanation, literary appreciation, writing training, cultural knowledge, etc., which tends to consolidate the knowledge content, system structure and teaching applicability of the teaching content, and gives a measurable and verifiable methodological basis for the intelligent construction of Chinese language teaching materials in colleges.

II. MODELING THE EDUCATIONAL CONTENT GENERATION TASK

The automatic generation of Chinese language teaching content in higher education can be formalized as a text generation task under multiple conditional constraints. The model input consists of course themes, knowledge points, instructional objectives, learner levels, and content types, and combines these with an external knowledge base to construct a set of instructional conditions [9]. Let the input text be $X = \{x_1, \ldots, x_n\}$, the instructional control conditions be C, the retrieved knowledge information be K, and the target instructional text be $Y = \{y_1, \ldots, y_m\}$. Then, the generation process is represented as follows:

$$\begin{aligned} Y^* &= \arg\max_Y P(Y \mid X, C, K) \\ &= \arg\max_Y \prod_{t=1}^{m} P(y_t \mid y_{<t}, X, C, K) \end{aligned} \tag{1}$$

where n and m represent the lengths of the input and output sequences, respectively; y_t is the tth generated word; $y_{<t}$ denotes the sequence of words generated up to that point; $P(\cdot)$ represents the conditional probability; and Y^* is the optimal generation result that meets the instructional requirements [10]. Through stages such as semantic parsing, knowledge retrieval, condition fusion, content generation, and quality optimization, the model transforms the original instructional requirements into instructional text that is structurally complete, factually accurate, and appropriately challenging. The specific task workflow is shown in Figure 1.

III. MODEL FOR AUTOMATIC GENERATION OF INSTRUCTIONAL CONTENT

A. OVERALL MODEL ARCHITECTURE

The automatic instructional content generation model consists of an input representation layer, a semantic encoding layer, a knowledge enhancement layer, a conditional control layer, a multi-granularity generation layer, and a content optimization layer (see Figure 2). The input representation layer evenly represents the subjects of the course, the knowledge points, and the teaching objectives; the semantic encoding layer extracts the semantic features of the context; the knowledge enhancement layer adds external knowledge retrieval to the literary knowledge and course knowledge; the condition control layer combines the learners' level and the type of content; the generation layer generates paragraphs, example problems, and exercises; and the optimization layer optimizes the output according to the feedback of consistency, accuracy, and appropriateness.

B. SEMANTIC ENCODING OF INSTRUCTIONAL TEXTS

The semantic encoding module is based on 18,600 college Chinese language textbooks, lesson plans, and course resources, comprising a corpus of approximately 3.2 million tokens. These tokens were uniformly truncated or padded to 512 tokens, and a 32,000-word vocabulary was constructed [12]. The initial representation of input tokens is:

$$e_i = E(w_i) + P_i + T_i \tag{2}$$

where $E(w_i)$ is the embedding of the ith token, P_i is the positional encoding, and T_i is the text type encoding [13]. The encoder uses multi-head self-attention to extract contextual relationships:

$$H = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where Q, K and V represent the query, key, and value matrices, respectively; d_k denotes the key vector dimension; and H denotes the semantic feature matrix [14]. This process captures thematic relationships, knowledge hierarchies, and discourse structures in instructional texts, providing a unified semantic input for knowledge enhancement and conditional control.

C. KNOWLEDGE AUGMENTATION GENERATION MECHANISM

The knowledge enhancement module constructs a domain knowledge base based on curriculum standards, literary works, author information, and teaching cases, containing approximately 46,000 knowledge entities and 128,000 relationship triples. The model retrieves candidate knowledge based on the semantic encoding results, and the relevance score is expressed as:

$$s_i = \frac{h^T k_i}{\|h\|\|k_i\|} \tag{4}$$

where h is the semantic vector of the instructional text, k_i is the representation vector of the ith knowledge item, and s_i is the cosine similarity [15]. The top 10 knowledge items are selected based on their scores and weighted fusion is performed via an attention mechanism:

$$z = \sum_{i=1}^{K} a_i k_i \tag{5}$$

$$a_i = \frac{\exp(s_i)}{\sum_{j=1}^{K} \exp(s_j)} \tag{6}$$

In the equation, K represents the number of candidate knowledge items, a_i represents the normalized weight of the ith knowledge item, and z represents the fused knowledge features [16].

D. INSTRUCTIONAL OBJECTIVE CONDITION CONTROL

The instructional objective condition control module encodes four types of objectives—knowledge mastery, skill training, aesthetic understanding, and cultural identity—into control vectors, and sets three learner levels, four content types, and five difficulty levels [17]. The comprehensive conditions are expressed as:

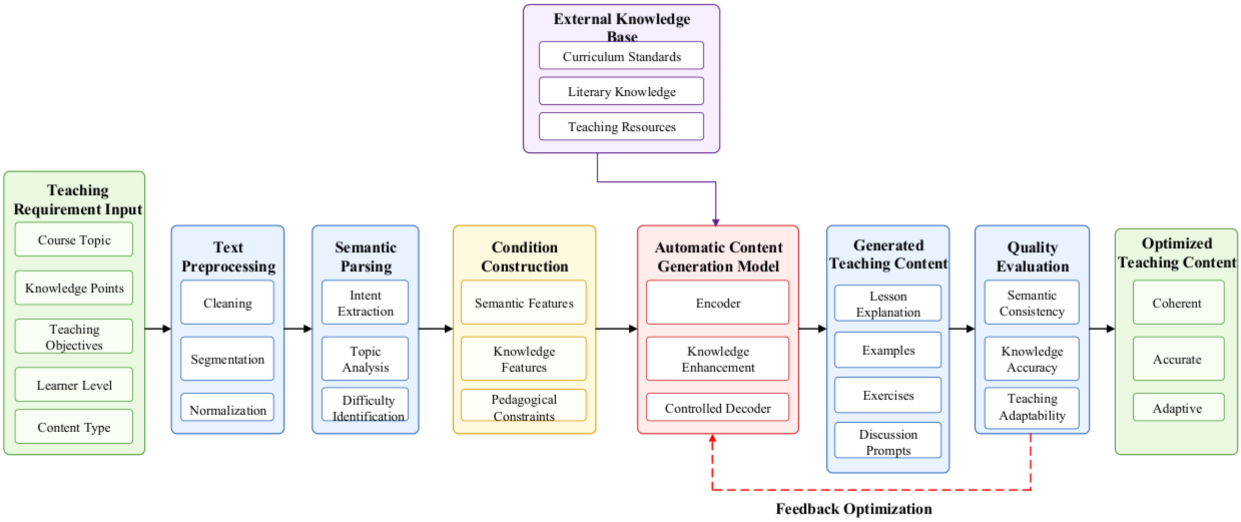

Fig. 1. Task Workflow for Automatic Instructional Content Generation.

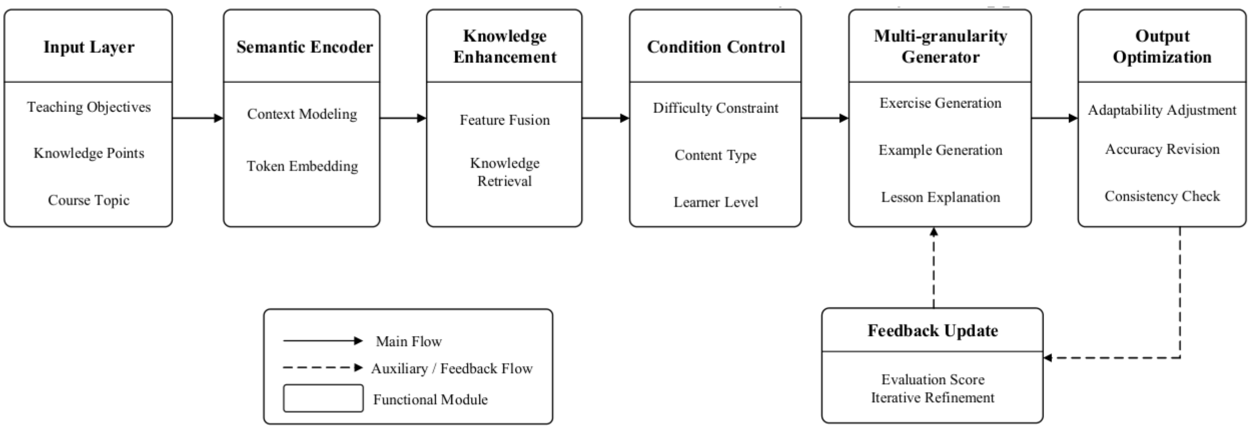

Fig. 2. Model Architecture Flowchart.

$$c = W_c[o; l; t; d] + b_c \tag{7}$$

where o, l, t and d are the vectors representing instructional objectives, learner levels, content types, and difficulty, respectively; $[\cdot]$ denotes vector concatenation; W_c and b_c are trainable parameters; and c is the comprehensive condition feature [18].

To modulate the influence of contextual information on semantic representations, a gated fusion mechanism is introduced:

$$g = \sigma(W_g[h; c] + b_g) \tag{8}$$

$$\widetilde{h} = g \cdot h + (1 - g) \cdot c \tag{9}$$

where h represents the text semantic features, g is the gating weight, σ denotes the Sigmoid function, $\cdot$ denotes element-wise multiplication, and $\widetilde{h}$ is the fusion representation constrained by the learning objective, which serves as the conditional input to the generative decoder [19].

E. MULTI-GRANULARITY CONTENT GENERATION STRATEGY

The content generation strategy of multi-granularity is to generate content step-by-step according to the four levels of "lexeme—sentence—paragraph—instructional unit", in which the prediction of candidate expressions is based on the result of semantic encoding in the lexeme layer, the syntactic completeness and knowledge consistency checking is done in the sentence layer, and the conceptual explanation and the guidance of problem-based are carried out in the paragraph layer, and the organic combination of the introduction, main body, exercise and summary is done in the instructional unit layer [20]. Shared semantic representation and inter-layer attention mechanism are used to transfer information between layers. The amount of content generated and the structure of it is dynamic according to the instructional goals, type of content, and difficulty level labels. This is shown in Figure 3.

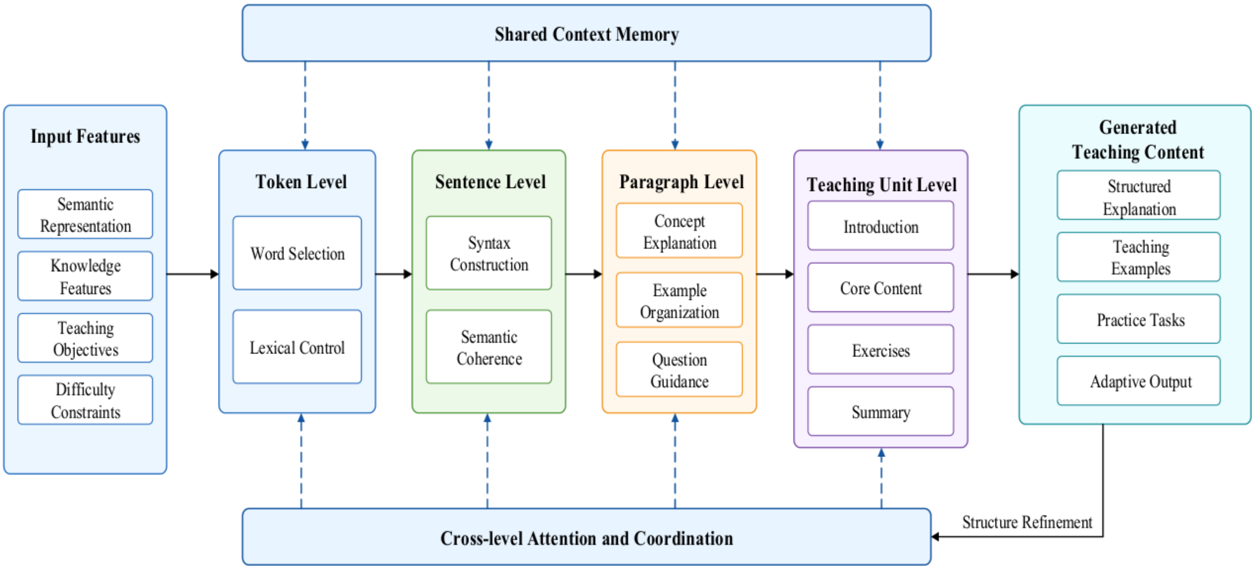

Fig. 3. Multi-granularity instructional content generation process.

F. MODEL TRAINING AND PARAMETER OPTIMIZATION

The model uses 18,600 instructional texts to construct training, validation, and test sets, divided in a 7:1:2 ratio. The batch size is set to 32, with a maximum of 30 training epochs, an initial learning rate of 2×10^{-5} , and a dropout rate of 0.1.

The training objective consists of the generation loss, knowledge consistency loss, and conditional control loss:

$$L = \lambda_1 L_{gen} + \lambda_2 L_{know} + \lambda_3 L_{ctrl} \tag{10}$$

where L_{gen}, L_{know} and L_{ctrl} represent the text generation, knowledge constraint, and instructional objective control losses, respectively, and λ_1, λ_2, λ_3 are the weight coefficients [21]. Parameters are optimized using AdamW with learning rate decay set as follows:

$$\eta_t = \eta_0 \left(1 - \frac{t}{T}\right) \tag{11}$$

where η_0 is the initial learning rate, t is the current training step, and T is the total number of steps [22]. Gradient clipping and an early stopping mechanism are combined to suppress overfitting; the relevant training parameters are shown in Table 1.

TABLE 1. Model Training Parameters and Loss Function Settings

Parameter Category	Parameter	Setting
Data Configuration	Training:Validation:Test Split	7:1:2
Input Configuration	Maximum Sequence Length	512
Training Configuration	Batch Size	32
Training Configuration	Maximum Number of Epochs	30
Optimization Configuration	Optimizer	AdamW
Optimization Configuration	Initial Learning Rate	2×10^{-5}
Regularization	Dropout Rate / Weight Decay	0.10 / 0.01
Training Stability	Gradient Clipping Threshold / Early-Stopping Patience	1.0 / 5
Loss Function	$\lambda_1, \lambda_2, \lambda_3$	0.60, 0.25, 0.15

IV. COURSE CONTENT: INTELLIGENT OPTIMIZATION METHODS

Intelligent optimization of instructional content emphasizes the output generated by the system, and takes "quality evaluation–problem identification–difficulty adjustment–regeneration" as a closed loop [23]. The system evaluates the content quality from five aspects: semantic consistency, knowledge accuracy, goal alignment, difficulty appropriateness and linguistic coherence to locate the problems of knowledge discrepancy, loose structure, word repetition and difficulty imbalance by low-scoring items. The optimization module calls the knowledge correction, content reorganization, redundancy compression and difficulty adjustment strategies for revising a candidate text partially or totally. The reworked results are then again sent back to the evaluation module until the comprehensive score reaches a pre-set threshold or the maximum number of iterations is attained, ensuring the instructional content is stable in terms of knowledge, structure and instructional suitability. The particular closed-loop system is displayed in Figure 4.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT CONFIGURATION

The experiments were performed on the Ubuntu 20.04 system using hardware platform of Intel Core i9 processor, 64 GB of RAM and NVIDIA RTX 3090 GPU. The software environment consisted of Python 3.10, PyTorch 2.0, Transformers 4.31 and CUDA 11.8. The experimental data were 18,600 college Chinese language textbooks, lesson plans, curriculum standards and teaching cases, which were divided into training set, validation set and test set according to the ratio of 7:1:2 [24]. The input sequence length was fixed to 512, the batch size fixed to equal 32 and the number

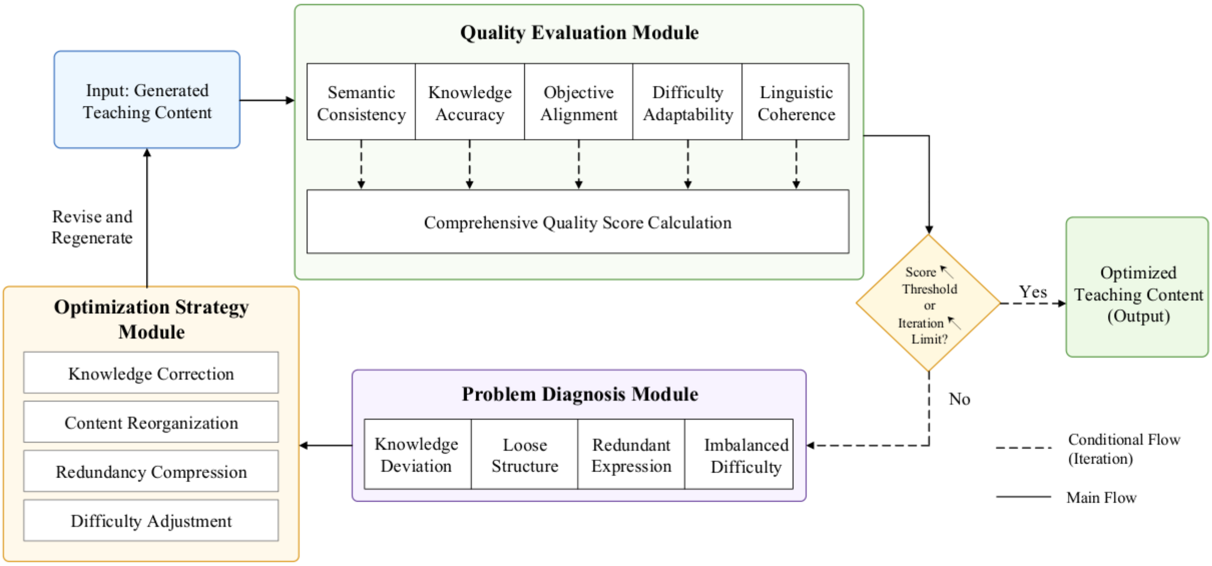

Fig. 4. Closed-Loop Framework for Intelligent Optimization of Instructional Content.

TABLE 2. Experimental Environment and Dataset Configuration

Configuration Category	Configuration Item	Parameter Setting
Hardware Environment	Processor	Intel Core i9
Hardware Environment	GPU / Memory	NVIDIA RTX 3090 / 64 GB
Software Environment	Operating System	Ubuntu 20.04
Software Environment	Development Framework	Python 3.10, PyTorch 2.0
Software Environment	CUDA / Transformers	CUDA 11.8 / Transformers 4.31
Dataset Configuration	Data Sources	University Chinese textbooks, lesson plans, curriculum standards, and teaching cases
Dataset Configuration	Dataset Scale	18,600 texts, approximately 3.2 million tokens
Dataset Configuration	Dataset Split	Training:Validation:Test = 7:1:2
Training Configuration	Sequence Length / Batch Size	512 / 32
Training Configuration	Maximum Epochs / Random Seed	30 / 42

of training iterations fixed to 30. To ensure experimental reproducibility, the random seed was set to 42 and AdamW optimizer was used.

B. DESIGN OF COMPARATIVE EXPERIMENTS

We selected Transformer, BART, T5, and GPT-2 as comparison models and conducted training and testing under identical dataset splits, input lengths, and training epochs. All models uniformly used the AdamW optimizer with a batch size of 32 and a fixed random seed of 42 to minimize the impact of parameter configuration differences on experimental results [25]. Content generation quality was evaluated using BLEU, ROUGE-L, and BERTScore; instructional adaptability was measured by knowledge accuracy, alignment with instructional objectives, difficulty matching, and teachers' comprehensive ratings; and statistical metrics such as the number of parameters, latency per text generation, and GPU memory usage were recorded. Each set of experiments was run independently five times, and results are reported as mean and standard deviation.

C. COMPARISON OF CONTENT GENERATION QUALITY

Comparative experiments were conducted to evaluate the text generation capabilities of different models for college Chinese language teaching texts under identical data split, training parameters, and decoding settings. BLEU-4, ROUGE-L, BERTScore, and perplexity were used to measure the lexical overlap, semantic consistency, and linguistic fluency of the generated texts. The results are shown in Table 3.

TABLE 3. Comparison of Generation Quality Across Different Models

Model	BLEU-4/% ↑	ROUGE-L/% ↑	BERTScore ↑	Perplexity ↓
Transformer	31.84±0.42	44.17±0.38	0.842±0.004	22.6±0.7
BART	34.26±0.37	46.53±0.41	0.861±0.003	19.8±0.5
T5	35.41±0.33	47.62±0.35	0.869±0.003	18.9±0.4
GPT-2	32.97±0.46	45.08±0.44	0.853±0.005	21.1±0.6
Proposed Model	38.12±0.29	50.36±0.31	0.887±0.002	16.7±0.3

The proposed model achieved BLEU-4 and ROUGE-L scores of 38.12% and 50.36%, respectively, representing improvements of 2.71 and 2.74 percentage points over T5; The BERTScore improved from 0.869 to 0.887, and the perplexity decreased to 16.7. The results indicate that knowledge enhancement and multi-granularity generation can improve semantic coherence in text and reduce topic drift and expressive inconsistencies.

D. ANALYSIS OF INSTRUCTIONAL ADAPTATION EFFECTIVENESS

The evaluation was conducted across four dimensions: knowledge accuracy, alignment with instructional objectives, difficulty adaptation, and teachers' comprehensive ratings. The test samples covered tasks such as literary appreciation, language knowledge, writing practice, and cultural understanding; the results are shown in Table 4.

The proposed model achieved a knowledge accuracy rate of 91.8%, an improvement of 3.9 percentage points over T5; the instructional objective alignment rate and difficulty fit reached 90.7% and 89.6%, respectively, with a teacher

TABLE 4. Comparison of Instructional Adaptation Performance Across Different Models

Model	Knowledge Accuracy/%	Objective Matching/%	Difficulty Adaptation/%	Teacher Score
Transformer	82.6±1.2	80.8±1.5	79.4±1.4	3.72±0.11
BART	85.9±1.0	84.7±1.2	83.5±1.1	3.96±0.09
T5	87.9±0.9	86.4±1.0	85.7±1.2	4.16±0.08
GPT-2	83.8±1.3	82.5±1.4	81.2±1.5	3.81±0.12
Proposed Model	91.8±0.7	90.7±0.8	89.6±0.9	4.41±0.06

TABLE 5. Results of Model Ablation Experiments

Model Configuration	BLEU-4/%	BERTScore	Knowledge Accuracy/%	Objective Matching/%
Full Model ($K = 10$)	38.12±0.29	0.887±0.002	91.8±0.7	90.7±0.8
Without Knowledge Enhancement	35.06±0.41	0.864±0.004	84.3±1.1	89.6±1.0
Without Objective Control	36.21±0.38	0.872±0.003	90.6±0.9	82.8±1.3
Without Multi-granularity Generation	35.48±0.44	0.868±0.004	89.4±1.0	87.2±1.1
Without Closed-loop Optimization	36.67±0.35	0.878±0.003	88.1±0.9	88.5±1.0
Knowledge Retrieval ($K = 5$)	37.34±0.32	0.881±0.003	89.9±0.8	89.8±0.9
Knowledge Retrieval ($K = 15$)	37.76±0.31	0.884±0.003	91.2±0.8	90.1±0.8

TABLE 6. Comparison of Model Efficiency

Model	Parameters/M	FLOPs/G	Generation Latency/ms	Peak GPU Memory/GB	Throughput/(samples · s^{-1})
Transformer	118	21.4	64.7±1.8	5.2	15.4
BART	139	28.9	72.8±2.1	6.1	13.7
T5	223	41.6	94.5±2.7	8.4	10.6
GPT-2	124	24.7	68.9±2.0	5.7	14.5
Proposed Model	246	39.8	88.2±2.3	8	11.3

rating of 4.41. Compared to Transformer, its teacher rating improved by 0.69 points, indicating that the condition control mechanism can effectively align instructional objectives with content difficulty.

E. ABLATION EXPERIMENTS AND PARAMETER ANALYSIS

The experiments used the same training set, optimizer, and random seed across all runs, focusing on evaluating generation quality and the model's ability to adhere to instructional constraints. The results are shown in Table 5.

After removing the knowledge enhancement module, the knowledge accuracy rate dropped from 91.8% to 84.3%; after removing the target control module, the target matching rate fell to 82.8%. When the number of retrieved items increased from 5 to 10, the BLEU-4 score improved by 0.78 percentage points; however, when it was further increased to 15, the score decreased to 37.76%, indicating that a retrieval count of 10 strikes a good balance between knowledge coverage and information redundancy.

F. MODEL EFFICIENCY AND SYSTEM PERFORMANCE

Under the same hardware platform and generation length, the number of statistical model parameters, computational load, latency per text generation, peak video memory usage, and throughput reflect the deployment costs of the model in batch-generation scenarios for educational content. The results are shown in Table 6.

The proposed model has 246 M parameters, with a peak GPU memory usage of 8.0 GB. The latency for generating a single text sample is 88.2 ms, which is lower than T5's 94.5 ms; its throughput reaches 11.3 samples/s. Although knowledge retrieval and conditional control increase the parameter scale, computational complexity is kept at 39.8 G through shared semantic encoding and candidate knowledge truncation.

With 1, 10, 30, 50, and 100 concurrent users configured to generate requests of uniform length for instructional content in succession, we measured the average response time, P95 latency, throughput, resource utilization, and error rate. The test results are shown in Table 7.

TABLE 7. Comparison of System Performance

Concurrent Users	Average Response Time/s	P95 Latency/s	Throughput (requests · s^{-1})	CPU Utilization/%	GPU Utilization/%	Error Rate/%
1	0.91	1.08	1.09	18.6	31.4	0
10	1.06	1.31	8.92	34.2	58.7	0
30	1.28	1.62	21.4	52.8	74.9	0.07
50	1.56	2.03	28.7	67.5	86.1	0.18
100	2.41	3.36	34.9	82.7	94.3	0.63

As the number of concurrent users increased from 1 to 50, the system throughput rose from 1.09 transactions/s to 28.70 transactions/s, with an average response time of 1.56 s and an error rate of only 0.18%. When the number of concurrent users reached 100, the throughput increased to 34.90 transactions/s, with an average response time of 2.41 s and an error rate of 0.63%; the system continued to operate stably.

VI. CONCLUSION

In response to the problems of generation and optimization of Chinese language teaching content in higher education, a technical framework of "semantic encoding" – "knowledge enhancement" – "teaching goal setting" – "multi-granularity Chinese language teaching content generation" – "closed-loop Chinese language teaching content quality evaluation" was designed, which can continuously process from requirement analysis to Chinese language teaching content generation, and then to Chinese language teaching content quality optimization. The main contribution is to combine domain knowledge constraints, teaching-objective control and hierarchical generation mechanisms in a single system, and to use multidimensional evaluation feedback to enhance knowledge accuracy, structural integrity and the pedagogical adaptability. However, the current model is still hindered by restricted coverage of its corpora, subjectivity of teacher's evaluation and insufficient cross-course transferability, and still cannot be fully used to represent complex context of literature and implicit intentions of teaching. Future research should increase the size of the multi-institutional and multidisciplinary corpora, introduce the dynamic updating of knowledge, individualized student profile, long-term feedback mechanism in teaching, and enhance the generalizability, interpretability and applicability of the framework in higher education.

REFERENCES

[1] M. Yu, "Application of an Artificial Intelligence-based adaptive learning system to chinese language education in universities," in Proceedings of the 2024 International Symposium on Artificial Intelligence for Education, 2024, pp. 416–421.

[2] Y. Shen, "Exploring the effectiveness of natural language processing technologies in automatically generating English teaching resource content," Journal of Computational Methods in Sciences and Engineering, vol. 25, no. 5, pp. 4817–4832, 2025.

[3] X. Huang, D. Zou, G. Cheng, et al., "Trends, research issues and applications of artificial intelligence in language education," Educational Technology & Society, vol. 26, no. 1, pp. 112–131, 2023.

[4] L. Law, "Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review," Computers and Education Open, vol. 6, p. 100174, 2024.

[5] G. Lu, N. B. Hussin, and A. Sarkar, "Navigating the future: Harnessing artificial intelligence generated content (AIGC) for enhanced learning experiences in higher education," in 2024 international conference on advances in modern age Technologies for Health and Engineering Science (AMATHE). IEEE, 2024, pp. 1–12.

[6] S. Yang, S. Yang, and C. Tong, "In-depth application of artificial intelligence-generated content AIGC large model in higher education," Adult and Higher Education, vol. 5, no. 19, pp. 9–16, 2023.

[7] M. Liu, Y. Ren, L. M. Nyagoga, et al., "Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools," Future in Educational Research, vol. 1, no. 1, pp. 72–101, 2023.

[8] X. Liu and Y. Xiao, "Chinese university teachers' engagement with generative AI in different stages of foreign language teaching: A qualitative enquiry through the prism of ADDIE," Education and Information Technologies, vol. 30, no. 1, pp. 485–508, 2025.

[9] L. Wang and B. Ren, "Enhancing academic writing in a linguistics course with Generative AI: An empirical study in a higher education institution in Hong Kong," Education Sciences, vol. 14, no. 12, p. 1329, 2024.

[10] T. Alqahtani, H. A. Badreldin, M. Alrashed, et al., "The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research," Research in social and administrative pharmacy, vol. 19, no. 8, pp. 1236–1242, 2023.

[11] O. Yunina, "Artificial intelligence tools in foreign language teaching in higher education institutions," The modern higher education review, no. 8, pp. 77–90, 2023.

[12] S. Alblooshi, "Artificial intelligence in higher education, opportunities, and challenges: A review," in Frontiers in Education. Frontiers, vol. 10, 2026, p. 1683968.

[13] J. Xia, Y. Ge, Z. Shen, et al., "The auxiliary role of artificial intelligence applications in mitigating the linguistic, psychological, and educational challenges of teaching and learning chinese language by non-chinese students," International Review of Research in Open and Distributed Learning, vol. 25, no. 3, pp. 116–133, 2024.

[14] Q. Huang, W. Li, M. Mokhtar Bin Muhamad, et al., "University english teaching evaluation using artificial intelligence and data mining technology," Scientific reports, vol. 15, no. 1, p. 30297, 2025.

[15] H. A. Younis, N. I. R. Ruhaiyem, W. Ghaban, et al., "A systematic literature review on the applications of robots and natural language processing in education," Electronics, vol. 12, no. 13, p. 2864, 2023.

[16] X. Li, "A personalized teaching system for college English based on big data and artificial intelligence," Scalable Computing: Practice and Experience, vol. 25, no. 6, pp. 5477–5485-5477–5485, 2024.

[17] H. Ma, L. Ismail, and W. Han, "A bibliometric analysis of artificial intelligence in language teaching and learning (1990-2023): evolution, trends and future directions," Education and Information Technologies, vol. 29, no. 18, pp. 25211–25235, 2024.

[18] H. Wu, S. Li, Y. Gao, et al., "Natural language processing in educational research: The evolution of research topics," Education and Information Technologies, vol. 29, no. 17, pp. 23271–23297, 2024.

[19] H. Crompton and D. Burke, "Artificial intelligence in higher education: the state of the field," International journal of educational technology in higher education, vol. 20, no. 1, pp. 1–22, 2023.

[20] W. Liu, Y. Lin, and Z. Yu, "A bibliometric analysis of artificial intelligence chatbots in language education," Social Education Research, pp. 59–78, 2024.

[21] A. W. Ou, C. Stöhr, and H. Malmström, "Academic communication with AI-powered language tools in higher education: From a post-humanist perspective," System, vol. 121, p. 103225, 2024.

[22] C. Zhou, "Integration of modern technologies in higher education on the example of artificial intelligence use," Education and Information Technologies, vol. 28, no. 4, pp. 3893–3910, 2023.

[23] J. C. Ruano-Borbalan, "The transformative impact of artificial intelligence on higher education: A critical reflection on current trends and futures directions," International Journal of Chinese Education, vol. 14, no. 1, p. 2212585X251319364, 2025.

[24] S. Ocen, J. Elasu, S. M. Aarakit, et al., "Artificial intelligence in higher education institutions: Review of innovations, opportunities and challenges," in Frontiers in Education. Frontiers Media SA, vol. 10, 2025, p. 1530247.

[25] R. Sahar and M. Munawaroh, "Artificial intelligence in higher education with bibliometric and content analysis for future research agenda," Discover Sustainability, vol. 6, no. 1, p. 401, 2025.