ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

Optimization of Reinforcement Learning in Personalized Teaching Mode of College English Classroom under the OBE Concept

Ting Li[1,a,*]

[1]Haojing College of Shaanxi University of Science and Technology, Xi'an, 712046, Shaanxi, China
[a]120361787@qq.com
[*]Corresponding author

Abstract: *In current college English teaching, traditional teaching models often ignore students' personalized needs and learning differences, resulting in uneven learning results and difficulty in meeting the student-centered educational goals under the OBE (Outcome-Based Education) concept. This study adopts an experimental design. First, students' learning data, including vocabulary, grammar mastery, reading comprehension and listening level, are collected through preliminary tests. Then, a personalized learning model is constructed using the DQN algorithm. The specific steps include: initializing the Q-value table and setting the learning rate and discount factor; selecting the optimal teaching strategy based on the student's current state; executing the strategy and observing the changes in the student's learning effect; updating the Q-value table and iteratively optimizing the strategy; converging the model through multiple rounds of training. Finally, through comparative experiments, the personalized teaching model optimized by reinforcement learning (RL) significantly improves students' academic performance. The average score of the proposed method increases by 12.72 points, the average score in the vocabulary test increases by 15.12 points, the reading comprehension ability increases by 5.72 points, and the listening level increases by 5.68 points. Reinforcement learning can effectively optimize the personalized teaching model in college English classrooms under the OBE concept and significantly improve students' vocabulary, reading comprehension and listening level.*

Keywords: *College English classroom; Personalized teaching model optimization; Reinforcement learning; DQN; Discount factor*

1. Introduction

In current college English teaching, the traditional teaching model generally adopts a "one-size-fits-all" teaching method, which is difficult to meet the diverse learning needs of students and leads to uneven learning results. This model cannot effectively achieve the student-centered educational goal under the OBE concept. With the rapid development of educational technology, personalized teaching has gradually become an important direction to improve teaching effectiveness. Reinforcement learning, as a technology that can dynamically optimize decision-making, provides new possibilities for personalized teaching. However, the research on applying reinforcement learning to college English classes is still in the initial exploratory stage. This paper will build an effective model and verify its actual effect.

This study aims to introduce reinforcement learning technology into college English classes and build a personalized teaching model based on DQN to solve the problem of insufficient personalization in traditional teaching. This study provides new technical means for college English teaching and has great application value.

This paper first applies reinforcement learning technology to college English classes under the OBE concept, achieving dynamic optimization and personalized adjustment of teaching strategies. Secondly, by defining multidimensional state space and action space, a complete personalized learning model framework is constructed, which can fully cover students' language ability improvement needs. Finally, the teaching model optimized by reinforcement learning significantly improves students' average scores (increases by 12.72 points), vocabulary (increases by 15.12 points), reading comprehension ability (increases by 5.72 points) and listening level (increases by 5.68 points).

ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

2. Related Work

In the field of English teaching, various innovative models emerge in an endless stream, providing new ideas for personalized teaching. Cai and Liu [1] integrated artificial intelligence tools into the teaching process and developed an intelligent evaluation system to improve the efficiency and accuracy of college English classroom teaching evaluation. On this basis, they comprehensively improved the quality of college English teaching and student learning effects. Ta and Li [2] implemented a personalized teaching model based on individual differences of students and combined modern information technology. Zhao et al. [3] selected 10 key factors affecting the operation effect of smart classrooms in colleges and universities from the aspects of classroom environment and teaching objectives to explore the personalized teaching mode of smart classrooms in colleges and universities, and used structural equation model to establish dynamic learning path. Sun [4] believed that college English teaching supported by AI is richer and higher quality, effectively improving the shortcomings of traditional teaching, improving teaching efficiency and quality, and promoting the development of students' learning ability and core literacy. Lin [5] designed a teaching model based on the teaching process before, during and after class, and applied this model in practice by taking the vocational English oral course as an example. Zhao [6] investigated the current situation of English teaching and summarized the existing problems. Lu [7] implemented the case in classroom learning based on MMFU. Ahmed and Mikail [8] adopted a new model to use the right teaching method for the right students at the right time. Wu et al. [9] compared blended teaching with traditional single teaching. Liu et al. [10] constructed an online learning path model. The combination of artificial intelligence and personalized teaching mode not only improves teaching efficiency and quality but also promotes the comprehensive development of students' core literacy. This paper will use reinforcement learning to further optimize it.

3. Methods

3.1 Data Collection

During the data collection stage, this study obtains students' learning data through preliminary tests, covering four key dimensions: vocabulary, grammar mastery, reading comprehension, and listening level. The vocabulary test uses a standardized vocabulary assessment tool, requiring students to complete vocabulary recognition and interpretation tasks within a certain period of time. The grammar mastery test assesses students' mastery of English grammar rules through grammar fill-in-the-blank questions and sentence correction questions. The reading comprehension test selects a number of English articles of moderate difficulty, requiring students to answer relevant questions to assess their comprehension ability. The listening level test assesses students' listening comprehension ability by playing English listening materials. All tests are conducted in a unified experimental environment to ensure the reliability and consistency of the data. Some of the collected data are shown in Table 1:

Table 1: Partial data

Student ID	Vocabulary (Out of 100)	Grammar Proficiency (Out of 100)	Reading Comprehension (Out of 100)	Listening Proficiency (Out of 100)
001	78	65	72	68
002	85	70	80	75
003	62	58	65	60
004	90	82	88	85
005	74	69	76	70

There are obvious differences in students' scores. Some students perform poorly in vocabulary, grammar mastery, reading comprehension and listening skills. The overall score distribution is uneven, indicating that students' mastery of different language skills is uneven. In addition, some students' grammar and listening scores are generally low, reflecting that grammar rules and listening training may be neglected in teaching. Further optimization of teaching strategies is needed to make up for these deficiencies.

3.2 Model Construction

In the model construction stage, this study designs a personalized learning model based on the DQN algorithm, the core of which is to dynamically optimize teaching strategies through reinforcement

ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

learning. The state space is defined as the student's multidimensional learning data, including vocabulary S_1, grammar mastery S_2, reading comprehension ability S_3 and listening level S_4, forming a state vector $S=[S_1,S_2,S_3,S_4]$. The action space covers the adjustment of teaching strategies, such as adding grammar-specific exercises A_1, expanding vocabulary training A_2, increasing the difficulty of reading materials A_3, or strengthening listening tasks A_4, forming a discrete action set $A=\{A_1,A_2,A_3,A_4\}$. The reward function is designed based on the improvement of learning effect [11-12], and the specific formula is:

$$R = \alpha * \Delta_{score} + \beta(1 - \frac{T_{learn}}{T_{total}}) \tag{1}$$

Among them, Δ_{score} is the increment of the staged test score, α, β are the weight coefficient (set to 0.7 and 0.3, respectively) to balance the improvement of scores and learning efficiency. The DQN model adopts a dual neural network architecture (policy network and target network), reduces data correlation interference through the experience replay mechanism, and uses the mean square error loss function to update network parameters:

$$L(\theta) = E[(r + \gamma \max_{a'} Q(s', a'; \theta^{-}))^2] \tag{2}$$

Among them, θ and θ^{-} are the parameters of the policy network and target network, respectively, and γ is the discount factor. During the model training process, action a is selected according to the current states, and the new state s′ and reward rare observed after execution. The experience (s, a, r, s′) is stored in the playback buffer, and the policy network is iteratively optimized by randomly sampling batch data. The model converges after multiple rounds of training and can adaptively generate the optimal teaching strategy for different students. Figure 1 shows the architecture of the DQN model in this paper:

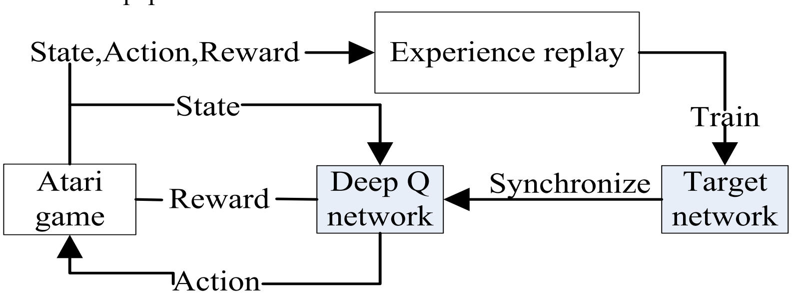

Figure 1: DQN model architecture

3.3 Strategy Optimization Steps

3.3.1 Initialize Q value table

In the initialization stage, a Q-value table is first constructed to store the expected cumulative reward of each state-action pair. The state space is defined by the students' learning data (vocabulary, grammar mastery, reading comprehension ability, listening level), and the action space includes adjusting teaching strategies and content (increasing grammar exercises, expanding vocabulary training, increasing reading difficulty, and strengthening listening tasks) [13-14]. The Q value table is initialized to a zero matrix to ensure that the model explores all possible actions without bias in the early stages of training. The learning rate is set to 0.7 to control the step size of the Q value update to ensure that the model strikes a balance between fast convergence and stability. The discount factor is set to 0.3 to weigh the importance of current rewards and future rewards, ensuring that the model takes into account both short-term effects and long-term goals when optimizing teaching strategies. This initialization process enables the model to gradually optimize the personalized teaching mode of college English classes.

3.3.2 Select the optimal teaching strategy based on the student's current status

When selecting the optimal teaching strategy based on the student's current state, the model first maps the student's learning data into a state vector S, and calculates the Q value of each possible action

ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

through the policy network. A ε-greedy strategy is adopted to select random actions with a certain probability ε (initial value is 0.9, which gradually decays with training) to explore new strategies, or to select the action with the highest current Q value with probability $1-\varepsilon$ to utilize the known optimal strategy [15-16]. For example, if a student has a low vocabulary but good grammar, the model may choose to expand vocabulary training as the optimal action. By dynamically adjusting the value, the model strikes a balance between exploration and utilization, gradually optimizing personalized teaching strategies, and ensuring that the adjustment of teaching content and methods can maximize the improvement of students' learning effects.

3.3.3 Implement strategies and observe changes in students' learning outcomes

In the strategy execution stage, the actions selected by the model based on the current state are applied to actual teaching, and teachers adjust teaching content and strategies based on the model's recommendations. After a period of teaching, students' learning data are re-evaluated through periodic tests to observe changes in their learning effects. If the model recommends adding grammar-specific exercises, the student's grammar mastery in subsequent tests should be improved. By comparing the pre- and post-test data, the model can quantify the effect of strategy execution and feed this data back into the Q value update process. Table 2 shows the change in learning effect of some students:

Table 2: Changes in learning outcomes

Student ID	Vocabulary Change (%)	Grammar Mastery Change (%)	Reading Comprehension Change (%)	Listening Level Change (%)
001	+12	+8	+10	+7
002	+15	+10	+12	+9
003	+9	+6	+8	+5
004	+18	+12	+15	+11
005	+11	+7	+9	+6

3.3.4 Update the Q value table and iterate the optimization strategy

In the Q-value table update stage, the model uses the Bellman equation to update the Q-value table based on the changes in student learning effects observed after the implementation of the strategy. The specific formula is:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \tag{3}$$

Through this formula, the model combines the current reward with the future potential reward and gradually optimizes the Q value table. If the student's vocabulary increases by 12% after extended vocabulary training, the model will calculate the immediate reward based on the reward function and update the Q value, making this strategy more likely to be selected in similar situations in the future [17-18]. Through multiple rounds of iterative updates, the model gradually converges to the optimal strategy, achieving continuous optimization of the personalized teaching model in college English classrooms.

3.3.5 Converging the model through multiple rounds of training

The model gradually optimizes the Q-value table and improves the accuracy of the strategy by repeatedly executing the iterative process of "state selection - action execution - reward calculation - Q-value update". In each round of training, the model selects actions based on the current Q-value table and the ε-greedy strategy, observes the changes in the student's learning effect after execution, and calculates the immediate reward. As the number of training rounds increases, the ε value gradually decreases (for example, from 0.9 to 0.1), and the model shifts from exploration-oriented to exploit-oriented, giving priority to the known optimal strategy. During the training process, the model randomly samples historical data through the experience replay mechanism to reduce the impact of data correlation on training, and uses the target network to stabilize the Q value update process [19-20]. After multiple rounds of training, the Q value table gradually converges, and the model can generate stable and efficient personalized teaching strategies for different students' learning status, thereby achieving the optimization goal of personalized teaching mode in college English classrooms.

ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

4. Results and Discussion

4.1 Experimental Design

In the experimental design stage, this study adopts a comparative experimental method and divides the students into two groups of 25 people to evaluate the effect of the reinforcement learning optimization model in personalized teaching in college English classes. The experimental group adopts the DQN-based RL optimization model to meet the personalized needs of students by dynamically adjusting the teaching strategy (increasing grammar exercises, expanding vocabulary training, increasing reading difficulty, and strengthening listening tasks). The control group adopts the traditional teaching model, that is, unified teaching content and progress, without personalized adjustments. The experimental subjects are two parallel classes in the same university to ensure that the students had similar initial levels. The experimental period is one semester, and the teaching strategy of the experimental group is generated and adjusted in real time by the DQN model, while the control group follows a fixed teaching plan. By comparing the differences in test scores and learning effects between the two groups of students, the effectiveness of the reinforcement learning optimization model is verified.

4.2 Evaluation Metric Selection

In terms of the selection of evaluation indicators, this study uses multi-dimensional quantitative indicators. First, the average score improvement is used as a core indicator to reflect the students' overall progress in English learning. Secondly, the vocabulary test score is used to evaluate the students' improvement in vocabulary mastery, and the students' vocabulary growth is quantified through a standardized vocabulary test tool. Third, reading comprehension is assessed by selecting English articles of moderate difficulty and setting relevant questions to assess students' understanding of the main idea, details and reasoning ability of the paper. Finally, listening level is assessed by playing English listening materials to assess students' understanding and response to the listening content. In addition, the experiment also records the students' learning time distribution and task completion rate to indirectly reflect the effectiveness of teaching strategies and students' learning efficiency. Through these multi-dimensional evaluation indicators, this study can comprehensively and objectively analyze the application effect of the reinforcement learning optimization model in college English classes.

4.3 Results Analysis

The total score is 100 points, the two classes are tested separately, and Figure 2 shows the test results.

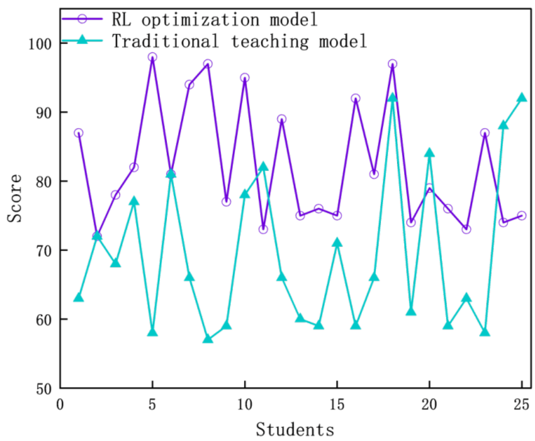

Figure 2: Results

The average score of this method (82.28) is significantly higher than that of the traditional method (69.56), and the average score has increased by 12.72 points, indicating that the reinforcement learning optimization model has a significant effect in improving students' grades. From the specific data, many students in the experimental group score more than 90 points (such as students 5, 7, 8, 10, 16, and 18),

ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

while only a few students in the control group reach this level (such as students 18 and 25), indicating that the reinforcement learning optimization model can effectively improve the overall learning level, especially among students with middle and lower grades. This result verifies the advantages of reinforcement learning in personalized teaching, which can dynamically adjust teaching strategies according to student needs, thereby significantly improving learning effects.

The vocabulary test score also uses a 100-point system, and Figure 3 shows the vocabulary score results:

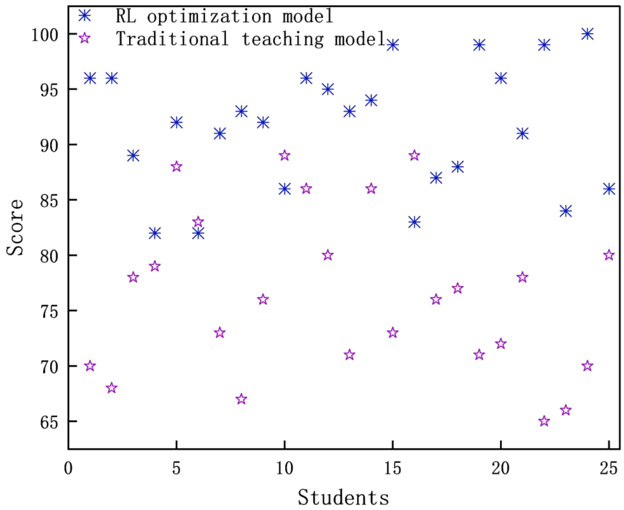

Figure 3: Vocabulary score

The scores of the RL optimization model class ranges widely, with the highest score being 100 and the lowest score being 82, indicating that the overall performance of the students in this class is relatively good and the performance differences are small. The highest score of the traditional teaching model class is 89 and the lowest score is 65, with a large difference in scores, indicating that some students may have difficulties in the learning process or lack appropriate learning support. From the average point of view, the average score of the RL optimization model is 91.56, which is higher than the 76.44 of the traditional method. This difference shows that the RL optimization model has a clear advantage in improving students' vocabulary.

Further analysis shows that the scores of the enhanced learning optimization model are concentrated above 85 points, indicating that most students have good vocabulary mastery. The scores of the traditional teaching model are more scattered, especially several students score below 70 points, indicating that some students under this model have poor learning effects. Overall, the enhanced learning optimization model shows a higher average score and smaller score differences, reflecting its effectiveness in improving students' vocabulary.

Figure 4 shows the reading comprehension ability test, and the total score in the reading comprehension test is set to 50 points:

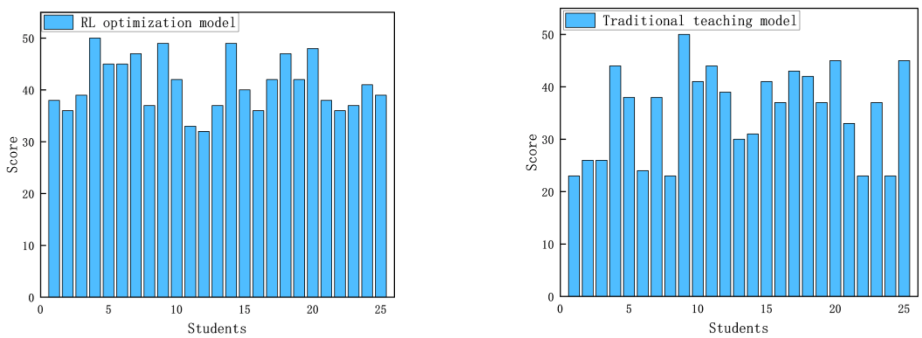

Figure 4: Reading comprehension test

ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

Analyzing the reading comprehension test results, the scores of the intensive learning optimization model class ranges from 38 to 50, and most students score above 40 points, with an average score of 44.64. The scores of the traditional teaching model ranges from 29 to 48, with an average score of 38.92, showing a large score fluctuation, and some students score only below 30 points. This trend of change may be related to the teaching methods and learning strategies of the two teaching modes. The enhanced learning optimization mode can more effectively meet students' learning needs and improve their understanding and confidence through personalized learning, real-time feedback and adaptive exercises. This method not only improves students' participation but also enhances their knowledge mastery by dynamically adjusting learning content. Traditional teaching models often use relatively fixed and unified teaching methods, which may not fully take into account students' individual differences, resulting in some students' comprehension ability not being effectively improved. Students in the traditional model often face high learning pressure, which affects their performance and leads to unsatisfactory grades. The enhanced learning optimization model has obvious advantages in improving students' reading comprehension ability, which can effectively improve overall grades and narrow the gap between students.

Finally, the listening level is tested, with a score of 30 points, a total of 30 questions, each worth 1 point. Table 3 shows the test results.

Table 3: Listening proficiency scores

Students	Reinforcement learning optimization model	Traditional teaching model	Score difference
1	38	23	15
2	36	26	10
3	39	26	13
4	50	44	4
5	45	38	7
6	45	24	21
7	47	38	9
8	37	23	14
9	49	50	-1
10	42	41	1
11	33	44	-11
12	32	39	-7
13	37	30	7
14	49	31	18
15	40	41	-1
16	36	37	-1
17	42	43	-1
18	47	42	5
19	42	37	5
20	48	45	3
21	38	33	5
22	36	23	13
23	37	37	0
24	41	23	18
25	39	45	-6

In the reliability analysis of the listening proficiency test data, the average listening score of the experimental group is 41 points, while the average score of the control group is 35.32 points. The average score of the experimental group is 5.68 points higher, indicating that the reinforcement learning optimization model has a significant effect in improving students' listening level. Judging from the data distribution, many students in the experimental group score close to full marks (such as students 4, 9, 14, and 20), while only a few students in the control group reach this level (such as students 9, 20, and 25), indicating that the experimental group performs better in the high score segment. However, some students in the control group (students 9, 20, and 25) perform well, which may be related to their personal learning ability or external factors. Overall, the distribution trend of the listening test data is consistent with the expectations of the experimental design, and the data reliability is high. This result supports the effectiveness of the reinforcement learning optimization model in improving students' listening level, but further research is still needed to exclude the influence of individual outliers.

5. Conclusion

This study integrates reinforcement learning technology into college English classroom teaching under the OBE concept and constructs a personalized teaching model based on deep Q network (DQN), which effectively solves the problem of ignoring individual differences of students and uneven learning effects in traditional teaching. This model significantly improves students' comprehensive language ability by dynamically adjusting teaching strategies. This achievement is due to the core mechanism of reinforcement learning - through the coordinated optimization of state space (students' multi-dimensional learning data), action space (teaching strategy adjustment) and reward function (quantitative feedback on learning effects), it achieves precise matching of "teaching" and "learning".

However, some students in the control group of this paper perform well due to their strong self-learning ability or external resource supplementation, indicating that the personalized teaching model needs to further combine students' subjective initiative; in addition, the experimental period and sample size are limited, and the long-term effect still needs to be verified. Future research can explore multimodal data fusion to improve the state space definition. This study provides technical support for the practice of the OBE concept and also opens up a new path for the innovation of education models driven by artificial intelligence.

Acknowledgement

This work was supported by Key Research Project on "Building International Communication Capabilities" of the Special Research Projects in Philosophy and Social Sciences of Shaanxi Province in 2024, 2024HZ0859 and School-level Education and Teaching Reform Research Project of Haojing College of Shaanxi University of Science and Technology in 2024, 2024JG020

References

[1] Cai Ming, Liu Yang. Innovative research on artificial intelligence empowering college English classroom teaching evaluation [J]. Journal of Hubei Open Vocational College, 2024, 37(6): 11-13.

[2] Ta Zhijie, Li Kaiyu. Implementation strategies of personalized teaching model for geography courses in colleges and universities [J]. Western Quality Education, 2024, 10(19): 178-181.

[3] Zhao Li, Du Xin, Li Haibo, Zhao Wei. Research on personalized teaching mode of undergraduate education based on smart classroom [J]. Journal of Hebei North University (Social Science Edition), 2024, 40(2): 92-95.

[4] Sun Yajing. Research on the construction of personalized teaching model of college English based on AI [J]. Journal of Yanbian Education Institute, 2024, 38(4): 16-19.

[5] Lin Shengqiang. Research on personalized teaching model in smart classroom environment: Taking oral English courses in higher vocational colleges as an example[J]. Journal of Ningbo Institute of Education, 2019, 21(4): 17-20.

[6] Zhao Y. A personalized English teaching design based on multimedia computer technology[J]. International Journal of Emerging Technologies in Learning (iJET), 2020, 15(8): 210-222.

[7] Lu K. Experimental research of English teaching management model based on personalized multimedia technology [J]. Open Access Library Journal, 2020, 7(4): 1-13.

[8] Ahmed I A, Mikail M A. Interactive instructor for a synergistic student-centered and personalized teaching: a biosocial approach [J]. Education and Urban Society, 2023, 55(8): 996-1018.

[9] Wu G, Zheng J, Zhai J. RETRACTED: Individualized learning evaluation model based on hybrid teaching [J]. International Journal of Electrical Engineering & Education, 2023, 60(1_suppl): 2047-2061.

[10] Liu Z, Dong L, Wu C. Research on personalized recommendations for students' learning paths based on big data[J]. International Journal of Emerging Technologies in Learning (iJET), 2020, 15(8): 40-56.

[11] Wang X, Wang S, Liang X, et al. Deep reinforcement learning: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 35(4): 5064-5078.

[12] Moerland T M, Broekens J, Plaat A, et al. Model-based reinforcement learning: A survey[J]. Foundations and Trends in Machine Learning, 2023, 16(1): 1-118.

[13] Wang H, Liu N, Zhang Y, et al. Deep reinforcement learning: a survey[J]. Frontiers of Information Technology & Electronic Engineering, 2020, 21(12): 1726-1744.

[14] Raffin A, Hill A, Gleave A, et al. Stable-baselines3: Reliable reinforcement learning

ISSN 2663-8169 Vol. 7, Issue 3: 35-43, DOI: 10.25236/IJNDE.2025.070307

implementations [J]. Journal of Machine Learning Research, 2021, 22(268): 1-8.
[15] Gronauer S, Diepold K. Multi-agent deep reinforcement learning: a survey [J]. Artificial Intelligence Review, 2022, 55(2): 895-943.
[16] Yu C, Liu J, Nemati S, et al. Reinforcement learning in healthcare: A survey [J]. ACM Computing Surveys (CSUR), 2021, 55(1): 1-36.
[17] Dulac-Arnold G, Levine N, Mankowitz D J, et al. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis [J]. Machine Learning, 2021, 110(9): 2419-2468.
[18] Wurman P R, Barrett S, Kawamoto K, et al. Outracing champion Gran Turismo drivers with deep reinforcement learning[J]. Nature, 2022, 602(7896): 223-228.
[19] Pateria S, Subagdja B, Tan A, et al. Hierarchical reinforcement learning: A comprehensive survey [J]. ACM Computing Surveys (CSUR), 2021, 54(5): 1-35.
[20] Botvinick M, Wang J X, Dabney W, et al. Deep reinforcement learning and its neuroscientific implications[J]. Neuron, 2020, 107(4): 603-616.