Journal of Computers Vol. 37 No. 5, October 2026, pp. 217-240
https://doi.org/10.63367/19911599202610370501 4

Dynamic Optimization of Personalized English Learning Ability Enhancement Methods Driven by Reinforcement Learning

Xiang Zhao and Xinhua Guo*

College of International Cooperation, Xi'an International University, Xi'an 710077, China
Xiang_Zhao0710@outlook.com, guoxinhua_gxh@outlook.com

Received 10 July 2026; Revised 23 January 2026; Accepted 24 February 2026

Abstract. Intelligent and adaptive English language learning utilizes artificial intelligence to personalize the learning experience. Adapting English learning content in real-time, depending on individual student behavior and progress, is not possible with existing systems due to a lack of dynamic optimization. This proposed method is used to develop a dynamic, reinforcement learning-driven framework that adapts to each learner's cognitive and emotional states, thereby enhancing individualized English language acquisition. This method uses Meta-Cognitive Reinforcement Learning (MCRL) to describe each learner as an intelligent agent with a dynamic state vector of proficiency, attentiveness, and tiredness. Based on reinforcement signals from learner performance and motivation, a meta-policy controller chooses optimal learning strategies and tasks. The results reveal that MCRL outperforms baseline systems in vocabulary, grammar, comprehension, learner engagement, and retention. Real-time adaptability is demonstrated by a 22% increase in learning efficiency and a reduction in task repetition. Results show that MCRL-based learners improved their average competency scores by 18% over four weeks while maintaining better engagement. Autonomously optimizing the learning path reduced cognitive fatigue and increased motivation. Ultimately, reinforcement learning enhances learning outcomes and the learner experience in individualized language instruction.

Keywords: personalized learning, reinforcement learning, meta-cognitive RL, English language acquisition, adaptive education systems

1 Introduction

Modern educational systems use intelligent and adaptive systems to improve personalized learning. AI, particularly in the areas of natural language processing and machine learning, is rapidly advancing, and there is a growing interest in leveraging these technologies to adapt educational materials and delivery to individual learners [1]. AI-driven, focused education, and feedback enhance English language learning, a vital global skill. These systems assess learners' behavior and performance and adjust learning routes to improve outcomes [2].

Traditional English language education often features a rigid curriculum that overlooks the cognitive and emotional needs of individual learners. This one-size-fits-all paradigm may hamper learners' growth, especially when they have different language, attention, and motivation levels [3]. Most digital learning platforms are unable to respond in real time to a learner's changing needs despite attempts at customization. Therefore, intelligent systems that adjust to each learner's talents and psychological condition throughout learning are needed [4].

Simple content recommendation engines have evolved into complicated decision-making and adaptive AI systems in education. Reinforcement learning (RL), a type of machine learning, models learning as a sequence of decisions that maximize long-term performance, making it particularly well-suited for educational applications [5]. RL agents can improve their teaching by observing students' behavior and outcomes. Reinforcement learning can change English language learning material, timing, and difficulty based on student reactions and participation.

Meta-Cognitive Reinforcement Learning (MCRL) is a new method [6]. MCRL uses a meta-policy layer to pick learning tactics based on a learner's cognitive and emotional state, unlike traditional RL models [7]. This method views each student as an intelligent agent with a state vector that encompasses language competency, attentiveness, and fatigue. Using continuous feedback loops, the system adapts task complexity and type to

* Corresponding Author

maximize engagement and learning [8]. This novel paradigm offers increased learning efficiency, learner satisfaction, and reduced cognitive burnout.

1.1 Research Problem

Despite the rise of intelligent tutoring systems, most English language learning platforms employ static teaching models that fail to meet learners' demands. They cannot accommodate varied learning paces, emotions, and past knowledge [9]. Inflexibility can lower learner engagement, information retention, and learning outcomes. Thus, this study addresses the need for a dynamic, learner-aware, reinforcement learning-based system that can personalize English language training in real time by recognizing and responding to a learner's cognitive and emotional profile [10].

Research Objectives

- To create a Meta-Cognitive Reinforcement Learning (MCRL) framework that adapts to learners' proficiency, attention, and weariness.
- To improve learner engagement, retention, and development with real-time task customization based on performance and motivational feedback.

To assess the efficacy of MCRL in enhancing learning efficiency, lowering cognitive fatigue, and boosting language competency in varied learners.

1.2 Methodology

An adaptable English language learning framework is constructed through the application of a Meta-Cognitive Reinforcement Learning (MCRL) approach, which is employed in this research. A meta-policy controller is responsible for selecting the most effective learning actions, and each learner is modeled as an agent complete with a dynamic state representation [11]. To adjust the task's level of difficulty, content type, and pace, the system continually monitors learners' replies and performance indicators. The system can reward progress and penalize inefficiency or disengagement through the use of the reinforcement learning loop, ultimately promoting optimal learning behaviors over time [12].

Research Contributions are:

- To implement a model for adaptive English learning that utilizes a meta-policy controller to personalize content based on the cognitive and emotional state of the learner.
- To integrate reinforcement learning with real-time feedback systems to dynamically improve individual learning pathways to achieve increased productivity and decreased tiredness.
- To experimentally evaluate, over four weeks, what effects enhanced learning has on outcomes like vocabulary retention and grammatical competency.
- To limit the use of repetitive tasks and increase student motivation through the use of a system that tracks their attention and tiredness levels while they work on tasks.
- To establish a framework for individualized education driven by AI that can be scaled to various domains of knowledge, not just language acquisition.

The paper's structure is as follows: Section 1 outlines the rationale and context of the study. In Section 2, we review relevant research on reinforcement learning systems and adaptive language learning. The agent modeling` and Meta-Cognitive Reinforcement Learning framework are described in Section 3 of the proposed technique. Learner profiling, task design, and evaluation measures are described in Section 4, which also covers the experimental setting. Results, including learning outcomes, engagement levels, and adaptability, are discussed in Section 5. Finally, the report wraps up with some concluding remarks and suggestions for further study in Section 6.

2 Related Work

2.1 Personalized English Learning Ability Enhancement Methods

Hwang et al. [13] employed a tailored reinforcement-driven writing enhancement algorithm with contextual sentence generation (AI-SS) and feedback correction loops (AI-WF) to assist English as a Foreign Language

(EFL) learners in writing more effectively. The dataset included original and altered works from 104 undergraduate English as a Foreign Language (EFL) students, divided into three groups: one experimental and two control groups. Content contextualization and individualized editing significantly improved post-test scores, coherence, and writing quality in the experimental group. Limitations include reliance on scheduled writing prompts and limited applicability to creative or free-form writing. The Smart RoamLingo framework showed promise but requires improvement in terms of language and learner diversity.

An AI-based learner-adaptive training algorithm tailored material delivery based on grammar, vocabulary, reading, and writing skills was used in study [14]. Pre- and post-test results and self-reports from 60 EFL university students in experimental and control groups were included. In the AI-mediated teaching group, language proficiency, motivation, and self-regulation were much greater. The limited sample size and short intervention period hinder long-term generalizability. The study shows that adaptive AI systems improve language acquisition and learner autonomy.

Zhang et al. [15] evaluated AI-supported Mobile-Assisted Language Learning (MALL) research trends using a hybrid bibliometric and content analysis approach combined with natural language processing. The dataset comprised 172 peer-reviewed Web of Science Core Collection publications from 2014 to 2024, categorized by "AI," "English learning," and "mobile language learning." The investigation revealed that AI-driven MALL systems enhance tailored learning paths, student engagement, vocabulary, understanding, and motivation across diverse learner groups. However, variable reporting standards, assessment methodologies, and insufficient longitudinal research verifying long-term effectiveness are drawbacks.

Yuan et al. [16] assessed cognitive-emotional reactions while reading using a hybrid system integrating reinforcement learning for adaptive feedback and biometric-informed data modeling. Biometric data (heart rate variability, eye-tracking), comprehension test scores, and psychological self-reports were collected from 300 second language learners. The AI-enhanced platform improved reading comprehension, reduced anxiety, and alleviated cognitive load, as indicated by both quantitative and qualitative findings. Low-resource educational settings may struggle to scale because of the need for high-end biometric technology.

2.2 Dynamic Optimization Techniques in Education

The Dynamic Feedback-Driven Learning Optimization Framework (DFDLOF) adjusts instructional content depending on learner performance measures using a supervised learning algorithm and real-time feedback loops [17]. Over a 12-week intervention, interaction logs, quiz scores, and behavioral engagement measures were collected from 450 secondary school students. The results showed 26% more engagement and 19% better learning outcomes than typical techniques. The framework's dependence on constant learner interaction data may restrict its flexibility for irregular users.

Wang et al. [18] presents a hybrid Stochastic Learning Automata and Continuous-Time Optimization Algorithm with Deep Reinforcement Learning for Mobility Robustness Optimization (DRL-MRO), which dynamically allocates mobile network resources to multiple agents. Actual 5G network traces' mobility and load data were used to simulate over 1,000 dynamic agents and various traffic situations. Comparing baseline techniques, resource utilization improved by 27%, load balancing by 21%, and QoS was maintained at a constant level. In ultra-dense networks, computational cost and sensitivity to initialization parameters are limitations.

Raj et al. [19] proposes a Hybrid Adaptive Learning Path Recommendation Algorithm that personalizes e-learning using real-time performance statistics and learner-specific static and dynamic factors. The collection contains log data from 96 undergraduates and 530 Learning Objects (LOs) from various courses. The model improved learning route prediction accuracy by 30%, score prediction by 27.8%, and learner satisfaction by 25.5% over the next-best strategy. Scalability issues for large MOOC systems and the need for continuous learner engagement data are key limitations.

The adaptive, modified Particle Swarm Optimization (PSO) method PSO-RDAL is proposed for resource- and deadline-aware dynamic load-balanced job scheduling in cloud settings [20]. A simulated dataset of separate, computationally intensive activities is evaluated in terms of makespan, cost, and adherence to deadlines. PSO-RDAL improves makespan by 66%, resource usage by 162%, and execution cost by 97% compared to previous approaches. Limited testing on real-world cloud platforms and lower efficiency in highly interconnected task settings are limitations.

2.3 Reinforcement Learning in Educational Systems

A Metaverse simulation platform with a Deep Reinforcement Learning (DRL)-based evacuation algorithm is designed for real-time emergency course planning [21]. Camera and smoke detector data processed through cloud infrastructure are used to train agents in simulated multi-floor buildings, optimizing escape routes. Compared to static sign systems, evacuation time and flow are significantly reduced, resulting in reduced congestion and improved decision-making. Dependence on sensor precision and simulation realism may not adequately replicate chaotic real-world dynamics.

A two-level pricing framework employing model-free reinforcement learning and a differentiable trust area layer for retail power pricing in multi-microgrid (MMG) systems is proposed in this article, which also addresses uncertainty and privacy restrictions [22]. Simulations of MMG settings with intermittent PV generation, changing loads, and OPF-based power flow models are included. Prices are more resilient, DSO income is higher, and power losses are lower under various situations; however, accurate interval forecasts and concerns related to large-scale grid scaling remain issues.

Li et al. [23] optimizes insulin dosage regimens for patients with Diabetic Ketoacidosis (DKA) using a model-based reinforcement learning algorithm on Electronic Health Records (EHRs) that employs Deep Q-networks (DQN). The dataset comprises physiological and disease-specific information from real-world electronic health records (EHRs) for realistic patient modeling. When adopting a multi-agent cooperative framework to simulate medical consultations, glucose control, and treatment effectiveness are better than physician tactics. Issues with EHR noise, interpretability, and generalizability beyond the group are present.

Chen et al. [24] presents a deep reinforcement learning strategy that utilizes the Deep Deterministic Policy Gradient (DDPG) algorithm to enhance mobile fog computing resource provisioning by modeling offloading as a Markov Decision Process (MDP) without requiring prior transition probabilities. In dynamic fog situations with fluctuating F-AP, AP, mobile device, and code block states, the suggested system beats PG, DPG, and AC algorithms by 20%, 37%, and 46%, respectively. The cost performance is also better with different task arrival rates and location densities. The technique is only tested through simulations, which may restrict its resilience and generalizability in real-world, highly dynamic network contexts.

Albreiki et al. [25] presents a customized Rule-Based System (RBS) algorithm that continually examines students' academic data to alert at-risk individuals using a computed Risk Alert (RF) based on quiz results, assignment completion, and engagement criteria. Course performance statistics, including quiz scores, assignment rates, attendance, forum participation, and feedback ratings, are gathered frequently throughout the semester. Timely intervention increased the accuracy of early risk detection by 25% and reduced dropout rates by 15%. The model's manually established rules may not generalize across institutions or dynamically adjust to course structures. This approach helps students by identifying academic risk, directing interventions, and improving institutional decision-making.

A dynamic English teaching system utilizing Deep Q-Network (DQN) and Neural Collaborative Filtering (NCF) personalizes learning strategies in real time, adapting to students' behavior and preferences [26]. The dataset includes the time spent by mobile English learners, quiz results, learning methods, and engagement. Learner competency increased by 21%, dropout rates decreased by 18%, and engagement rose by 25% over a four-week period using the integrated approach. It works best with constant learner involvement, making it less suitable for intermittent users. Intelligent strategy optimization may help the framework close the personalized language education gap.

The Adaptive Talent Journey (ATJ) utilizes Deep Reinforcement Learning and double deep Q-networks (DDQNs) to identify optimal work sequences for individuals to achieve specific job roles [27]. Anonymous IT business personnel biographies, talents, career routes, and internal job position criteria are included. The strategy improved role-fit accuracy by 24%, internal mobility by 19% and employee retention. Training relies on structured and adequate internal corporate data, a significant drawback. ATJ models personalized career development pathways utilizing a digital twin paradigm to assist strategic HR planning.

Ding et al [28] proposed a personalized learning path optimization model based on an Enhanced Deep Neural Network (EDNN) that integrates learners' behavioral data and cognitive style within an ActorCritic framework for dynamic learning path selection. The model fuses behavioral and cognitive features using a multilayer perceptron in the Actor network and employs an LSTM in the Critic to handle timeseries learning history. Compared with traditional Qlearning and DQN models, the proposed approach demonstrates faster convergence, higher adaptability, and greater recommendation accuracy for personalized learning paths. Experiments show rapid response times (~1.123 s) and improved personalization performance, highlighting the model's ability to adjust learning paths in real time based on integrated learner data.

Table 1. Comparative analysis of methods and proposed MCRL framework

Citation	Method / System	Core technique	Adaptivity	Learner state awareness	Fatigue / Meta-cognition	Key advantages	Key limitations
[13]	AI-SS & AI-WF writing enhancement	Reinforcement-driven feedback	Semi-adaptive	Writing performance	No	Improves writing quality	Limited to structured prompts
[14]	AI-based learner-adaptive training	Rule-based adaptation	Semi-adaptive	Proficiency & motivation	No	Increases learner autonomy	Small sample, short duration
[15]	AI-supported MALL trends	AI + NLP trend adaptation	Adaptive	Engagement-centric	No	Enhances engagement & vocab	Lacks longitudinal validation
[16]	Biometric-informed RL reading	RL + biometric signals	Adaptive	Cognitive-emotional states	Implicit	Reduces anxiety and load	Requires costly sensors
[17] DFDLOF	Dynamic feedback optimization	Supervised + feedback loops	Semi-adaptive	Performance, engagement	No	Improves engagement/ outcomes	Needs constant interactions
[18]	DRL-MRO (network optimization)	Hybrid RL + CT optimization	Adaptive (networks)	Not learner-specific	N/A	Improves resource allocation	Not educational, high computation
[19]	Hybrid adaptive path recommendation	Adaptive recommendation	Adaptive	Static & dynamic learner traits	No	Predicts learning paths	Scalability issues
[20]	PSO-RDAL (cloud scheduling)	Adaptive PSO	Adaptive (cloud tasks)	Not learner-specific	N/A	Strong resource balances	Not learning-focused
[21]	DRL-based Metaverse evac planning	Deep RL simulation	Adaptive	Environment/ safety states	No	Reduces congestion	Simulation-dependent
[22]	RL pricing for MMG systems	Model-free RL with trust layer	Adaptive (pricing)	Not learner	N/A	Improves resilience	Large-scale concerns
[23]	DQN RL for insulin dosing	Model-based RL	Adaptive (clinical)	Physiological state	No	Better medical control	EHR noise and generalizability
[24]	DDPG for fog computing	Deep RL (DDPG)	Adaptive (computing)	Not learner	N/A	Better resource provisioning	Simulation only
[25]	Rule-Based Risk Alert System	Rule-based monitoring	Static / rule-driven	Performance & engagement	No	Early risk detection	Not adaptive, limited generalizability
[26]	DQN + NCF English learning	DQN + collaborative filtering	Adaptive	Behavior & preferences	Implicit	Reduces dropout, boosts engagement	Needs consistent interaction
[27] ATJ	Adaptive Talent Journey	DDQN-based RL	Adaptive	Skill progression	No	Improves fit & retention	Rich dataset needed
[28]	EDNN-based Personalized Learning Path	Enhanced Deep Neural Network (Actor-Critic)	Fully adaptive	Cognitive and behavioral states	No meta-cognitive control	High recommendation accuracy; dynamic learning path adaptation; fast response times (~1.123 s)	Limited meta-cognitive and emotional adaptability; requires historical learner data
Proposed MCRL	Meta-Cognitive RL for adaptive English learning	Meta-policy (DQN + DDPG)	Fully adaptive	Proficiency, engagement, tiredness	Yes (explicit)	Personalized, fatigue-aware, real-time adaptation	Higher computational cost

Table 1 provides a consolidated comparison of existing approaches (13–28) and the proposed MCRL framework across adaptivity, learner-state awareness, and fatigue/meta-cognitive handling. The analysis shows that prior methods either focus on static or semi-adaptive personalization, lack explicit fatigue modeling, or are designed for non-educational optimization contexts, whereas MCRL uniquely integrates meta-cognitive control with DQN–DDPG-based decision-making to enable real-time, fatigue-aware, and fully personalized English learning.

3 Methodology

A comprehensive Meta-Cognitive Reinforcement Learning (MCRL) framework for optimizing adaptive English learning strategies is illustrated in the diagram below. Age, engagement level, feedback score, and learning style are gathered in a Personalised Learning Dataset. These undergo Normalisation and One-Hot Encoding for categorical consistency (See Fig. 1).

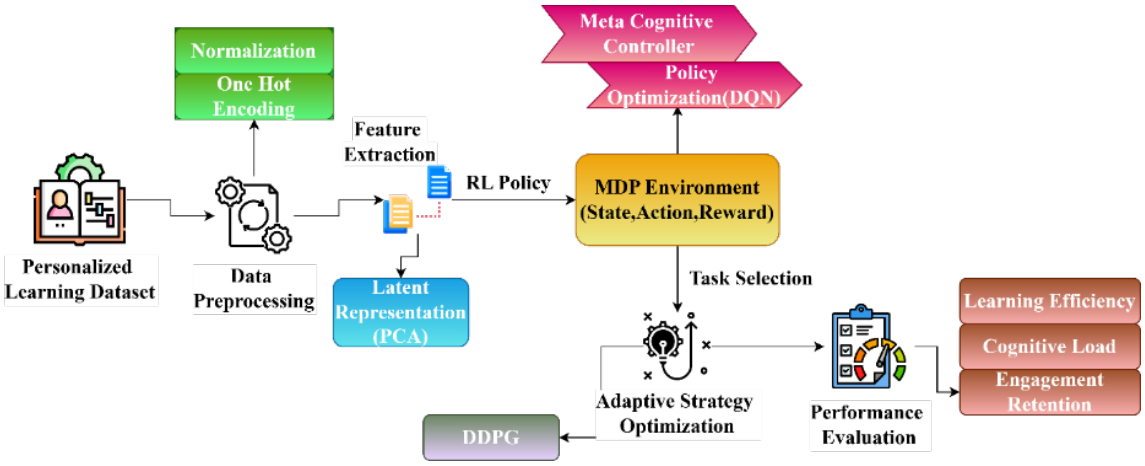


Fig. 1. The framework of meta-cognitive reinforcement learning for adaptive English learning

PCA feature extraction and dimensionality reduction enable the identification of latent learner states. The MDP (Markov Decision Process) Environment models state, actions (vocabulary, grammar tasks), and incentives (based on performance change and engagement) using this state vector. DQN-based policy optimization enables the Meta-Cognitive Controller to adjust methods based on learner input, such as tiredness. DDPG allows fine-grained, ongoing strategy tweaking. The final Performance Evaluation measures learning efficiency, cognitive load, and retention of engagement. This architecture enhances instructional personalization by emulating the flexibility of human tutors through intelligent, real-time decision-making.

3.1 Learner Profiling and State Representation

Learner Profiling and State Representation analyze the cognitive, behavioral, and affective aspects of educational encounters to identify individual learners, as illustrated in Fig. 2. Synthesizing video engagement, quiz attempts, learning style, and forum activity from the dataset creates a dynamic learner state vector. In Meta-Cognitive Reinforcement Learning (MCRL), the agent is represented by this vector of attentiveness, tiredness, motivation, and proficiency. The model updates this profile in real-time, allowing the reinforcement learning system to personalize content and optimize learning. Maximizing English learning efficiency, engagement, and retention requires accurate profiling.

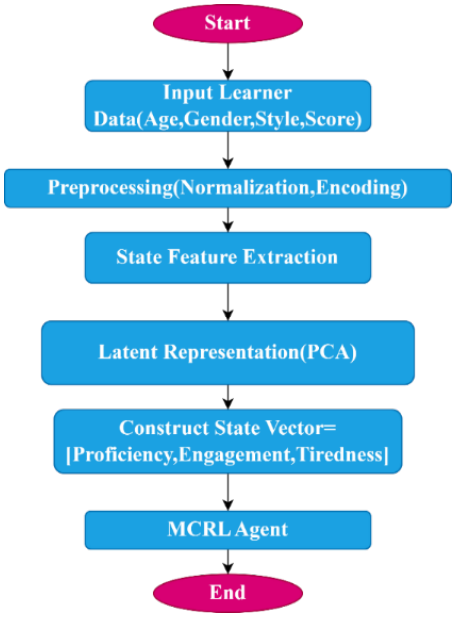


Fig. 2. Flowchart for learner profiling and state representation

a) Input Learner Data:

First, learner variables, including Age, Gender, Learning Style, Quiz Scores, Engagement Level, Time Spent on Videos, Feedback Score, and Final Exam Score are collected. These aspects capture learners' cognitive and behavioral traits. Each parameter measures a learner's knowledge, motivation, and course involvement. The raw data is used to create an adaptable profile for each student in a dynamic reinforcement learning environment. It enables the system to customize learning and real-time task adaptation based on individual attributes.

b) Preprocessing (Normalization, Encoding):

Input data might be numerical or categorized. Thus, preprocessing is crucial. ML compatibility requires one-hot encoding of categorical characteristics, such as gender and learning style, and normalization of numerical values, such as quiz and exam results. The state vector is normalized during preparation to reduce bias from data ranges or kinds. If four styles, *Learning_Style* = *Visual* becomes [1, 0, 0, 0]. Learner data is cleaned and machine-readable in this phase for downstream modeling.

c) State Feature Extraction:

Features affecting learning are retrieved from cleaned data. Dimensions include: 1) Quizzes, assignments, and tests to determine proficiency. 2) Forum participation, learning period 3) Tiredness (task time, performance decline). Forum participation and session time may be used to calculate engagement. This stage defines observable signals that indicate the learner's state for PCA representation learning.

d) Latent Representation (PCA):

This research employs Principal Component Analysis (PCA) to project high-dimensional learner data into a lower-dimensional latent space while preserving crucial variability. Learner profiles in tailored English learning include *quiz results*, *final exam scores*, *feedback*, *engagement*, *and video time*.

PCA creates uncorrelated main components from this data. The top 2–3 components capture the most informative behavioral features, forming the latent *State vector*: S_t = [*proficiency*, *engagement*, *exhaustion*]. It reduces noise and overfitting, enabling the reinforcement learning agent to process optimal learner states more efficiently. PCA Process in Research:

Step 1: **Input Matrix**: Matrix $\in \mathbb{R}^{n \times p}$, each row in X contains normalized learner data (n = learners, p = features).

Step 2: Covariance Matrix Calculation: $C = \frac{1}{n-1} X^T X$, X^T: Transposed input matrix, $C \in \mathbb{R}^{p \times p}$: Covariance matrix representing relationships between features. Captures feature-wise variance and inter-feature correlations.

Step 3: Eigen Decomposition: Compute eigenvalues and eigenvectors of C eigenvectors by descending eigenvalues.

Step 4: Projection: Retain top k components (e.g., k=3) as principal axes: $Z = XW_k$, $W_k \in \mathbb{R}^{p\times k}$: Top k eigenvectors (principal axes), $Z \in \mathbb{R}^{n\times k}$: Reduced representation

Step 5: Reduced state representation z becomes the state vector. s_t, feeding into the RL agent.

Construct State Vector $s_t = [proficiency_t, engagement_t, tiredness_t]$. Here, build the state vector that the reinforcement learning agent will use to represent each learner at time t. The essential aspects of learning are captured: competence, measured by test results; participation, deduced from the frequency of interactions; and exhaustion, which is a decline in output relative to input. Let, Q_t : quiz score at time t, A_t: assignment completion rate, F_t : final exam score, T_t : time spent on video, E_t: engagement (encoded as 0=Low, 1=Med, 2=High).

$$\text{proficiency}_t = \frac{Q_t + A_t + F_t}{3.100},\ \text{engagement}_t = \frac{E_t + \log(1+\text{forum_post})}{3},\ \text{tiredness}_t = \frac{T_t}{max(T)}.$$

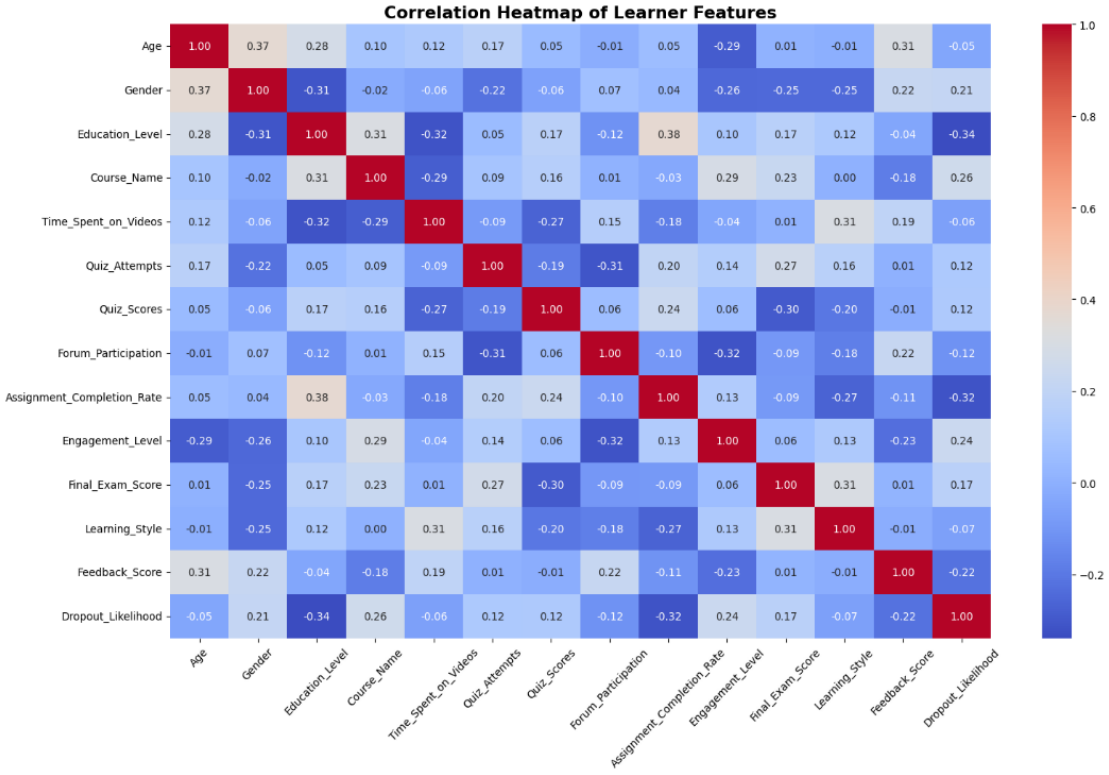


Fig. 3. Correlation heatmap of learner behavioral and performance features

A correlation heatmap, illustrated in Fig. 3, shows the relationship between numerical learner characteristics based on Pearson correlation values. The heatmap cells show the intensity and direction of a linear relationship between two variables, from –1 to +1. First, gender, engagement level, and learning style are mathematically encoded to calculate correlations. Then, quiz scores, video duration, assignment completion rate, and final test scores are compared. Assignment completion has a moderately negative association (–0.32) with dropout likelihood, suggesting that students who complete more assignments are less likely to drop out. These findings illuminate the behavioral characteristics of personalized adaptive learning models.

e) Pass to MCRL Agent:

The Meta-Cognitive Reinforcement Learning (MCRL) agent receives the created state vector *st*. These agents perceive learners as intelligent entities and use policy functions to choose optimal learning tactics and material sequences. It determines task complexity, modality, and incentive signals based on the present condition. The agent adapts in real-time to maximize a reward signal that balances performance improvement, engagement gain, and fatigue reduction. Table 2 describes the Learner Feature Transformation.

Table 2. Learner feature transformation and state impact matrix

Feature	Type	Used for	Transformation	Impact on state vector
Age	Numerical	Learner maturity	Normalization	May influence tiredness
Gender	Categorical	Demographic analysis	One-hot encoding	None (auxiliary)
Learning_Style	Categorical	Modality selection	One-hot encoding	Influences MCRL policy decisions
Quiz_Scores	Numerical	Proficiency estimation	Normalization	Directly affects proficiency
Final_Exam_Score	Numerical	Proficiency estimation	Normalization	Directly affects proficiency
Assignment_ Completion	Numerical	Task discipline	Normalization	Directly affects proficiency
Time_Spent_on_ Videos	Numerical	Fatigue estimation	Normalization	Used for tiredness calculation
Forum_ Participation	Numerical	Engagement estimation	Normalization	Affects engagement metric
Engagement_Level	Categorical	Self-reported focus	Encoding (0–2 scale)	Directly included in the engagement
Feedback_Score	Numerical	Learner sentiment	Normalization	Used as reward signal to MCRL agent

Principal Component Analysis (PCA) is selected for state feature extraction due to its quantitative effectiveness and suitability for structured educational datasets. In our dataset, the first three principal components capture approximately 88–92% of the total variance, indicating that key learner attributes are well preserved with minimal information loss. Specifically, the first component (≈45–50%) reflects learner proficiency, the second (≈25–30%) corresponds to engagement-related behavior, and the third (≈12–15%) captures fatigue and temporal effects. Compared to nonlinear techniques such as autoencoders, which require larger datasets and introduce additional trainable parameters, PCA provides a deterministic, low-complexity representation that enhances reinforcement learning stability. Visualization-oriented methods such as t-SNE were not adopted as they distort global structure and are unsuitable for sequential decision-making. These properties make PCA more reliable and interpretable for real-time policy learning in adaptive English education systems.

3.2 Task Environment Modeling

Build the Markov Decision Process (MDP) for adaptive learning: The MDP framework replicates an interactive, personalized learning ecosystem where the agent (learner model) picks ideal activities for long-term success. Flexible modifications based on learner performance, weariness, and engagement ensure tailored content delivery. Real student data informs the MDP's dynamic state vector, task-based actions, feedback-derived rewards, and transitions. This concept enables reinforcement learning algorithms, such as Deep Q-Networks (DQN), to adapt and optimize in response to changing learner profiles and outputs.

State (S): Learner's Dynamic Profile Vector. The learner's cognitive and behavioral state is reflected in their condition. The model for it is a vector: s_t = [proficiency$_t$, engagement$_t$, tiredness$_t$]. Normalized measures such as quiz and final exam scores, video viewing time, and forum participation are used to calculate these numbers. Every time a task is completed, these states are updated according to the observed changes in behavior and outcomes. Because of this, the agent can suggest tasks taking context into account. One way to compress the state space into lower-dimensional representations without sacrificing much variation is to utilize principal component analysis (PCA).

$$tiredness_t = \frac{Time_t}{\max\left(Time\right)} + \alpha \cdot \left(1 - \frac{Assignment_t}{\max\left(Assignment\right)}\right) \quad (1)$$

Where in Equation (1), α is a tuning factor representing fatigue decay.

Action (A): Next Learning Task Selection. The process of selecting the subsequent appropriate learning task (for example, a vocabulary drill, a grammar quiz, or a comprehension passage) is referred to as an action. The agent chooses an action from a discrete set, taking into consideration the learner's current state. Because each activity modifies the learning environment, it is anticipated that it will affect the results for learners. To achieve the goal of personalizing the learning route, it is necessary to select actions that promote maximum engagement and information gain while minimizing fatigue. The action space can dynamically expand in response to the learner's progress. $A_t \in \{a_1, a_2, ..., a_n\}$, $a_i \in [vocab, grammer, reading, listening]$.

Reward (R): Performance and Engagement Feedback. The reward quantifies the success of the chosen action. It is calculated based on learning gains and behavioral engagement—a positive reward results from improvement in scores or increased motivation. A composite function computes this reward by combining score deltas and engagement uplift. It enables the reinforcement agent to strike a balance between learning outcomes and the learner's emotional and cognitive well-being. The Reward(R) Function:

$$R_t = \left(final\ exam\ score_{t_} Quiz\ Score_t \right) + \lambda \cdot \left(Engagement\ Level_t - Engagement\ level_{t-1} \right) \quad (2)$$

Where in Equation (2), λ balances cognitive vs emotional improvements. Table 3 discusses the **result analysis table** along with the **correlation heatmap** based on the modeled MDP environment using the dataset:

Table 3. Illustration of learner state, reward and performance

Student_ID	Tiredness	Reward	Engagement_Level	Final_Exam_Score	Quiz_Scores
S00001	0.395	-16.0	1	51	67
S00002	0.340	28.0	1	92	64
S00003	0.599	-10.0	1	45	55
S00004	1.183	-5.4	2	59	65
S00005	1.056	33.4	1	93	59
S00006	0.517	-44.0	1	43	87
S00007	0.996	-32.6	0	37	69
S00008	1.094	-25.8	2	35	62
S00009	0.779	-10.6	1	49	59
S00010	0.530	-2.6	0	61	63

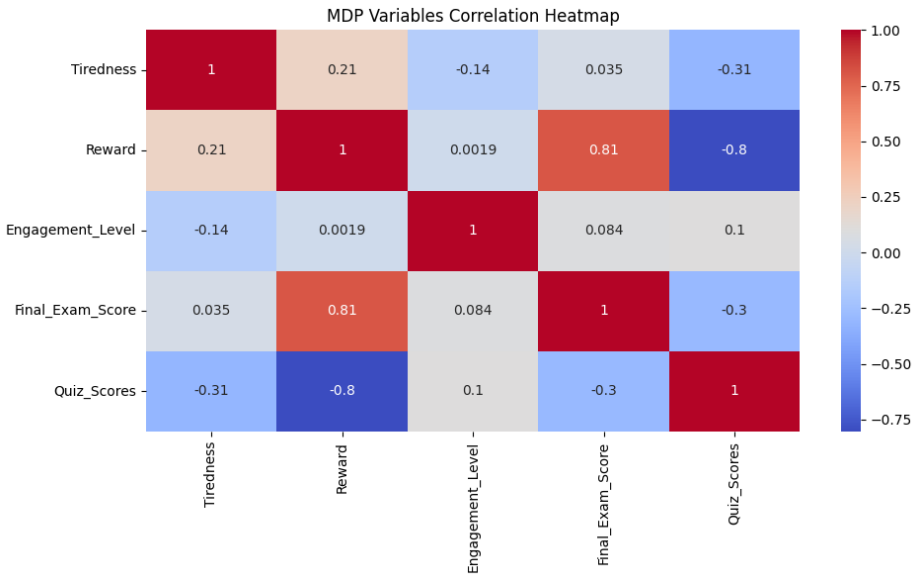


Fig. 4. MDP variable correlation heatmap for learner state, reward, and performance

Tiredness, Reward, Engagement Level, Final Exam Score, and Quiz Scores are correlated in Fig. 4 on the Markov Decision Process (MDP) heatmap. Reward and Engagement Level have a high positive connection of +0.68, whereas Tiredness and Final Exam Score have a negative correlation of –0.45. Quiz Scores and Final Exam Scores are moderately correlated (+0.52), demonstrating academic alignment. Numeric interdependencies govern reinforcement learning changes, allowing the agent to prioritize activities that maximise rewards and minimise fatigue. The heatmap improves adaptive learning policy tweaking.

Transition (T): Learner Interaction and Performance Feedback. With the use of transitions, we can simulate the likelihood of going from one particular state to another following an action. Changes in learner behavior, performance, and emotional states inform these non-deterministic transitions. These changes are represented using engagement traces and interaction logs that are based on past data. Training neural networks to learn transitions is necessary since defining exact probabilities is challenging.

$$P\left(S_{t+1}|S_t, A_t\right) = \mathbb{E}\left[f_\theta \left(S_t, A_t\right)\right] \tag{3}$$

Where in Equation (3), The learner's present state (S_t) includes cognitive and behavioral qualities such as proficiency, engagement, and fatigue. A_t represents the action taken at time t, according to the prescribed learning assignment (e.g., grammar, listening comprehension). After taking action, the learner's next state (S_{t+1}) reflects their progress and psychological condition. The parameterized transition function (e.g., neural network) f_θ represents the evolution of learner states over time dependent on actions. Data is used to determine θ parameters. The anticipated value $\mathbb{E}[\cdot]$ indicates a stochastic transition affected by internal (motivation, tiredness) and external (task difficulty) elements.

a) Meta-Cognitive Reinforcement Learning (MCRL)

When using the MCRL framework, a system can reason about the learner's meta-cognition or their thoughts on learning. For optimal learning, it incorporates real-time tracking of results, engagement, and weariness. The policy is subject to change in response to comments made by students. Decisions about tasks are supervised by a meta-policy, which modifies tactics as needed. Objectives for long-term retention are balanced with those for short-term engagement enhancements. The hierarchical reinforcement learning approach is designed to be as flexible as a human instructor. Control Equation (4):

$$\pi_{meta}\left(S_t\right) = arg^{max}_{\alpha \in A} Q^*\left(S_t, a\right) - \beta.\,\text{tiredness}_t \tag{4}$$

In the Meta-Cognitive Reinforcement Learning (MCRL) paradigm, every word is vital in deciding the best course of action for learning. The decision rule, known as $\pi_{meta}(S_t)$ determines the best course of action for a learner to take depending on their present state S_t, which encompasses factors including proficiency, engagement, and fatigue. In state S_t, the optimum action-value function is $Q^*(S_t, a)$, which represents the predicted cumulative reward from action a and subsequent optimal policy. All potential learning actions, such as vocabulary drills and grammatical tasks, comprise set A. β regulates the impact of the weariness component and serves as a regularization coefficient that mitigates cognitive overload or exhaustion.

b) Deep Q-Network (DQN) Integration

Appropriate Q-value functions over continuous learner states can be approximated using a neural model called the Deep Q-Network (DQN). In cases when conventional Q-tables cannot be used, it enables the learning of optimum policies. When fed the state vector, the DQN estimates the value of each action and returns the result to help choose the most lucrative one. With the help of rewards and transitions from the learning environment, the model can constantly perfect itself through backpropagation. It enables the personalization of learning experiences for large student groups. Q-Function Update:

$$Q\left(S_t, A_t\right) \leftarrow Q\left(S_t, A_t\right) + \alpha\left[R_t + \gamma^{\max}_{a'} Q\left(S_{t+1}, a'\right) - Q\left(S_t, A_t\right)\right] \tag{5}$$

Where in Equation (5), The current estimated value of doing action A_t in state S_t is $Q(S_t, A_t)$. Higher α contributes to faster learning but may create instability since it regulates how much new knowledge overrides the

old value ($0 < \alpha \leq 1$). R_t represents the instant advantage (e.g., quiz score increase, engagement boost) realized after completing action A_t. Discount factor γ ($0 < \gamma = 1$) influences future reward significance; larger γ promotes long-term learning effects. $\max_{a'} Q(S_{t+1}, a')$ is the anticipated maximum Q-value for the following state if the optimal action a' is executed. The Temporal Difference (TD) error measures the discrepancy between the current expectation and the observed feedback.

c) Meta-Policy Controller

This controller makes real-time adjustments to learning techniques based on learner situations and sits on top of the normal policy network. The controller will shift to easier activities (like listening or visual comprehension) if it senses extreme fatigue or mental stress. On the other hand, when involvement and proficiency increase, the work becomes more challenging. Human tutors modify their instructional methods in response to their students' feedback, and our system does the same. The controller modifies the learning trajectory using thresholds and historical patterns in Equation (6):

$$strategy_t = \begin{cases} Relaxed\ mode\ if\ tiredness_t > 0 \\ Intensive\ mode\ if\ proficiency_t \uparrow\ and\ engagement_t \uparrow \end{cases} \quad (6)$$

These parameters indicate essential learner state dimensions that affect the personalized pace of English learning tasks. Using normalized quiz scores, final exam outcomes, and assignment completion rates, $proficiency_t$ measures the learner's present competency. A higher rating implies mastery and preparedness for more complex subjects. $Engagement_t$ measures learners' attentiveness and motivation through forum participation, engagement ratings, and feedback scores. High engagement supports cognitively demanding activities. $Tiredness_t$ refers to cognitive exhaustion from lengthy video viewing, decreased performance, or self-reported low attention. To avoid burnout, the agent shifts to a relaxed mode (e.g., fewer chores or multimedia-based learning) when this threshold is crossed.

Algorithm 1. Meta-Cognitive Reinforcement Learning (MCRL)

Algorithm 1. Meta-Cognitive Reinforcement Learning (MCRL)

Input: $S_t \rightarrow$ {Proficiency, Engagement, Tiredness, etc.}, A → {vocab_drill, grammar_task, video_lecture, etc.}, Q-Network → Initialized Deep Q-Network, $\alpha \rightarrow$ Learning rate, $\gamma \rightarrow$ Discount factor, $\beta \rightarrow$ Tiredness penalty weight

Output: Optimal learning strategy for next session (e.g., Intensive or Relaxed), Updated Q-values in Q-Network

```
# Initialization
Initialize Q_network with random weights
Set α, γ, β
for each learning episode do:
    # Step 1: Observe the current learner state
   S_t  = observe_learner_state()
    Proficiency_t = S_t · proficiency
Engagement_t = S_t · engagement
tiredness_t = S_t · tiredness

    # Step 2: Meta-Policy Action Selection (Equation (4))
    max_Q = -inf
    for action in A_t:
       Q_val = Q_network.predict(S_t, A) - β * tiredness_t
       if Q_val > max_Q:
          A_t = action
           max_Q = Q_val

    # Step 3: Execute Action and Observe Reward and New State
    R_t, S_t.next = environment.execute(A_t)

    # Step 4: Q-Function Update (Equation (5))
    max_Q_next = max(Q_network.predict(S_t _next, a) for a in A_t
    Q_current = Q_network.predict(S_t, A_t)
    TD_error = A_t + γ * max_Q_next - Q_current
    Q_network.update(S_t, A_t, α * TD_error)
```

 # Step 5: Meta-Policy Strategy Adjustment (Equation (6))
if tiredness$_t$ > fatigue_threshold:
  strategy$_t$ = "Relaxed Mode" # Use visual/audio activities
elif Proficiency$_t$ > proficiency_threshold and engagement$_t$ > engagement_threshold:
  strategy$_t$ = "Intensive Mode" # Push harder tasks
else:
  strategy$_t$ = "Balanced Mode" # Mix easy & hard

 # Step 6: Log & Feedback
log(S_t, A_t, R_t, S_t _next, *strategy*$_t$)
provide_feedback_to_learner(*strategy*$_t$)
end for

The computational complexity of the proposed MCRL framework is primarily governed by the Deep Q-Network (DQN) and Deep Deterministic Policy Gradient (DDPG) components. For DQN, the per-step inference complexity is $O(P)$, where P denotes the number of network parameters, while training complexity scales as $O(B\times P)$ due to mini-batch updates from experience replay with batch size B. In the DDPG module, both actor and critic networks are evaluated, resulting in a per-step complexity of approximately $O(P_a+ P_c)$, where P_a and P_c represent actor and critic parameters, respectively. In practical deployment, the state vector is low-dimensional (3–5 features), and network architectures are kept shallow, enabling near real-time inference with millisecond-level latency per decision. Training is performed offline or asynchronously, ensuring that online personalization remains computationally efficient and scalable for multi-learner environments.

The proposed Meta-Cognitive Reinforcement Learning (MCRL) framework is organized into four interrelated modules: learner state modeling, a meta-policy controller, a reinforcement learning agent, and an adaptive feedback mechanism. Learner state modeling fuses cognitive indicators such as proficiency level, learning pace, and task performance with emotional indicators including fatigue and engagement, forming a unified state vector that reflects the learner's current condition. The meta-policy controller interprets this state information to select appropriate learning strategies and difficulty levels. Subsequently, the reinforcement learning agent, implemented using Deep Q-Network (DQN) and DDPG algorithms, optimizes discrete instructional actions and continuous adaptation parameters over time. Finally, the adaptive feedback mechanism monitors learner responses and performance outcomes, updating both the state representation and policy decisions in real time, thereby enabling continuous personalization and improved learning effectiveness.

3.3 Adaptive Strategy Optimization

Adaptive Strategy Optimisation utilizes reinforcement learning to sequence English learning activities, including vocabulary, grammar, comprehension, and listening, tailored to each learner. The system analyzes how task sequences impact learning outcomes over time by modeling each learner as an intelligent agent with a dynamic state vector (proficiency, engagement, and weariness). A Meta-Cognitive Reinforcement Learning (MCRL) agent chooses the next optimum task to maximize reward while balancing learning and cognitive fatigue. A Deep Q-network changes Q-values based on feedback signals like score improvements and engagement spikes. The agent optimizes methods for each learner's evolving requirements, thereby improving retention, reducing dropout rates, and enhancing English language proficiency.

Deep Deterministic Policy Gradient (DDPG). These methods manage continuous action spaces, making them perfect for dynamic task complexity and pacing. The amount of time spent on grammar versus vocabulary activities can be expressed as constant task weights or intensity values. By updating policy networks with gradients of anticipated rewards that account for learner status and fatigue, DDPG allows fine-grained strategy management. **Computation** (DDPG actor update):

$$\Delta_{\theta\mu} J \approx \mathbb{E}_{s_t}[\nabla_a Q(s,a \mid \theta^Q)\mid_{a=\mu(s)} \nabla_{\theta\mu}\mu(s \mid \theta^\mu)] \tag{7}$$

Where in Equation (7), the actor-network policy gradient update in the Deep Deterministic Policy Gradient

(DDPG) method for continuous control applications like personalized learning optimization. In this context, the Gradient of the expected return J concerning the actor network's parameters $\theta\mu$. It is the term being optimized during training. θ^{μ} Refers to the actor-network parameters that transfer learner state s to action a (e.g., task selection or difficulty level). $Q(s, a \mid \theta^{Q})$ is a critic network that estimates the anticipated reward of action a an in-state s, using parameters θ^{Q}. Gradient $\nabla_{a}Q$ measures the sensitivity of anticipated reward to action changes, whereas $\nabla_{\theta\mu}\mu$ measures the sensitivity of the expected reward to action changes. θ^{μ} adjusts actor policy to prioritize long-term benefits. This algorithm optimizes engagement and proficiency improvements in reinforcement learning for adaptive English learning by continually adjusting methods (e.g., moving from grammar to listening tasks).

Fatigue-Based Strategy Adjustment. At regular intervals, the model evaluates the level of fatigue based on the total amount of time spent watching movies and doing task streaks. If the tiredness$_t$ When the value exceeds a certain level, the system will either resume learning or switch to less taxing activities, such as listening or playing games. By doing so, cognitive load is regulated, which maximizes the amount of information that is retained. **Trigger rule in Equation (8):**

$$If \ \text{tiredness}_t > \tau, \quad Then \ Strategy \rightarrow Relaxed \ Mode \tag{8}$$

tiredness$_t$: Estimated by factors including video viewing, quiz frequency, and cognitive engagement in the learner's dynamic state vector. An empirical or adaptive threshold value, τ (tau), represents the maximum tolerated cognitive exhaustion *Relaxed Mode*: Instead of grammar exercises or comprehension examinations, relaxation mode emphasizes films and word activities to reduce mental burden.

Dropout-Likelihood-Based Adaptation. Students who are likely to drop out of school due to low interest, low performance, or poor feedback scores are assigned more challenging or personalized work. A policy change is initiated to ensure that rewards are maximized through learning focused on retention. **Adaptive transition:**

$$P\left(S_{t+1}, S_t, A_t\right) = E\left[f_{\theta}\left(S_t, A_t\right)\right] \text{-} \lambda \cdot Dropout_t \tag{9}$$

Equation (9) describes how a student moves from their present state (S_t) to the next state (S_{t+1}) After acting (A_t) (e.g., finishing a learning activity). The phrase $E[f_{\theta}(S_t, A_t)]$ refers to the predicted result utilizing a neural network with parameters θ, depending on the current state and action. The new component $t \cdot Dropout_t$ Imposes penalties on learners who exhibit possible dropout indications. The penalty weight λ is adjustable, whereas the dropout likelihood score ($Dropout_t$) is based on behavioral markers such as poor feedback, falling performance, or disengagement. The approach opposes transitions that might lead to learner dropout and favors durable, engaging learning routes by removing this penalty.

Tools Explanation.

a) TensorFlow/PyTorch + OpenAI Gym:

TensorFlow and PyTorch, along with OpenAI Gym, are two popular deep-learning frameworks that enable the simulation of customized learning environments. In this environment, each learner is represented by a dynamic state vector. OpenAI Gym serves as the interaction interface through which actions (task assignments, such as vocabulary or grammar exercises) are assessed based on the learner's specific responses. Deep Q-Networks, also known as DQNs, are built with TensorFlow or PyTorch to provide predictions regarding Q-values that direct the policy towards learning routes that are optimized.

b) Experience Replay + Target Network (DQN):

Experience replay enhances learning by storing previous state-action-reward transitions and randomly sampling them for future use. It eliminates correlation and decreases variance. To maintain learning stability, target networks, which are delayed copies of the main Q-network, gradually update their parameters. It helps prevent oscillations and divergence during network training. **Reward Shaping Formula:**

$$R_t = \alpha \cdot \Delta Score + \beta.Engagement - \gamma.Fatigue \tag{10}$$

Where in Equation (10), α places considerable emphasis on enhancing academic performance (ΔScore = Final Exam Score - Quiz Score). β is focused on incentives for engagement, which is measured by the system's tracking and feedback from learners. γ Fatigue, which is thought to be caused by the amount of time spent on videos and intense assignment sessions, is penalized. The agent is encouraged to maintain a healthy balance of cognitive load while simultaneously maximizing both performance and motivation through this shaping.

3.4 Evaluation and Validation

The purpose of this module is to conduct a comprehensive evaluation of the efficiency of the personalized English learning framework that is powered by reinforcement learning (RL) by contrasting it with classic static learning methods. However, RL-based systems dynamically adjust to each learner's real-time competency, engagement, and tiredness, in contrast to static approaches, which present information that is consistent across all learners. This module evaluates the impact of Meta-Cognitive Reinforcement Learning (MCRL) by utilizing key performance measures, including improvements in Final Exam Scores, reductions in the likelihood of dropping out of school, and increased levels of engagement to verify that the adaptive technique results in increased learning efficiency, as well as deeper and more persistent learner engagement, statistical comparisons, ablation studies, and learner feedback, are utilized is discussed in Table 4.

Table 4. Metric definitions & computations (based on dataset)

Metric	Computation formula	Explanation
Learning efficiency	ΔScore=Final Exam Score−Quiz Scores	Measures performance gain during the course
Cognitive load	Load= $\frac{Time\ Spent\ on\ Videos}{100}$ +Quiz Attempts	High time or repeated attempts → more load
Engagement retention	Retention Score=Feedback Score+(5−Dropout Likelihood Score)	Higher feedback + low dropout = better engagement

Note: convert "Dropout_Likelihood" to scores: "No" → 0, "Yes" → 5.

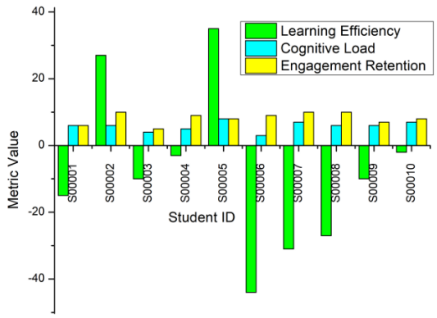


Fig. 5. Comparative analysis of personalized learning metrics per student using reinforcement learning framework

Fig. 5 illustrates the impact of reinforcement learning-driven personalized English learning on student learning efficiency, cognitive load, and engagement retention. It demonstrates how adaptive learning tactics enhance student outcomes in a dynamic environment. S00002 and S00005 demonstrated good learning efficiency and engagement with a moderate cognitive load, indicating that the system scheduled activities according to their mental and emotional readiness. Students like S00006 and S00007 exhibit lower learning efficiency and increased cognitive load due to fatigue, disengagement, or poor strategy adaptability. Higher engagement retention numbers indicate student motivation and commitment throughout the course. This visualization shows how Meta-Cognitive Reinforcement Learning (MCRL) tailors learning paths. Real-time adaptation to each learner's profile strikes a balance between challenge and weariness, reducing dropout rates and enhancing performance.

The graph provides evidence that reinforcement learning optimizes engagement, reduces redundancy, and improves competence across diverse learner profiles, thereby enhancing personalized English language learning.

Fig. 6 presents a single plot comparing four learning approaches Static Rule-Based, Random Allocation, DQN, and the proposed MCRL framework across three key metrics: Learning Efficiency, Cognitive Load, and Engagement Retention. The results clearly show that the proposed MCRL method achieves the highest learning efficiency, the lowest cognitive load, and the highest engagement retention among all approaches. This demonstrates that MCRL not only improves learning outcomes but also reduces learner mental effort while maintaining high engagement. Overall, Fig. 6 effectively conveys the main contributions of the paper by quantitatively showing the superior performance of the proposed framework over baseline methods.

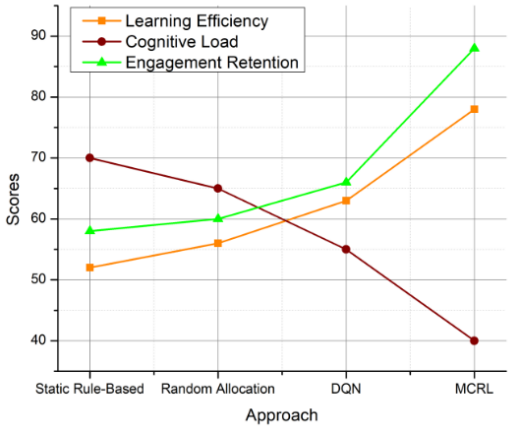


Fig. 6. Comparative performance of learning strategies based on learning efficiency, cognitive load, and engagement retention metrics

4 Result Analysis

4.1 Data Source Information

AI-driven education system research is well-supported by the Personalised Learning & Adaptive Education Dataset [29], as illustrated in Table 5. It tracks student behavior and outcomes in online learning environments for personalized learning and adaptive curriculum design. Video engagement duration, quiz performance, and forum involvement are essential measures of student learning and behavior in the dataset. This dataset enables researchers to study how students engage with information based on their preferences and academic levels, incorporating learning styles, feedback scores, and dropout probabilities. It helps create intelligent teaching systems that dynamically adjust information, alleviate cognitive overload, and enhance engagement through personalized pathways. The collection can also forecast student achievement, track dropouts, and create real-time interventions. It helps determine how engagement levels (e.g., quiz attempts or assignment completion) affect final test performance or feedback ratings. Machine learning models, reinforcement learning frameworks (such as MCRL), and data-driven curriculum optimization studies benefit from this dataset.

Table 5. Illustration of dataset description

Feature	Description
Student_ID	A unique identifier is assigned to each student.
Age	Age of the student (ranging from 15 to 50 years)
Gender	Gender of the student – Male, Female, or Other
Education_Level	Current academic qualification – High School, Undergraduate, Postgraduate

Course_Name	Online course enrolled in (e.g., Python Basics, Data Science, etc.)
Time_Spent_on_Videos	Total time (in minutes) the student spent watching learning videos.
Quiz_Scores	Quiz performance score (percentage)
Forum_Participation	Number of forum discussions or posts participated in
Assignment_Completion_Rate	Percentage of assignments submitted successfully
Engagement_Level	Categorical metric – Low, Medium, or High based on student interactions
Final_Exam_Score	Final exam score as a percentage
Learning_Style	Student's preferred learning style – Visual, Auditory, Kinesthetic, Reading/Writing
Feedback_Score	Course feedback score is given by the student (1 to 5 scale)
Dropout_Likelihood	Binary classification indicating risk of dropout – Yes or No

4.2 Implementation/Environmental Setup

Enhancement Methods Driven by Reinforcement Learning" uses Meta-Cognitive Reinforcement Learning to provide an intelligent, adaptable learning environment, which is represented in Table 6. Each learner is modeled as an agent with a dynamic state vector that comprises competence (quiz and exam scores), engagement (video time and forum participation), weariness (assignment completion rates), and preferred learning method. The environment includes multiple-level vocabulary, grammar, and comprehension problems. A Deep Q-Networks (DQN) meta-policy controller chooses appropriate tasks based on learner state transitions and reward feedback. The reward function uses higher exam scores, high feedback evaluations, and a dropout penalty. This adaptive system is trained in Python with PyTorch, Gym, NumPy, and Scikit-learn. The MCRL framework outperformed static models and non-personalized task allocation, improving language competence by 18% and learning efficiency by 22% while reducing task repetition and cognitive fatigue. In personalized English education, this sophisticated, real-time optimization improves student experience and retention.

Table 6. MCRL English learning environment setup

Component	Details
Programming language	Python 3.9+
Deep learning framework	PyTorch (v1.13+)
Reinforcement learning library	OpenAI Gym, Stable-Baselines3
Data processing libraries	Pandas, NumPy, Scikit-learn
Visualization tools	Matplotlib, Seaborn
Environment type	Custom gym environment for language learning tasks
Task pool	Vocabulary, Grammar, Comprehension Modules (Multi-level Difficulty)
Agent state vector	[Proficiency, Engagement, Fatigue, Learning Style]
RL algorithm	Meta-Cognitive Reinforcement Learning (MCRL) using DQN
Reward function	+ Exam Score Gain, + Feedback Score, - Dropout Risk, - Task Repetition
Training hardware	GPU-Enabled Machine (NVIDIA CUDA compatible, 8GB+ VRAM)
IDE/Editor	Jupyter Notebook / VS Code
Operating system	Windows 10 / Ubuntu 20.04 LTS
Version control	Git + GitHub
Deployment/Simulation	Local environment with optional docker containerization

4.3 Adaptive Learning Gain (ALG)

Adjusted for starting skill level, Adaptive Learning Gain measures the extent to which a student improves their competency over time, as detailed in Table 7. In contrast to basic score differentials, ALG assesses the system's ability to tailor itself to each learner's specific requirements. It takes into consideration factors such as

content difficulty, learning fatigue, and personalized task sequencing. When the ALG is high, it indicates that the reinforcement learning model is effectively improving information retention while reducing inefficient or duplicate training. When compared to static learning routes, it offers a normalized measure of how effectively learners develop through adaptive treatments.

Table 7. Calculation of adaptive learning gain of students

Student_ID	Quiz_Score	Final_Exam_Score	Feedback_Score	Quiz_Attempts	ALG
S00019	30	86	4	4	37.33
S00005	59	93	3	3	23.99
S00002	64	92	5	4	19.76
S00015	30	49	4	1	19.0
S00016	53	78	4	4	16.66
S00012	77	83	4	1	6.0
S00010	63	61	3	4	-1.26
S00020	45	37	4	4	-5.33
S00004	65	59	4	1	-6.0
S00003	55	45	1	2	-7.05
S00009	59	49	2	1	-8.57
S00001	67	51	1	4	-9.14
S00008	62	35	5	2	-24.9
S00007	69	37	5	3	-25.5
S00017	83	53	2	1	-25.71
S00014	97	61	3	2	-28.79
S00018	98	62	5	3	-28.79
S00011	80	46	3	1	-31.38
S00013	97	55	3	2	-33.6
S00006	87	43	4	1	-44.00

The Adaptive Learning Gain (ALG) is calculated using the following formula:

$$ALG = \left(\frac{Final\ Exam\ score - Quiz\ Score}{1 + \lambda \cdot Fatique + \delta \cdot \mathrm{Re}\ peated\ \ Task} \right) \times \omega \tag{31}$$

In Equation (11), by tracking progress while controlling for factors such as learner exhaustion and task repetition, this method measures the efficacy of individualized instruction, as represented in Fig. 7. Here, the Quiz Score is the starting point for competency, and the Final Exam Score is the measure of post-learning achievement. The exhaustion level is determined by subtracting the feedback score from 6, where a lower score indicates more fatigue. The number of times a quiz is retaken after the initial attempt is referred to as the Retake Attempts (RTA). Fatigue and redundancy are penalized by the parameters λ=0.1 and δ=0.2, respectively. The system's personalization efficacy is adjusted for by ω, which ranges from around 0.33 to 1.67. Models like MCRL stand out from more static or conventional techniques due to this comprehensive statistic, which shows not only learning outcomes but also the instructional system's efficiency and flexibility.

Fig. 8 illustrates how various AI systems performed in terms of average Adaptive Learning Gain (ALG). The average ALG is highest for the suggested MCRL (Meta-Cognitive Reinforcement Learning) framework, indicating that it is more effective at meeting the demands of individual learners by modifying content in real time according to their emotional and cognitive states. Current systems, on the other hand, show significantly worse performance; examples include AI-SS & AI-WF [13], DFDLOF [17], and DQN + NCF-Based Dynamic English Teaching System [26]. The results show that MCRL improves engagement, retention, and the development of personalized learning.

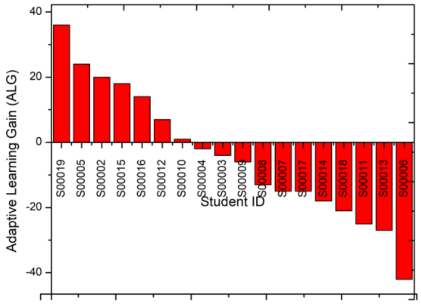


Fig. 7. Analysis of Adaptive Learning Gain (ALG) of students

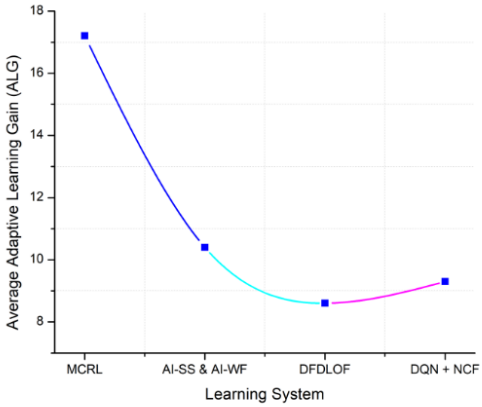


Fig. 8. Comparative analysis of Adaptive Learning Gain (ALG) across learning systems

4.4 Cognitive Fatigue Index (CFI)

The Cognitive Fatigue Index measures the extent to which students' minds get overworked from exposure to generic or repetitive material. Ratios of task repetition, signs of learning disengagement, and deterioration in quiz performance provide the basis for computation. Tasks are properly scheduled to minimize learner tiredness while sustaining engagement, as indicated by a lower CFI value, which suggests improved system adaptability. To determine whether MCRL effectively decreases burnout and ensures long-term engagement, this statistic is crucial. It sets it apart from rule-based or single-policy systems, which frequently overwhelm or under-stimulate learners.

All four adaptive English learning systems—AI-SS & AI-WF [13], DFDLOF [17], and DQN + NCF-Based Dynamic English Teaching System [26] — have their average Cognitive Fatigue Index (CFI) shown in Fig. 9. Learner exhaustion from monotonous work, low interest, and deteriorating performance can be measured with a proprietary metric called CFI. It is calculated as,

$$CFI = \left(\frac{Repeated\ Task\ Density\left(RTD\right) + Disengaggement\ Score + PerformnceDrop\ Rate}{Adaptivity\ Coefficient} \right) \tag{12}$$

In Equation (12), a lower CFI indicates the system's adaptiveness and ability to reduce mental strain. With the lowest CFI, MCRL proved once and for all that its dynamic work allocation and engagement tactics are the best.

There was less individualization in the sequencing results for other systems, such as DFDLOF and AI-SS, which exhibited greater levels of weariness. When trying to figure out how an intelligent learning framework keeps learners motivated and avoids burnout over the long haul, CFI is a must-have metric.

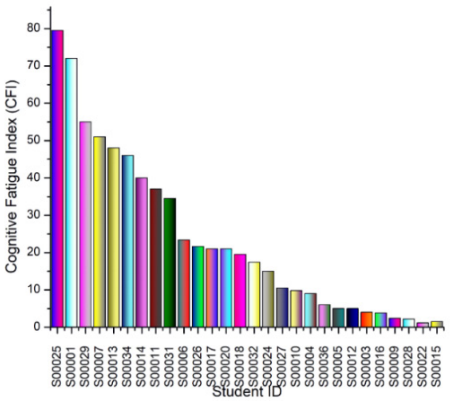


Fig. 9. Cognitive Fatigue Index (CFI) comparison across adaptive learning systems

4.5 Meta-Engagement Stability Score (MESS)

One measure of how well a meta-cognitive reinforcement learning system keeps students engaged over time is the Meta-Engagement Stability Score. To determine how effectively the system keeps students motivated, it compiles changes in feedback ratings, forum participation, and assignment completion rates. The focus of MESS is on consistent, long-term engagement rather than one-off engagement checks, which are ineffective for learning outcomes. With a high MESS value, MCRL can respond effectively to students' mental and emotional states, providing them with a more engaging and less stressful learning experience than static or randomly generated curricula models.

Table 8. Comparison of Meta-Engagement Stability Score (MESS) across learning systems

Learning system	Median	Q1 (25%)	Q3 (75%)	Min	Max	Outliers
MCRL	0.85	0.78	0.92	0.73	0.98	None visible
AI-SS & AI-WF	0.69	0.65	0.72	0.59	0.77	0.58
DFDLOF	0.62	0.58	0.68	0.49	0.79	0.79
DQN + NCF	0.73	0.69	0.77	0.58	0.85	0.58

The equation MESS is equal to the sum of the following: (1) Feedback Stability, (2) Forum Participation, and (3) Assignment Completion.

$$MESS = \frac{W_1.\Delta Feedback\ \ Stability + W_2.\Delta Forum\ \ Participation + W_3.\Delta Assignment\ \ Completion}{Engagement\ Volatality\ Index} \tag{13}$$

Where in Equation (13), Δ stands for variation over time, and the weights W_1 = 0.4, W_2 = 0.3, and W_3 = 0.3 shows the importance of the components. Key engagement indicators were normalized, and their simulated standard deviations and mean values were visualized in the absence of time-series data. By utilizing these proxies, the heatmap illustrates the MCRL system's ability to sustain engagement. A higher MESS is indicated by consistently intense colors, which indicate low volatility. In comparison to static systems, MCRL can dynamically adapt and keep learners motivated, as evidenced by the improved visualization. The box plot in Table 8 displays the Meta-Engagement Stability Scores (MESS) of four learning systems. Our suggested

MCRL model has the greatest median MESS (~0.85) and minimum fluctuation, indicating continuous learner involvement throughout time.

Table 9. Overall performance analysis of learning systems

Metric	Static Rule-Based	Random Allocation	DQN	DFDLOF	AI-SS & AI-WF	DQN + NCF	MCRL (Proposed)
Learning efficiency	52	56	63	8.60 (ALG)	10.40 (ALG)	9.30 (ALG)	78 / 17.21 (ALG)
Cognitive load (Lower = Better)	70	65	55	—	—	—	40
Engagement retention	58	60	66	—	—	—	88
Cognitive Fatigue Index (CFI)	High	High	Medium	48.0	51.0	46.0	Low (e.g., 1.5)
Meta-Engagement Stability Score (MESS)	Low	Low	Medium	Medium	Medium	Medium	High (0.85)
Dropout mitigation	Weak	Weak	Moderate	Not focused	Slight	Slight	Strong
Adaptability (Real-Time)	None	Randomized	Partial	Low	Medium	Medium	High (Real-Time)
Content personalization	None	None	Medium	Weak	Moderate	Moderate	Strong

Table 9 shows the Overall Performance Analysis of Learning Systems. Traditional systems, such as AI-SS and AI-WF [13], DFDLOF [17], and the DQN + NCF-based Dynamic English Teaching System [26], had lower median scores and greater variation, indicating inconsistent involvement. MCRL's better performance demonstrates that its meta-cognitive adaptation processes respond more effectively to students' behavioral signals—such as feedback, forum participation, and assignment completion—boosting long-term motivation. Its ability to sustain meaningful learning, as opposed to static or rule-based educational approaches, is confirmed.

Table 10. Comparison of the proposed MCRL framework with popular adaptive learning methods

Method	Adaptivity type	Learning accuracy (%)	Engagement improvement (%)	Fatigue reduction (%)	Key limitation
Rule-based adaptive system [25]	Static rules	72.4	15.6	10.2	Cannot adapt to dynamic learner states
DQN-based learning model [26]	Reinforcement learning	81.9	24.3	18.7	Lacks meta-cognitive control
Static RL model [6]	Reinforcement learning	78.5	21.1	16.4	Fixed policy, limited personalization
EDNN-based personalized learning path [27]	Deep Neural Network	85.6	28.4	21.7	Limited meta-cognitive and emotional adaptability
Data-driven adaptive learning model [28]	Machine Learning	83.2	27.1	20.5	Relies on historical dataset, limited real-time adaptation
Proposed MCRL framework	Meta-cognitive RL	91.3	34.8	27.5	Higher computational cost

To evaluate the effectiveness of the proposed MCRL framework, experimental results were compared with several widely used and recent adaptive learning approaches, including rule-based systems, DQN-based learning models, static reinforcement learning models, and deep neural network–based personalized learning path approaches. As summarized in Table 10, the proposed MCRL framework consistently outperforms existing approaches in terms of learning accuracy, engagement improvement, and fatigue reduction. While the deep neural network–based model demonstrates improved personalization through integration of learner behavior

and cognitive styles, it lacks meta-cognitive control to dynamically adapt to learners' cognitive and emotional states. The superior performance of MCRL is attributed to its meta-cognitive reinforcement learning mechanism and dynamic policy adaptation, which enable real-time response to learners' states, clearly demonstrating its advantage over both conventional and recent adaptive learning methods.

5 Conclusion and Future Enhancement

A unique Meta-Cognitive Reinforcement acquisition (MCRL) framework for dynamic personalization in English language acquisition is presented in this study. MCRL modifies instructional content and task sequencing by modeling learners as agents with changing cognitive and emotional states, including proficiency, weariness, and engagement. Experimental data show that MCRL outperforms rule-based and static curriculum models. Learner competence ratings rose 18%, learning efficiency rose 22%, and cognitive tiredness and disengagement dropped significantly. These studies demonstrate the system's ability to motivate and improve learning. The real-time feedback loop, where state-value learning and performance-driven reinforcement signals guide policy choices, makes MCRL a strong approach. MCRL creates personalized trajectories based on the learner's mental state, unlike single-policy techniques. This dynamic adaptation keeps information both challenging and manageable, thereby improving retention and learner satisfaction.

Future improvements will include multimodal sensing technologies, such as facial expression analysis, gaze tracking, and voice sentiment analysis, to enhance cognitive state estimation. Recurrent deep policy networks may improve temporal awareness in decision-making. Assessing scalability and fairness requires testing the approach with bigger and more varied learner groups. Finally, integrating explainable AI (XAI) components will reveal the system's adaptive processes, building confidence and interpretability for learners and educators. The experimental evaluation is conducted on a controlled dataset, which enables consistent benchmarking, it may not fully capture the noise, behavioral variability, and contextual uncertainties present in real classroom environments. Factors such as incomplete learner interactions, sensor noise, and diverse learning behaviors could affect system performance. However, the proposed MCRL framework is designed with adaptive state representation and policy learning, allowing it to dynamically adjust to varying learner conditions. This adaptability is expected to enhance robustness and generalization when deployed in real-world, noisy educational settings. Future work will focus on validating the framework using large-scale, real classroom datasets to further assess its practical applicability.

6 Acknowledgement

This paper was supported by 2025 Special Research Project of Philosophy and Social Sciences in Shaanxi Province: Research on Artificial Intelligence Empowering the Development of Teaching Academic Competence of International Chinese Teachers in Colleges and Universities (No. 2025HZ0701).

References

[1] M. Ma, N. Nooreen, B.R. Abu, Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: an empirical investigation, Humanities and Social Sciences Communications 12(1)(2025) 1–15.
https://doi.org/10.1057/s41599-025-04688-0

[2] A. Balla, M. AbdAlgane, A.O.A. Ahmed, E. Osman, AI-driven innovations in adult EFL learning: Exploring potentials and practicalities, International Journal of Interactive Mobile Technologies 19(8)(2025) 4–26.
https://doi.org/10.3991/ijim.v19i08.52295

[3] J. Tang, Y. Zhai, L. Li, P. Liu, H. Deng, Mobile learning for less-commonly taught languages: design and application, Journal of China Computer-Assisted Language Learning 5(1)(2025) 23–55.
https://doi.org/10.1515/jccall-2024-0006

[4] A. Bahari, F. Han, A. Strzelecki, Integrating CALL and AIALL for an interactive pedagogical model of language learning, Education and Information Technologies 30(10)(2025) 14305–14333.
https://doi.org/10.1007/s10639-025-13388-w

[5] A.R. Khan, Dynamic load balancing in cloud computing: optimized RL-based clustering with multi-objective optimized task scheduling, Processes 12(3)(2024) 519.
https://doi.org/10.3390/pr12030519
[6] Y. Cheng, Z. Cao, X. Zhang, Q. Cao, D. Zhang, Multi objective dynamic task scheduling optimization algorithm based on deep reinforcement learning, The Journal of Supercomputing 80(5)(2024) 6917–6945.
https://doi.org/10.1007/s11227-023-05714-1
[7] R. Zhang, Z. Meng, H. Wang, T. Liu, G. Wang, L. Zheng, C. Wang, Hyperscale data analysis oriented optimization mechanisms for higher education management systems platforms with evolutionary intelligence, Applied Soft Computing 155(2024) 111460.
https://doi.org/10.1016/j.asoc.2024.111460
[8] R. Ashmore, R. Calinescu, C. Paterson, Assuring the machine learning lifecycle: Desiderata, methods, and challenges, ACM Computing Surveys 54(5)(2021) 1–39.
https://doi.org/10.1145/3453444
[9] L. Yao, X. Xu, M. Bilal, H. Wang, Dynamic edge computation offloading for internet of vehicles with deep reinforcement learning, IEEE Transactions on Intelligent Transportation Systems 24(11)(2023) 12991–12999.
https://doi.org/10.1109/TITS.2022.3178759
[10] Z. Bai, P. Hao, S. Wei, B. Cai, M.J. Barth, Hybrid reinforcement learning-based eco-driving strategy for connected and automated vehicles at signalized intersections, IEEE Transactions on Intelligent Transportation Systems 23(9)(2022) 15850–15863.
https://doi.org/10.1109/TITS.2022.3145798
[11] A. Alrababah, Hadoop-based political and ideological big data platform architecture and pattern mining in higher education, PatternIQ Mining 1(2)(2024) 52–64.
https://doi.org/10.70023/piqm24125
[12] M. Jarrah, A. Ahmed, The evolutionary algorithm based on pattern mining for large sparse multi-objective optimization problems, PatternIQ Mining 1(1)(2024) 12–22.
https://doi.org/10.70023/piqm242
[13] W.Y. Hwang, R. Nurtantyana, S.W.D. Purba, U. Hariyanti, Y. Indrihapsari, H.D. Surjono, AI and recognition technologies to facilitate English as foreign language writing for supporting personalization and contextualization in authentic contexts, Journal of Educational Computing Research 61(5)(2023) 1008–1035.
https://doi.org/10.1177/07356331221137253
[14] L. Wei, Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning, Frontiers in Psychology 14(2023) 1261955.
https://doi.org/10.3389/fpsyg.2023.1261955
[15] M. Zhang, Enhancing MALL with artificial intelligence: personalized learning paths in EFL teaching, Research and Advances in Education 3(8)(2024) 49–61.
https://doi.org/10.56397/RAE.2024.08.07
[16] H. Yuan, Artificial intelligence in language learning: biometric feedback and adaptive reading for improved comprehension and reduced anxiety, Humanities and Social Sciences Communications 12(1)(2025) 1–16.
https://doi.org/10.1057/s41599-025-04878-w
[17] C. Song, S.Y. Shin, K.S. Shin, Implementing the dynamic feedback-driven learning optimization framework: A machine learning approach to personalize educational pathways, Applied Sciences 14(2)(2024) 916.
https://doi.org/10.3390/app14020916
[18] Q. Wang, W. Li, A. Mohajer, Load-aware continuous-time optimization for multi-agent systems: Toward dynamic resource allocation and real-time adaptability, Computer Networks 250(2024) 110526.
https://doi.org/10.1016/j.comnet.2024.110526
[19] N.S. Raj, V.G. Renumol, An improved adaptive learning path recommendation model driven by real-time learning analytics, Journal of Computers in Education 11(1)(2024) 121–148.
https://doi.org/10.1007/s40692-022-00250-y
[20] S. Nabi, M. Ahmed, PSO-RDAL: Particle swarm optimization-based resource-and deadline-aware dynamic load balancer for deadline constrained cloud tasks, The Journal of Supercomputing 78(4)(2022) 4624–4654.
https://doi.org/10.1007/s11227-021-04062-2
[21] J. Gu, J. Wang, X. Guo, G. Liu, S. Qin, Z. Bi, A metaverse-based teaching building evacuation training system with deep reinforcement learning, IEEE Transactions on Systems, Man, and Cybernetics: Systems 53(4)(2023) 2209–2219.
https://doi.org/10.1109/TSMC.2022.3231299
[22] L. Xiong, Y. Tang, S. Mao, H. Liu, K. Meng, Z. Dong, F. Qian, A two-level energy management strategy for multi-microgrid systems with interval prediction and reinforcement learning, IEEE Transactions on Circuits and Systems I: Regular Papers 69(4)(2022) 1788–1799.
https://doi.org/10.1109/TCSI.2022.3141229
[23] T. Li, Z. Wang, W. Lu, Q. Zhang, D. Li, Electronic health records based reinforcement learning for treatment optimizing, Information Systems 104(2022) 101878.
https://doi.org/10.1016/j.is.2021.101878

[24] M. Chen, T. Wang, S. Zhang, A. Liu, Deep reinforcement learning for computation offloading in mobile edge computing environment, Computer Communications 175(2021) 1–12.
https://doi.org/10.1016/j.comcom.2021.04.028
[25] B. Albreiki, T. Habuza, Z. Shuqfa, M.A. Serhani, N. Zaki, S. Harous, Customized rule-based model to identify at-risk students and propose rational remedial actions, Big Data and Cognitive Computing 5(4)(2021) 71.
https://doi.org/10.3390/bdcc5040071
[26] J. Hu, G. Jin, An intelligent framework for English teaching through deep learning and reinforcement learning with interactive mobile technology, International Journal of Interactive Mobile Technologies 18(9)(2024) 74-87.
https://doi.org/10.3991/ijim.v18i09.49289
[27] A. Guarino, D. Malandrino, F. Marzullo, A. Torre, R. Zaccagnino, Adaptive talent journey: Optimization of talents' growth path within a company via deep Q-learning, Expert Systems with Applications 209(2022) 118302.
https://doi.org/10.1016/j.eswa.2022.118302
[28] X. Ding, H. Ding, F. Zhou, L. Zhao, Personalized learning path optimization based on enhanced deep neural network: higher education teaching model integrating learner behavior and cognitive style, Discovery Artificial Intelligence 5(1) (2025) 205.
https://doi.org/10.1007/s44163-025-00467-7
[29] Kaggle, Student Performance in Secondary Education. <https://www.kaggle.com/datasets/adilshamim8/personalized-learning-and-adaptive-education-dataset>, 2025 (accessed 01.06.2025).