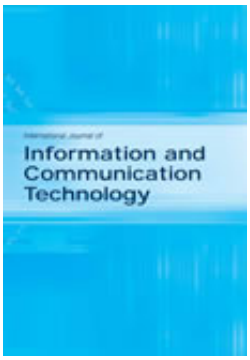
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Personalised learning path generation mechanism based on RL and knowledge graph

Zhao Liu, Yanan Shang

DOI: 10.1504/IJICT.2026.10075958

Article History:

Received:	27 September 2025
Last revised:	23 October 2025
Accepted:	29 October 2025
Published online:	09 February 2026

Copyright © 2026 Inderscience Enterprises Ltd.

Personalised learning path generation mechanism based on RL and knowledge graph

Zhao Liu and Yanan Shang*

Cangzhou Normal University,
Cangzhou, 061000, China
Email: liuzhao2514@caztc.edu.cn
Email: shangyanan@caztc.edu.cn
*Corresponding author

Abstract: This article proposes a personalised learning path generation mechanism that combines reinforcement learning and knowledge graphs. It constructs a knowledge graph that includes knowledge points, student status, and historical behaviour. Using this image and learning trajectory, it constructed a student model. Then, the RL algorithm optimises the learning path based on real-time feedback. An experiment targeting 300 college students compared the proposed method with traditional methods. The results showed that reinforcement learning based methods improved learning outcomes by 12.5%, increased learning satisfaction by 23.7%, accelerated knowledge acquisition by 15%, and shortened average learning time by 8%. These findings confirm the effectiveness of this mechanism in improving learning outcomes and meeting individual student needs.

Keywords: personalised learning; reinforcement learning; knowledge graph; learning pathways; data analysis.

Reference to this paper should be made as follows: Liu, Z. and Shang, Y. (2026) 'Personalised learning path generation mechanism based on RL and knowledge graph', *Int. J. Information and Communication Technology*, Vol. 27, No. 8, pp.74–93.

Biographical notes: Zhao Liu earned his Bachelor's in Network Engineering from Xingtai University in 2015 and later obtained his Master's in Computer Technology from Shenyang Ligong University in 2018. He is currently a faculty member at the School of Computer Science and Engineering, Cangzhou Normal University, with research interests in computer course pedagogy reform, wireless sensor networks, and machine learning.

Yanan Shang holds a Master's in Art from Kunming University of Science and Technology (2016) and PhD in Philosophy (Education) from the University of São Paulo (2023). She serves as an Associate Professor in the Publicity Department at Cangzhou Normal University, specialising in education, cultural industry development, and new media and digital networks.

1 Introduction

With the continuous development of educational informatisation, personalised learning has become an important direction to improve the teaching effect and meet students' diverse needs (Afzali and Shamsinejadbabaki, 2025). Traditional teaching methods often

Copyright © The Author(s) 2026. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

adopt a unified curriculum arrangement, ignoring students' differences in learning interests, ability, and progress. This 'one size fits all' approach can't meet students' individual needs, and it is difficult to optimise the learning effect. Therefore, how to formulate an accurate learning path for students through intelligent means has become an important topic in current educational research. Personalised learning path generation not only helps to improve students' learning efficiency but also enhances students' learning motivation and interest, thus improving the overall educational level.

Reinforcement learning (RL), as an adaptive intelligent algorithm, can continuously optimise decision-making strategies through interaction with the environment and has significant advantages in generating personalised learning paths (Buciuman and Potra, 2025). Through RL, the system can dynamically adjust learning strategies according to students' feedback and performance, making the learning path more flexible and adaptable (Aritonang et al., 2025). However, the existing RL methods mostly rely on a single reward mechanism and do not fully consider the differences in students' knowledge mastery, interests, and preferences. Therefore, how to combine RL with other intelligent technologies to further improve the accuracy and practicality of learning path generation has become an important topic of current research.

Knowledge graph (KG) is a knowledge representation method based on graph structure, which can effectively show the semantic relationship and structured information between knowledge points (Chen et al., 2025b). In personalised learning, a KG can be used as a tool for the organisation and management of learning content. Revealing the relationship between different knowledge points can help the system better understand students' learning progress and knowledge mastery. By combining KGs, personalised learning paths can more accurately recommend appropriate learning content for students, avoiding knowledge blind spots and incoherent problems in traditional methods. Especially when dealing with large-scale learning data, KGs can effectively integrate multi-dimensional information and provide strong support for generating learning paths.

This paper proposes a personalised learning path generation mechanism integrating RL and KGs. Firstly, the method generates a preliminary learning path for each student by constructing a personalised learning model, combining students' learning behaviour data and subject knowledge points in the knowledge map. Then, the RL algorithm is used to dynamically optimise and adjust the learning path according to the students' real-time feedback and learning effect. Compared with traditional methods, the mechanism proposed in this paper can more accurately match students' individual needs and optimise students' learning progress in real time. Through experimental verification, this study shows that the personalised learning path generation mechanism integrating RL and KG significantly improves learning effect and efficiency.

2 Theoretical basis and related research

2.1 Basic theory of RL and KG

RL is a machine learning method that continuously learns the optimal decision strategy by interacting with the environment. In RL, the agent perceives the current state, selects an action, and updates its strategy according to the environment's feedback (reward or punishment) (Lu et al., 2025). The core elements of RL include state, action, reward,

policy, and value function. The goal of RL is to maximise the cumulative reward that the agent receives over the long term. In the field of education, RL is widely used in the generation of personalised learning paths. Constantly adjusting learning content and progress, students optimise learning paths according to their learning progress and feedback and improve learning effects (Chen et al., 2025a).

The balance between exploration and exploitation is a key issue in RL. Exploration means the agent tries new and possibly unfamiliar actions to discover better strategies. The utilisation is based on the known optimal strategy for decision-making (Ghadiri and Hajizadeh, 2025). Exploration can help the system discover new learning content or methods in personalised learning. At the same time, utilisation can ensure that students can continue to learn efficiently in the knowledge areas they have mastered (An et al., 2025). Through reasonable exploration and utilisation strategies, RL can provide students with a challenging learning path that conforms to their learning ability and improves the efficiency and effectiveness of the learning process.

KG is a way of representing knowledge through a graph structure that aims to show relationships between entities. In personalised learning, the function of a KG is to represent subject knowledge structurally and intuitively display the correlation and dependency between knowledge points to help the system understand students' knowledge mastery and learning needs (Murray et al., 2025). By combining students' learning progress with the relationship between knowledge points in the knowledge map, the system can more accurately infer students' knowledge gaps and recommend appropriate learning content. KGs can not only help the education system understand students' learning status but also provide effective semantic support in the process of generating personalised learning paths.

In generating personalised learning paths, the combination of RL and KG provides powerful capabilities for the system. RL can adjust the learning path according to students' real-time feedback through the decision optimisation of agents. At the same time, the KG provides rich subject background information to ensure that the recommended learning content is structural and logical (Nguyen et al., 2025). By combining RL and KG, a dynamically optimised learning path generation mechanism can be realised, which enables students to constantly adjust and improve their learning routes in the learning process, thus improving learning efficiency and effect. This integration mechanism provides a new solution for personalised education, which has important theoretical and practical significance.

2.2 Current status of personalised learning paths of RL and KG

With the development of educational technology, personalised learning has become an important direction to improve the learning effect. RL and KG have been widely used in personalised learning path generations as two powerful, intelligent technologies in personalised learning path generation. RL generates personalised learning paths by simulating students' learning processes and dynamically adjusting according to students' learning status and feedback. However, although existing RL approaches can optimise learning pathways through reward mechanisms, they often ignore the intrinsic structure of knowledge and the specific learning needs of students, resulting in a lack of systematicness and coherence in the generated learning pathways in some cases. Therefore, combining KG and RL has become a hot spot in current research, aiming to

compensate for existing methods' shortcomings and improve the effect of personalised learning path generation (Dai et al., 2024).

The application of KGs in personalised learning paths is gradually recognised. By constructing the relationship network between subject knowledge points, the KG can provide students with more structured and visual learning content (Park et al., 2025). A KG can help students understand the relationship between different knowledge points and reveal the relationship between the current and the knowledge points they have not mastered, thus providing strong support for generating personalised learning paths. At present, many personalised learning systems based on KGs have been able to recommend learning content according to students' learning progress and knowledge structure. However, they still face the challenge of accurately matching students' personalised needs and KG content.

Although RL and KG have achieved initial results in personalised learning path generation, it is still challenging to effectively combine the two (Sun et al., 2025a). In existing research, RL is often used independently, focusing on continuously optimising paths through the interaction between agents and the environment. In contrast, KGs are mainly used to support static knowledge structure. However, how to organically integrate the two, which uses the dynamic adjustment ability of RL and fully mines the semantic information in the KG, has become a key issue in current research. Some studies have begun to try to combine KGs with RL to achieve the optimisation and dynamic update of personalised learning paths. However, there are still many challenges in balancing the roles of the two and improving the system's overall performance.

The personalised learning path generation mechanism combining RL and KG is still in the exploratory stage. Most research focuses on using RL for path optimisation while providing auxiliary support through KG. Studies have shown that KGs can help RL to evaluate students' knowledge status better and choose the best learning path by providing students with a structured view of knowledge (Sun et al., 2025b). Some experimental results show that the personalised learning path generation mechanism combining RL with KG can effectively improve the learning effect, especially in improving learning efficiency and student satisfaction. However, the existing system has not been widely used in practical teaching, and it still needs to be further improved, especially in data collection and processing, model generalisation ability, and real-time response ability of the system. Therefore, future research must deeply explore optimising the model combining RL with KG to achieve more accurate and dynamic personalised learning path recommendations (Valko and Kudenko, 2025).

This study uses DQN and A3C RL algorithms to optimise personalised learning paths. By defining a composite state vector that includes knowledge mastery level, learning behaviour characteristics, progress, and historical trajectory, and combining it with a multidimensional action space that recommends knowledge points, resource selection, difficulty adjustment, and rhythm control, a reward function is designed that combines learning progress, efficiency, participation, performance improvement, and personalised matching weighting. The exploration and utilisation are balanced in a dynamic environment, and the cumulative reward is maximised through iterative interaction to generate an efficient personalised learning path that adapts to the evolving needs of students.

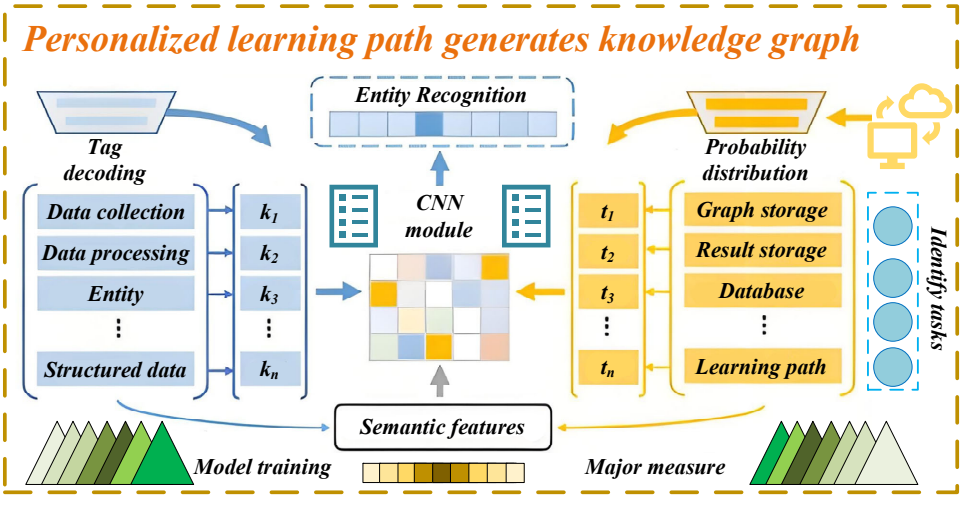
3 Establishment of personalised learning path model based on RL and KG

3.1 KG model construction

In building the KG, this paper adopts a top-down method. Since the collected datasets mainly come from textbooks and online resources, the amount of data is huge and unstructured, so the structure is scattered, resulting in low screening efficiency when using traditional knowledge extraction methods (Wang et al., 2025a). Therefore, this paper introduces the RCBC entity recognition and GCN relationship extraction models and realises triples' automatic extraction. In the context of personalised learning path generation, the process includes entity extraction, relationship extraction, and KG storage. For the named entity recognition task, since there is no clear word segmentation boundary in Chinese text, it is necessary to segment the text first, construct feature vectors, and train and test the type. The identified entities will be saved and used to complete the extraction of relationships between entities. Finally, the results of relationship recognition are stored in the neo4i graph database together with the entities obtained by entity recognition to complete the KG construction required for personalised learning path generation.

In this paper, the RCBC model (i.e., RoBERTa-CNN-BiLSTM-CRF model) is proposed, which consists of four modules, namely the RoBERTa module, CNN module, BiLSTM module, and CRF module. The RoBERTa module is responsible for pre-training the input text to extract the semantic features of the data and take these semantic features as the input of the CNN module to extract local features further and output them (Wang et al., 2025b). These features will then be fed into the BiLSTM module, which is used to predict the probability distribution of entity tags. Finally, the CRF module decodes the probability distribution of these entity tags, and the final entity tag output is obtained. The construction process of the personalised learning path in this paper is shown in Figure 1.

Figure 1 Personalised learning path generation KG construction process (see online version for colours)



In this paper's personalised learning path generation task, this model is used to identify and extract key entities and relationship information in the learning path, thus providing a basic KG for generating personalised learning paths. The semantic feature extraction formula is shown in (1).

$$S_{t+1} = f\left(S_t, a_t\right) \tag{1}$$

where S_t represents the state at the current time, a_t represents the action taken at the current time, and f represents the state transition function. The local feature extraction formula is shown in (2).

$$h_v = Emb_v\left(W_v\right) \tag{2}$$

where h_v represents the representation vector of node v, Emb_v represents the node embedding function, and W_v represents the feature matrix of node v. Firstly, the student's learning behaviour data are input, and the input learning data text is decomposed into the corresponding learning behaviour sequence by querying the learning feature table and input into the RoBERTa module to obtain the learning feature vector. Then, the learning feature vector is input to the CNN module, and the local features of each learning behaviour are extracted by CNN and then input to the BiLSTM module. The input hidden information is calculated by BiLSTM (Wang et al., 2025c). Finally, the CRF module is used to decode and solve the output of the BiLSTM module, and the prediction label of the personalised learning path is output. Combine CNN and BiLSTM, that is, use CNN coding to obtain the embedding of each learning behaviour, and then send it to LSTM to obtain the LSTM hidden layer vector (Ye et al., 2025). CNN can effectively externalise the local features of input data, and the bi-directionality of BiLSTM can enable the model to better consider past and future learning information to better grasp the context relationship in the learning process. Such a combination can learn local and global features and adaptively extract features, thus exhibiting excellent performance when processing learned data. Therefore, combining the ability of BLSTM to capture context information with the ability of CRF to model label dependencies can enhance the sequence labeling ability of the model, improve the effect and accuracy of personalised learning path generation, and ensure that the predicted learning path has high accuracy. The global information processing formula is shown in (3).

$$L = \sum_{i=1}^{n} \lambda_i \cdot f_i\left(s_i\right) \tag{3}$$

where L represents the total objective of the generated path, λ_i represents the weight of the i state in the path, and $f_i(s_i)$ represents the learning objective function of state s_i. The learning path prediction formula is shown in (4).

$$y = CRF(h) \tag{4}$$

Among them, h represents the hidden state vector sequence of all learning behaviours output by the BiLSTM module, y represents the personalised learning path prediction label sequence obtained by decoding by the *CRF* module, and *CRF* represents the conditional random field module. The learning behaviour vector vector is the vectors generated by Roberta at the input layer, the learning path vector, and the position vector (Yu et al., 2025; Zhang et al., 2025). The input data is transformed by querying the

learning feature table through the learning behaviour vector. The learning path vector represents the textual information of the learning path and is used to distinguish different learning activities. The position vector is coded by the position information corresponding to each learning behaviour, which can distinguish the semantic information of learning behaviours in different positions (Zhang and Li, 2025). The learning behaviour vector query formula is shown in (5).

$$L_{path} = \sum_{i=1}^{n-1} d\left(s_i, s_{i+1}\right) \tag{5}$$

where L_{path} represents the total length of the learning path, $d(s_i,s_{i+1})$ represents the distance measure from state s_i to state s_{i+1}, and n represents the number of samples. The input layer vector summation formula is shown in (6).

$$v_{input} = v_{behavior} + v_{path} + v_{pos} \tag{6}$$

Among them, $v_{behaviour}$ represents the learning behaviour vector, v_{path} represents the learning path vector, v_{pos} represents the position vector, and v_{input} represents the final vector of RoBERTa model input.

This study achieves continuous optimisation of the RL model through a closed-loop feedback mechanism: real-time reward signals are generated based on student test performance, and student status is dynamically updated by combining participation data. A daily incremental training mode is used to prioritise sampling key learning trajectories using an experience replay buffer, and an advantage weighted regression adjustment strategy is used to automatically increase the exploration probability of relevant teaching strategies when multiple students encounter similar knowledge difficulties. Finally, policy network parameters are updated through gradient descent to form a closed-loop loop from learning interaction to path optimisation, enabling personalised learning paths to quickly respond to students' evolving needs and constructing an adaptive educational experience based on teaching evidence.

Compared to collaborative filtering and Bayesian optimisation, the RL-KG framework explicitly models the learning domain structure through KGs, combines RL to achieve dynamic, long-term sequential decision-making, and adjusts paths based on real-time student status. It demonstrates better adaptability in handling complex and state based requirements generated by learning paths, becoming a more principled and comprehensive solution.

3.2 Algorithm improvement of CNN

Since CNN only considers the length of the learning path while ignoring the semantic relationship between learning behaviours when extracting learning features, it only uses maximum pooling in the pooling layer, considering the maximum eigenvalue of each feature map without considering other factors, which may lead to the loss of some key information (Zhou et al., 2025). Therefore, this paper proposes a multi-convolution and pooling neural network structure for improving CNN networks to extract semantic information in learning paths better and enhance the effect of personalised learning path generation. The multi-convolution operation formula is shown in (7).

$$C_i = Conv\left(v_{input}, W_k, b_k\right) \tag{7}$$

Among them, C_i represents the k convolution feature map output by the i convolution layer, W_k represents the convolution kernel parameter, b_k represents the convolution bias term, and *Conv* represents the multi-convolution operation formula. The multi-pooling operation formula is shown in (8).

$$P_{final} = Concat\left(P_1, P_2, ..., P_K\right) \tag{8}$$

Among them, P_1, P_2, …, P_K represents the feature map after multiple pooling operations, P_{final} represents the final pooling result, and *Concat* represents the splicing operation. Since graph neural network (GNN) is an effective method to solve multi-hop relationship reasoning problems, applying neural network to learning path graph structure can directly obtain the dependency information between nodes, thus alleviating the influence of multi-hop relationship on learning path generation (Dai et al., 2024; Ge et al., 2024). Using self-attention mechanisms can capture richer semantic information from students' learning data and enable models to understand the strength of associations between learning behaviours, thereby better using dependencies in learning paths. Therefore, this paper combines the attention mechanism with a GNN to strengthen the relationship between students' learning behaviours and the representation of implicit information to optimise the generation of personalised learning paths and effectively solve the relationship extraction problem in learning path recommendation. The weighting formula for the strength of the dependency relationship between nodes is shown in (9).

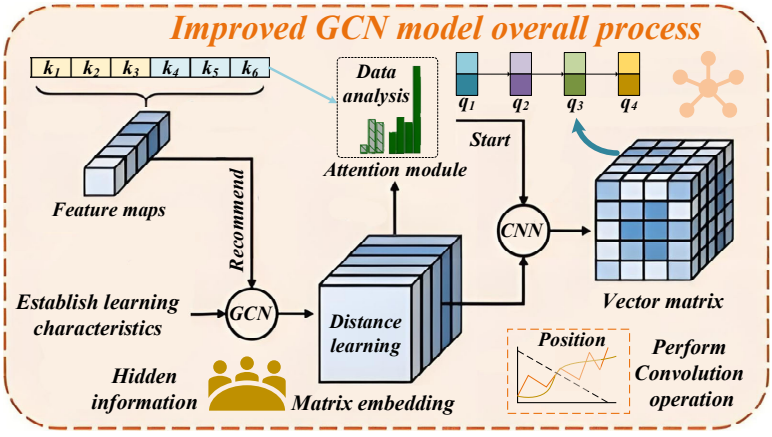
$$h_v' = \sum_{u \in \mathcal{N}(v)} \alpha_{vu} h_u \tag{9}$$

Among them, h_v' represents the feature representation of node v after dependency weighting, α_{vu} represents the strength of dependency relationship between nodes, and h_u represents the feature representation of neighbour node u. The personalisation score formula is shown in (10).

$$P\left(s_t\right) = w_1 \cdot f_1\left(s_t\right) + w_2 \cdot f_2\left(s_t\right) \tag{10}$$

where $P(s_t)$ represents the personalisation score of state s_t, w_1, w_2 represent the weight of the score function, and $f_1(s_t)$, $f_2(s_t)$ represent the functions of different personalisation features. The improved GCN model consists of two modules: the GCN module and the attention module, as shown in Figure 2. Two types of attention blocks are designed to obtain better node feature representations by plotting the global context information of local learning features (Li et al., 2025; Lv et al., 2024). A convolution layer is added at the top of the GCN module, and three filters are used to perform convolution operations to generate feature maps, resulting in a new learning behaviour vector matrix with the same dimensions. This will retain more detailed information without adding additional parameters (Makanda et al., 2025). Then, the new learning node feature matrix is input into two parallel attention modules to better capture the key features and dependencies in students' learning paths and optimise the generation effect of personalised learning paths.

Figure 2 GCN module and attention module (see online version for colours)



The model input is the embedding of the learning behaviour nodes and the node adjacency matrix, and the GCN represents the hidden information in the nodes, indicating that the results are input to the attention module through the pooling layer (Rasti-Meymandi et al., 2025). The attention module consists of two parallel modules: the positional attention module and the relational attention module. Different operations obtain the attention matrices of these two modules. By establishing the correlation between learning features and attention mechanisms and exploring global context information, long-distance learning context information can be adaptively aggregated, thus improving the representation ability of node features, improving the accuracy of personalised learning path recommendation, and optimising the effect of relationship extraction (Reddy and Kumar, 2024). The output formula of the pooling layer is shown in (11).

$$Q_{path} = \frac{\sum_{i=1}^{n} f(s_i)}{L_{path}} \tag{11}$$

Here, Q_{path} represents the quality evaluation of the path, $f(s_i)$ represents the eigenvalue of the state s_i, and L_{path} represents the path length.

This personalised learning path framework has the potential for cross educational scenario promotion due to its universal architecture of RL dynamic adaptation and KG structured knowledge representation. The high school education scene needs to reconstruct the KG to match the high school curriculum and adjust the complexity of knowledge points, while calibrating the RL reward function to adapt to high school goals; Online education platforms need to build domain KGs and customise RL state representations to adapt to diverse learner backgrounds and goals. Its containerisation deployment and LMS integration capabilities further support cross scenario deployment, and in the future, it will explore targeted implementation in high school, vocational education, and MOOC environments.

This study explores the use of trade-offs to introduce randomness through RL, combined with recent learning history in state representation to avoid recommending similar content repeatedly. It also utilises the rich connectivity of KGs to provide multiple effective paths, allowing for different knowledge sequences, resource selection, and focus

adjustment. This ensures personalised, efficient, and diverse paths while maintaining learning objectives, thereby maintaining students' long-term engagement and freshness in learning experiences.

4 Experimental results and analysis

To statistically validate the learning improvement, a pre-test and post-test evaluation design was implemented. Before the experiment commenced, all 300 students undertook a pre-test to establish their baseline knowledge level. Following the three-month intervention, an equivalent post-test was administered. The academic performance improvement was calculated based on the normalised gain from pre-test to post-test, with the 12.5% improvement representing a statistically significant difference ($p < 0.05$) between the experimental and control groups.

Throughout the entire data collection and model training process, strict measures were taken to ensure the privacy and data security of students. All student data, including learning records and behaviour logs, have been anonymised by deleting personal identification information and replacing it with random identifiers to prevent re identification. The research protocol has been approved by the institutional ethics committee and written informed consent has been obtained from all participants. The data is stored on encrypted servers with strict access control, and model training uses aggregation and differential privacy techniques to minimise privacy risks. These measures comply with data protection regulations and ethical guidelines in educational research.

The dataset used in this experiment mainly comes from public data and real teaching scenarios in the field of education, covering students' learning records, behaviour logs, course content, and feedback information, including multi-dimensional data such as learning time, progress, error records, interests, and knowledge point mastery. These data are used to build personalised learning models and generate learning paths matching student characteristics. Throughout the three-month experiment, each structured learning session had a duration of approximately 60 minutes. Student learning progress was tracked continuously via the system's logging capabilities, which recorded granular interaction data such as time-on-task, completion rates for learning activities, and error patterns. Furthermore, knowledge acquisition was formally assessed at the beginning and end of each thematic learning module, supplemented by bi-weekly milestone quizzes to evaluate longer-term retention and path adaptation effectiveness. The software and hardware facilities required for the experiment include high-performance computing platforms, servers equipped with multi-core processors and high-performance GPUs, and large-scale data processing and deep-learning model training. Regarding software, TensorFlow and PyTorch deep learning frameworks combine GNNs and RL algorithms for model training and optimisation, and the Neo4j graph database is used to manage KGs to achieve efficient data access and query operations. The recommendation effect comparison of personalised learning paths is shown in Table 1.

The table shows the effect of students with different subject types on personalised learning path recommendations. There are fewer recommended paths for physics students. Still, the recommendations' accuracy and path overlap are high, indicating that the learning paths of physics students are easier to recommend accurately. Mathematics students have more path recommendations, but the correct recommendation rate and path

coincidence degree are relatively low, which indicates that the learning path of this type of student may be complicated. Computer science students have the largest number of recommended paths, and the accuracy rate is high, indicating that this subject group's personalised learning path system performs well.

Table 1 Comparison of recommendation effects of personalised learning paths

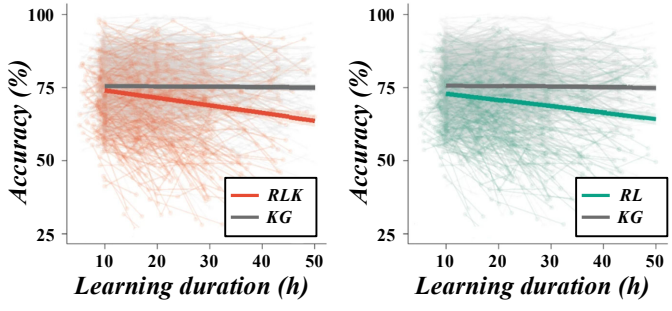
Type of student	*Number of recommended paths*	*Correct recommendation rate*	*Path coincidence degree*
Physics student	12	85%	92%
Students of mathematics	15	80%	87%
Computer science students	18	88%	90%

The personalised learning path generation system is deeply integrated with Moodle LMS through LTI protocol and runs on a server cluster equipped with dual Intel Xeon Silver 4216 processors, 256GB RAM, and 4 NVIDIA Tesla T4 GPUs. It adopts Ubuntu 20.04+Docker containerisation architecture and integrates TensorFlow 2.8, PyTorch 1.12, and Neo4j 4.4 technology stacks. It supports concurrent access by 300 students with a response time of less than 2 seconds, achieving real-time collection of learning behaviour data, model training, and path generation in a closed loop. While ensuring smooth teaching, it verifies its feasibility in real educational scenarios.

Through comparative experiments, rule-based systems have an advantage in initial path generation speed, but they have shortcomings such as insufficient flexibility, manual rule updates, and high maintenance costs for dynamic scenes; Although the RL-KG method has a slightly longer initial time consumption, it can dynamically optimise the path through real-time feedback, reduce manual intervention, and improve long-term efficiency. Moreover, its CPU utilisation increases linearly with user load, demonstrating better scalability and adaptability compared to rule-based systems that experience exponential growth in resource consumption in complex scenarios.

This paper analyses the accuracy rate of personalised learning path recommendations based on different disciplines to show the difference in accuracy rate. The results are shown in Figure 3.

Figure 3 Accuracy of personalised learning path recommendation based on different disciplines (see online version for colours)



It can be seen from the figure that on the left side of the figure, with the increase in learning time, the accuracy rate improves. Still, the increase is small, indicating that

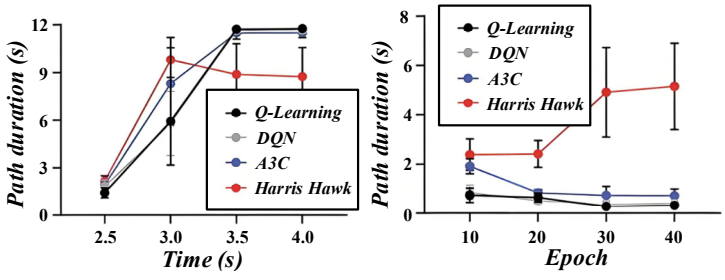
combining RL and KG may be more effective in the initial stage. At the same time, it tends to stabilise when the learning time is longer, and the accuracy rate remains at about 75%. In the chart on the right, the recommendation mechanism using only RL shows a relatively stable trend. With the increased learning time, the accuracy rate also tends to be about 75%, indicating that the mechanism shows high stability in long-term learning. From the overall trend, the RLK model shows a certain gain in the recommendation of learning paths in different disciplines, while using KG alone plays a small role in improving accuracy.

The system customises learning paths for three types of students: those with weak foundations are given priority in strengthening knowledge points such as 'basic algebra', accompanied by step-by-step videos and slow paced tests, resulting in an 18% improvement in post test scores for basic questions; Skip redundant foundations for interest driven individuals and introduce higher-order content and project resources such as 'Physics Application Calculus' to achieve a 20% increase in module completion speed and an accuracy rate of 94%; The integration of collaboration enthusiasts into group discussions and interactive activities, maintaining a moderate difficulty gradient, achieved the highest satisfaction rate of 5.0 and a 15% increase in participation, empirically verifying the precise personalised ability of path generation.

The satisfaction with learning is measured using a standardised five-point Likert scale survey. The survey includes 10 items that evaluate various aspects of the learning experience, including content relevance, adaptability of the learning path, engagement, perceived usefulness, and overall satisfaction with the personalised system. The Cronbach's alpha of the survey is 0.89, indicating high internal consistency.

This paper analyses the influence of RL on the learning path generation time to show the influence of different RL algorithms and verify the efficiency difference of RL methods. The results are shown in Figure 4.

Figure 4 Effect of RL on the duration of learning path generation (see online version for colours)

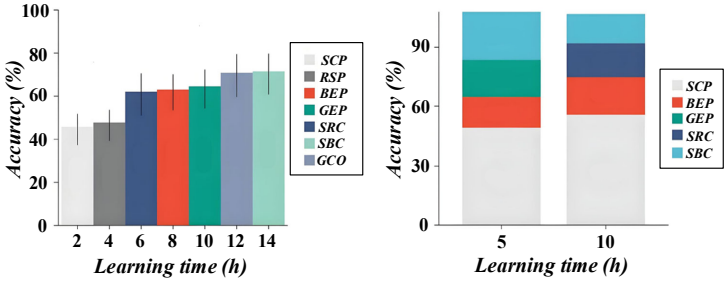


The data in the graph shows that when the learning time is between 2.5 seconds and 4.0 seconds, the Q-Learning algorithm shows a clear upward trend in the duration of the generated path, gradually increasing from nearly 3 seconds to about 9 seconds. This shows that the Q-Learning algorithm is greatly affected by the time change in the generation of the learning path. In contrast, the generation duration of the DQN and A3C algorithms varies less throughout the period. It always remains in the lower range, indicating that these two algorithms are more stable in learning path generation and less affected by time delay. Harris Hawk algorithm shows great fluctuations, especially in the period close to 4 seconds, and the path generation time almost reaches the maximum

value, surpassing other algorithms, reaching the highest of about 12 seconds, showing high delay and instability.

To show the change in learning path recommendation accuracy before and after applying KG optimisation and verify the optimisation effect of KG on path generation, this paper compares the change of recommendation accuracy before and after personalised path optimisation based on KG and the results are shown in Figure 5.

Figure 5 Changes in recommendation accuracy before and after personalised path optimisation based on KG (see online version for colours)



As can be seen from the figure, SCP stands for discipline map, RSP stands for resource map, BEP stands for behaviour map, GEP stands for synthetic map, SRC stands for discipline-resource, SBC stands for discipline-behaviour, and GCO stands for integrated optimisation. In the left chart, SCP and GEO have lower accuracy at shorter learning times, but with increasing learning time, the accuracy improves significantly, eventually reaching about 80%. However, BEP and RSP maintain a high accuracy rate in the whole learning process; the highest can reach more than 70%. In the stacked bar chart on the right, the combined performance of the five path recommendation methods is also shown, in which the combination of SCP and BEP occupies a larger proportion, showing the comprehensive performance of these methods under different learning times. Overall, the recommendation accuracy before and after path optimisation has been significantly improved, especially under the application of methods such as SCP and GEP.

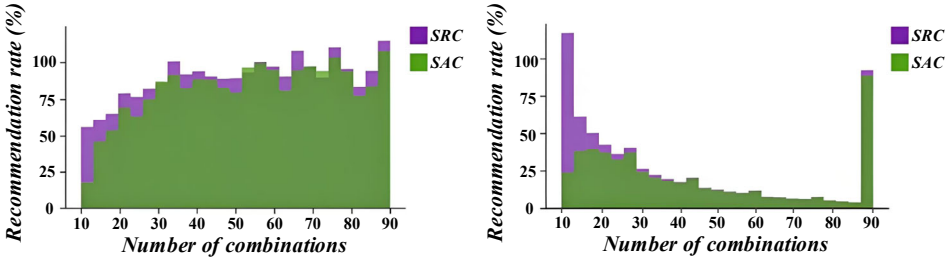
Table 2 Analysis of learning path generation time for RL optimisation

Optimisation algorithm	*Average generation time (seconds)*	*Minimum build time (seconds)*	*Maximum build time (seconds)*
Q-learning	3.2	2.5	4.1
DQN	5.1	4.0	6.5
A3C	4.3	3.2	5.8

Table 2 analyses the learning path generation time of RL optimisation and shows the temporal performance of the three RL algorithms when generating the learning path. The Q-learning algorithm has the shortest average generation time and small time fluctuation, suitable for quick response scenarios. The DQN algorithm takes the longest to generate, probably because it requires more computing resources and its complex deep neural network structure. The A3C algorithm is somewhere in between and is suitable for balancing time and accuracy.

This paper analyses the number and accuracy of path recommendations with different parameter combinations to analyse the influence of different parameter combinations on personalised learning path recommendations and evaluate which parameter combinations can bring the best path recommendation effect. The analysis results are shown in Figure 6.

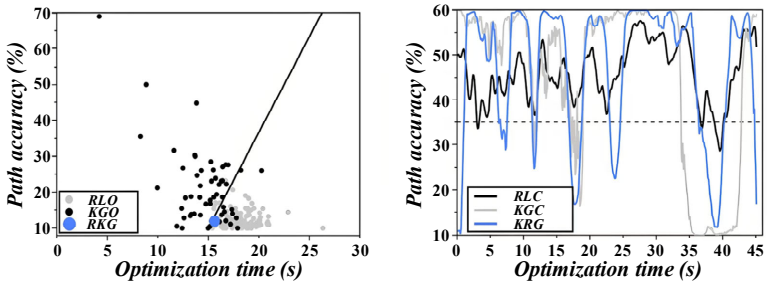
Figure 6 Influence of different parameter combinations on path recommendation number and accuracy (see online version for colours)



The data in the figure shows that the recommendation accuracy rate of SRC is relatively stable, about 75%. With the increase in combinations, the recommendation accuracy rate gradually increases, approaching 100%. In contrast, the accuracy of the SAC method is significantly higher, basically maintaining above 85%, especially when the number of combinations is greater than 60; the SAC method shows a relatively robust recommendation effect. The chart on the right shows that under a small number of combinations, the accuracy rate of SRC fluctuates greatly, and the recommendation accuracy rate fluctuates greatly, even dropping to about 50%. At the same time, SAC still maintains a high accuracy rate in this case, which is stable at more than 70%. Only when the number of combinations exceeds 70 the accuracy rate quickly jumps to nearly 100%. This shows that the SAC method has strong stability and high efficiency in the path recommendation task, especially for more parameter combinations.

This paper compares the recommendation path generation accuracy based on RL and KG optimisation to verify the recommendation effect of the combination. The results are shown in Figure 7.

Figure 7 Comparison of recommended path generation accuracy based on RL and KG optimisation (see online version for colours)

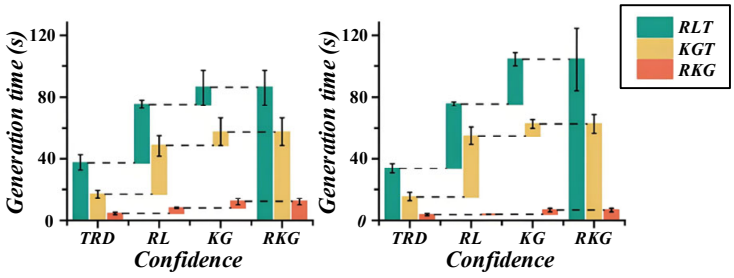


The chart shows that in the scatter plot on the left, RLO shows lower accuracy with shorter optimisation time, and the path accuracy mostly remains below 20%, while KGO

and KRG show higher accuracy; with increasing optimisation time, the accuracy gradually improves, and approaches 70%. This shows that the optimisation method based on RL and KG can effectively improve the accuracy of path recommendation, especially under the KRG method; with the extension of time, the accuracy rate reaches a high level.

This paper compares the generation time of personalised learning paths for different learning styles to test the influence of different learning styles. The results are shown in Figure 8.

Figure 8 Personalised learning path generation time based on different learning styles (see online version for colours)



It can be seen from the chart that there are significant differences in the generation time of personalised learning paths under different learning styles. The generation time of TRD is longer at all confidence levels, especially in the low confidence case, where the generation time is close to 90 seconds. However, the generation time of the RL method is generally short, about 40 seconds at a low confidence level. With the increase in confidence, the time increases, but it is still lower than TRD. The generation time of the KG method at moderate and high confidence levels is also relatively fast, usually around 50 seconds. The RKG method exhibits the shortest generation time at all confidence levels, especially under high confidence conditions; the generation time is less than 10 seconds. Overall, the RKG method has significant advantages in improving the generation efficiency, especially at high confidence levels, which can significantly reduce the generation time of learning paths.

Table 3 Impact of KG optimisation on learning path

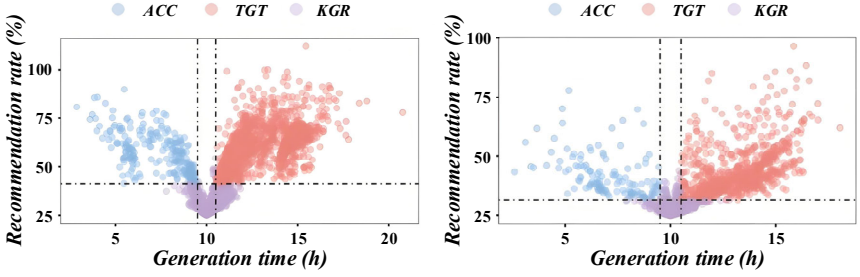
KG types	*Path accuracy before optimisation*	*Path accuracy after optimisation*	*Promotion rate (%)*
Subject knowledge map	72%	85%	18.06%
Learning resource KG	65%	78%	20.00%
Behavioural data KG	80%	90%	12.50%

The impact of KG optimisation on the learning path is shown in Table 3. The table shows the effects of different types of KGs before and after personalised learning path optimisation. After optimising the subject KG, the path accuracy rate increased by 18.06%, indicating that this KG type greatly affects the subject recommendation. The learning resource KG has the highest improvement rate, reaching 20%, which indicates that this graph type helps optimise resource selection and recommendation of learning

paths. The improvement rate of the behavioural data KG is relatively low, but it still shows effectiveness after optimisation.

This paper analyses the impact of the KG on the accuracy of personalised path recommendations in a multidisciplinary context to analyse how it can improve this accuracy. See Figure 9 for specific results.

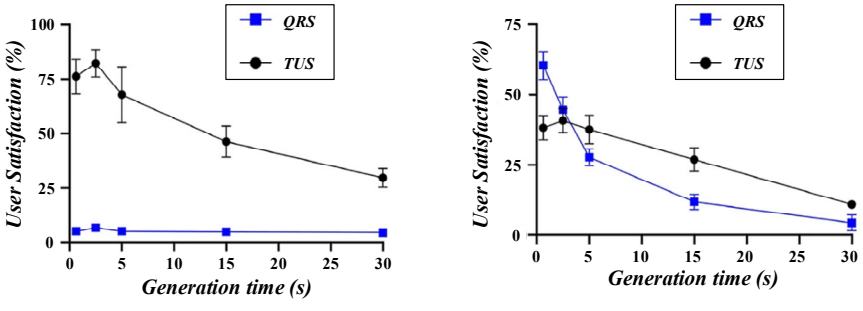
Figure 9 Impact of KG on path recommendation accuracy in multidisciplinary context (see online version for colours)



It can be seen from the figure that in the left chart, the recommendation accuracy of the ACC method is maintained at around 50% when the generation time is short. Still, with the increase in generation time, the recommendation accuracy is significantly improved, up to 75%. In contrast, the recommended accuracy of the TGT method is higher, above 60%, and with the increase of generation time, the accuracy gradually approaches 80%. The KGR method has the highest recommendation accuracy rate, and the accuracy rate exceeds 75% in a short generation time, and the accuracy rate is close to 90% in 10 hours. The chart on the right shows the changing trends in different discipline backgrounds. Overall, the KGR method is superior in improving the recommendation accuracy, especially in the case of complex discipline backgrounds, which can significantly improve the recommendation accuracy and achieve higher accuracy quickly.

This paper compares the relationship between user satisfaction generated by learning path and recommendation quality to verify the impact of path recommendation accuracy on user experience. The results are shown in Figure 10.

Figure 10 Relationship between user satisfaction and recommendation quality generated by learning path (see online version for colours)



In the left chart, the QRS method achieves user satisfaction close to 75% when the generation time is short. Still, with the increase of generation time, the satisfaction

decreases rapidly and eventually approaches 25%. In contrast, the TUS method maintained user satisfaction at around 50% in all generation periods and changed little after the generation time exceeded 15 seconds, maintaining between 40% and 50%. The chart on the right further shows the impact of recommendation quality. The QRS method's user satisfaction still shows a rapid decline trend with the extension of generation time, but within 5 seconds, the satisfaction remains high. The TUS method is relatively stable. The longer the generation time, the slower the user satisfaction decreases, and finally remains at a level close to 40%. Overall, the QRS method provides higher user satisfaction in a short generation time. Still, the satisfaction decreases rapidly with the increase of time, while the TUS method maintains relatively stable user satisfaction for a longer time.

5 Conclusions

The personalised learning path generation mechanism of RL and KG proposed in this paper shows significant advantages in improving learning effects and optimising learning processes. The following is a detailed data analysis of the main conclusions of this paper:

1 Improvement of academic performance: in the experimental group of 300 students, the learning path generated by this method improved academic performance by 12.5% compared with the traditional method. This improvement was measured using course-specific assessments, comprising identical pre-test and post-test examinations for both groups. The post-test scores were normalised against the pre-test baselines, and the reported 12.5% represents the average gain difference between the experimental and control groups. This result shows that combining RL and KG's personalised path-generation mechanism can significantly improve students' learning effects. Especially in terms of in-depth mastery of subject knowledge and reasonable regulation of learning progress, the combination of RL and KG shows high adaptability. It can be dynamically optimised according to students' real-time feedback, thus improving the learning effect to a greater extent.

2 Improvement of learning satisfaction: the average satisfaction score of students in the experimental group is 4.7, 23.7% higher than the traditional method score of 3.8. This significant improvement reflects students' high recognition of the personalised learning path generation mechanism, indicating that the learning paths generated through this mechanism can better meet students' learning needs but also enhance students' learning motivation and interest. The combination of RL and KG can accurately match students' interests, preferences, and knowledge mastery, making the learning experience more personalised and efficient.

3 Improvement of learning efficiency: the experimental data show that the knowledge mastery speed of students in the experimental group is increased by 15%, and the average learning time of each student is reduced by 8% compared with the traditional method. This result proves that RL and KG's personalised learning path generation mechanism can improve students' learning effect, optimise the learning process, and improve learning efficiency. By adjusting the learning path intelligently, students can complete more learning tasks quickly to master what they have learned more efficiently.

The personalised learning path generation mechanism proposed in this paper, which integrates RL and KG, shows significant advantages in improving learning effect, enhancing student satisfaction, and improving learning efficiency. This is achieved through the dynamic adjustment of a globally pre-trained RL model, which is fine-tuned in real-time based on individual student parameters and interactions, combined with the structured support of the KG.

Future work will focus on improving the scalability and efficiency of models to handle larger datasets, expanding KGs to cover more interdisciplinary topics, and enhancing system interoperability with a broader ecosystem of educational platforms. In addition, exploring more complex reward functions and transfer learning techniques to improve the adaptability and generalisation ability of models remains a key direction.

The personalised learning path generation mechanism proposed in this paper, which integrates RL and KG, shows significant advantages in improving learning effect, enhancing student satisfaction, and improving learning efficiency through dynamic adjustment of RL and structured support of KG. This research not only provides a new solution for personalised education but also lays a solid foundation for the development of intelligent education.

Declarations

All data generated or analysed during the study are available from the corresponding author by request.

The authors declare that they have no conflicts of interest.

References

Afzali, A. and Shamsinejadbabaki, P. (2025) 'PHiFL-TL: personalized hierarchical federated learning using transfer learning', *Future Generation Computer Systems*, Vol. 166, p.107672.

An, D., Yang, Y., Gao, X., Qi, H., Yang, Y., Ye, X., Li, M. and Zhao, Q. (2025) 'Reinforcement learning-based secure training for adversarial defense in graph neural networks', *Neurocomputing*, Vol. 630, p.129704.

Aritonang, P.K., Wiryono, S.K. and Faturohman, T. (2025) 'Hidden-layer configurations in reinforcement learning models for stock portfolio optimization', *Intelligent Systems with Applications*, Vol. 25, p.200467.

Buciuman, C-F. and Potra, S. (2025) 'Revolutionizing education in Industry 4.0: eye-tracking and AI for personalized learning', *Procedia Computer Science*, Vol. 253, pp.1658–1667.

Chen, D., Yu, P., Wang, G., Liu, X., Ding, Y. and Jin, J. (2025a) 'Design of a hybrid-mode piezoelectric actuator for compact robotic finger based on deep reinforcement learning', *Mechanical Systems and Signal Processing*, Vol. 227, p.112401.

Chen, Z., Zhang, Y., Fang, Y., Geng, Y., Guo, L., Chen, J., Liu, X., Pan, J.Z., Zhang, N., Chen, H. and Zhang, W. (2025b) 'Knowledge graphs for multi-modal learning: Survey and perspective', *Information Fusion*, Vol. 121, p.103124.

Dai, G., Tang, J., Zeng, J., Hu, C. and Zhao, C. (2024) 'Road network traffic flow prediction: a personalized federated learning method based on client reputation', *Computers and Electrical Engineering*, Vol. 120, p.109678.

Ge, H., Pokhrel, S.R., Liu, Z., Wang, J. and Li, G. (2024) 'PFL-DKD: modeling decoupled knowledge fusion with distillation for improving personalized federated learning', *Computer Networks*, Vol. 254, p.110758.

Ghadiri, H. and Hajizadeh, E. (2025) 'Designing a cryptocurrency trading system with deep reinforcement learning utilizing LSTM neural networks and XGBoost feature selection', *Applied Soft Computing*, Vol. 175, p.113029.

Li, H., Wang, X., Cao, P., Li, Y., Yi, B. and Huang, M. (2025) 'FedCPG: a class prototype guided personalized lightweight federated learning framework for cross-factory fault detection', *Computers in Industry*, Vol. 164, p.104180.

Lu, L., Si, G., Liang, X., Li, M. and Zhou, F. (2025) 'A survey on dynamic scene understanding using temporal knowledge graphs: From scene knowledge representation to extrapolation', *Neurocomputing*, Vol. 635, p.129854.

Lv, F., Qian, P., Lu, Y. and Wang, H. (2024) 'Personalized federated learning on long-tailed data via knowledge distillation and generated features', *Pattern Recognition Letters*, Vol. 186, p.178–183.

Makanda, I.L. D., Jiang, P. and Yang, M. (2025) 'Personalized federated unsupervised learning for nozzle condition monitoring using vibration sensors in additive manufacturing', *Robotics and Computer-Integrated Manufacturing*, Vol. 93, p.102940.

Murray, L., Castillo, T., Diego, I.M. de, Weber, R., González-Olabarria, J.R., García-Gonzalo, J., Weintraub, A. and Carrasco-Barra, J. (2025) 'Deep reinforcement learning for optimal firebreak placement in forest fire prevention', *Applied Soft Computing*, Vol. 175, p.113043.

Nguyen, X-B., Phan, X-H. and Piccardi, M. (2025) 'Fine-tuning text-to-SQL models with reinforcement-learning training objectives', *Natural Language Processing Journal*, Vol. 10, p.100135.

Nie, M., Chen, D., Chen, H. and Wang, D. (2025) 'Automtnas: automated meta-reinforcement learning on graph tokenization for graph neural architecture search', *Knowledge-Based Systems*, Vol. 310, p.113023.

Park, G., Jung, W., Han, S., Choi, S. and Sung, Y. (2025) 'Adaptive multi-model fusion learning for sparse-reward reinforcement learning', *Neurocomputing*, Vol. 633, p.129748.

Rasti-Meymandi, A., Sajedi, A. and Plataniotis, K.N. (2025) 'FedPnP: personalized graph-structured federated learning', *Pattern Recognition*, Vol. 163, p.111455.

Reddy, B.R.R. and Kumar, R.L. (2024) 'A fusion model for personalized adaptive multi-product recommendation system using transfer learning and Bi-GRU', *Computers, Materials and Continua*, Vol. 81, No. 3, pp.4081–4107.

Sun, C., Yang, J., Cao, Z., Yang, Z., Yang, Y. and Shu, J. (2025a) 'Fast convergent actor-critic reinforcement learning based interference coordination algorithm in D2D networks', *Ad Hoc Networks*, Vol. 171, p.103788.

Sun, L., Ding, A. and Ma, H. (2025b) 'Multi-agent reinforcement learning system framework based on topological networks in Fourier space', *Applied Soft Computing*, Vol. 174, p.112986.

Valko, D. and Kudenko, D. (2025) 'Hybrid pathfinding optimization for the lightning network with reinforcement learning', *Engineering Applications of Artificial Intelligence*, Vol. 146, p.110225.

Wang, J., Liang, Q., Li, M., Qu, Z. and Zhang, Y. (2025a) 'Dynamic scheduling in flexible and hybrid disassembly systems with manual and automated workstations using reward-shaping enhanced reinforcement learning', *Engineering Applications of Artificial Intelligence*, Vol. 150, p.110588.

Wang, J., Yan, Y., Hu, Y., Yang, X. and Zhang, L. (2025b) 'A transfer reinforcement learning and digital-twin based task allocation method for human-robot collaboration assembly', *Engineering Applications of Artificial Intelligence*, Vol. 144, p.110064.

Wang, S., Wang, R., Liu, Y., Zhang, Y. and Hao, L. (2025c) 'Dynamic modeling and control of pneumatic artificial muscles via deep Lagrangian networks and reinforcement learning', *Engineering Applications of Artificial Intelligence*, Vol. 148, p.110406.

Ye, Z., Qiu, D., Li, S., Fan, Z. and Strbac, G. (2025) 'Federated reinforcement learning for decentralized peer-to-peer energy trading', *Energy and AI*, Vol. 20, p.100500.

Yu, J., Zhang, Y. and Sun, C. (2025) 'End-to-end multi-task reinforcement learning-based UAV swarm communication attack detection and area coverage', *Knowledge-Based Systems*, Vol. 316, p.113390.

Zhang, S., Guan, Z., Wang, X., Tan, P. and Jiang, H. (2025) 'Reinforcement learning-based automatic block decomposition of solid models for hexahedral meshing', *Computer-Aided Design*, Vol. 182, p.103850.

Zhang, Z. and Li, R. (2025) 'Q-value-based experience replay in reinforcement learning', *Knowledge-Based Systems*, Vol. 315, p.113296.

Zhou, X., Guan, X., Sun, D., Zhang, X., Zhang, Z. and Ohtsuki, T. (2025) 'Heterogeneous multi-agent deep reinforcement learning based low carbon emission task offloading in mobile edge computing', *Computer Communications*, Vol. 234, p.108089.