ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on dynamic assessment model of educational legal Risk based on reinforcement learning

Jianfeng Zhang[1,*]

1 School of International Exchange, Gansu University of political science and law, Lanzhou, Gansu, 730070, China

SUMMARY: *Effective education legal risk governance needs a dynamic assessment model to transform heterogeneous compliance evidence into traceable and reviewable early warning decisions. This paper proposes a reinforcement learning model ERL-Risk for educational legal risk assessment. The model encodes policy terms, complaint records, teacher behavior events, contract texts, platform access logs and historical disposal results into continuous risk states, and uses PPO Actor-Critic agent to select early warning actions and review priorities. This paper constructed a data set containing 8,640 anonymous risk events, 42,300 behavior logs, and 3,120 policies and contract terms from simulated school governance records and compliance cases annotated by experts. ERL-Risk was compared with rule matching, XGBoost, BiLSTM and Transformer classifiers under the same feature pool. The proposed model achieves 93.4% risk level accuracy, 91.8% Macro-F1, 94.6% high risk recall, and controls the average warning delay at 0.41 seconds. The research results provide a more robust strategy update and reliable support for educational decision-making in the digital campus compliance scenario.*

KEYWORDS: *Reinforcement learning; Educational legal risk; Dynamic evaluation; Risk warning*

1 Introduction

Educational legal risk presents multi-source interweaving and continuous evolution in digital campus. Student rights, teacher behavior, contract performance, data authorization, learning records, and complaint disposition, all may form computable risk signals. Traditional review relies on experience and fixed rules, which is difficult to track event diffusion, subject association and disposal feedback. Reinforcement learning transforms risk identification into a process of state perception, action selection, reward feedback and strategy update, which can modify the evaluation strategy in dynamic scenarios. Franceschelli and Musolesi studied the policy optimization and reward modeling of reinforcement learning in generative artificial intelligence, which provided the algorithm foundation [1]. Fritz-Morgenthal et al. proposed a trusted artificial intelligence risk management framework, emphasizing that model outputs retain audit paths [2]. Memarian and Doleck sort out the issues of fairness, accountability and ethics in artificial intelligence in higher education, and provide reference for the setting of responsibility boundaries [3].

Abnormal access, semantic change of complaints or deviation of contract terms need to be judged jointly by combining historical records, subject identity and behavior frequency. Ismail

*zhanglaoshi202506@163.com
https://doi.org/10.65102/is2026995
Published: 30 April, 2026

and Yusof summarized machine learning techniques in student flow prediction, showing that educational data can be used for classification and early warning by feature engineering [4]. Marx et al. studied the cognitive model and attitude of middle school students toward artificial intelligence, suggesting that the system needs to consider the use of subject differences [5]. Farhi et al. analyzed students' views on the ethics of ChatGPT usage and pointed out that data compliance, content responsibility and usage boundaries would enter the risk chain after generative tools entered the learning process [6]. Educational legal risk assessment needs to incorporate rule clauses, behavior logs and feedback results into the computational framework.

This paper constructs the ERL-Risk model for educational legal risk assessment. The model performs entity recognition, clause matching, time alignment and semantic encoding on policy clauses, complaint texts, teacher behavior events, contract agreements, platform logs and historical disposal results, and maps multi-source information into state vectors. Ocak et al. proposed the pattern recognition method enhanced by artificial intelligence, which reflected the feasibility of modeling the spatio-temporal characteristics of educational events [7]. Rodway and Schepman studied the influence of artificial intelligence education technology adoption on curriculum satisfaction expectation, indicating that the model output would affect the judgment of technical responsibility [8]. In this paper, risk subject, event type, clause correlation, behavior frequency and disposal result are taken as state variables, and low risk marking, medium risk review, high risk warning, manual audit and disposal sequencing are taken as action variables.

The existing methods of educational legal risk assessment mostly rely on manual review, rule matching or static classification models, and the time accumulation, subject association and disposal feedback of risk events are insufficiently utilized. Some models can complete text classification or risk label prediction, but it is difficult to update the judgment strategy after new events continue to enter the system, and it is difficult to explain the computational relationship between early warning actions and historical disposal results. Khosravi et al. proposed the mechanism that learners, educators and machines jointly participate in content generation, indicating that educational data has the characteristics of multi-agent collaborative generation [9]. Priyambada and Usagawa proposed a two-layer ensemble prediction model integrating learning behaviors and domain knowledge, which provided a method reference for the joint input of rule constraints, behavior sequences and risk labels [10]. Based on this, this paper builds an ERL-Risk dynamic evaluation model and compares it with rule matching, XGBoost, BiLSTM and Transformer. Further, the effects of state coding, action space, reward function and strategy update module on risk level recognition, high risk early warning and response delay are tested through ablation experiments. Make education legal risk assessment form a trainable, traceable and updatable computing link.

2 Literature Review

Educational AI research has shifted from learning support to data-driven interaction modeling, content generation, and behavior prediction. For the dynamic assessment of educational legal risk, although the existing results do not directly deal with compliance events, they provide a computational foundation in dialogue interaction, text generation, behavior recognition, automatic scoring and artificial intelligence literacy measurement. The risk object concerned in this paper is not ordinary teaching evaluation, but the risk state formed by platform use, data processing, content generation, teacher-student interaction and contract execution, which needs to be sorted out from the perspectives of data coding, sequence modeling and feedback

update.

Annamalai et al. studied the application of chatbots for English learning in higher education, and showed that the dialogue system can record learner input, system feedback and interaction trajectory, and such time series data can provide modeling entry for data authorization, content liability and platform reply boundary in educational legal risk [11]. Sanusi et al. proposed a teaching design for machine learning understanding of middle school students in African schools, emphasizing that artificial intelligence learning needs to combine age, task and classroom situation, and their research provides reference for the setting of subject differences in risk assessment models [12]. Chomphooyod et al. proposed an English grammar multiple choice question generation method based on Text-to-Text Transfer Transformer, and proved that Transformer can transform educational texts into structured task outputs. This idea can be transferred to the semantic parsing of contract clauses, complaint texts and compliance instructions [13]. Abichandani et al. studied learning benefits and attitude coding in artificial intelligence and computer vision education, and showed that visual tasks, behavioral performance and evaluation data can enter the machine learning framework, which has enlightenment for multimodal risk event recognition [14]. Bezirhan and von Davier proposed to use the OpenAI large language model to automatically generate reading materials, indicating that the large model can complete the generation and difficulty control of educational texts, and also suggesting that the traceability, copyright boundary and applicable object of generated content need to enter the risk assessment process [15].

In order to further sort out the relationship between the above research and the dynamic assessment of educational legal risk, Table 1 summarizes four aspects: calculation method, main contribution, modeling relationship with this paper and remaining deficiencies. It can be seen that existing research provides a technical basis for text parsing, behavior recognition, interaction recording and automatic scoring in educational scenarios, but most methods still stay at the level of single task processing, and a continuous decision-making framework for legal risk state transfer, disposal action selection and feedback reward update has not been formed.

Table 1: Modeling relationship between related research and this paper

Reference	Computational Method	Main Contribution	Relation to This Study	Remaining Limitation
[11]	Chatbot	Records dialogue and feedback	Supports interactive risk modeling	Lacks legal risk labels
[12]	Machine learning instructional design	Characterizes cognitive differences	Supports state feature construction	Does not involve dynamic assessment
[13]	T5 text generation	Generates structured items	Supports clause semantic parsing	Lacks disposal feedback
[14]	Visual recognition and evaluation coding	Integrates behavior and evaluation data	Supports multimodal event encoding	Insufficient action space
[15]	Large language model generation	Controls text quality	Supports content compliance review	Lacks a reinforcement learning closed loop

Kajiwara et al. proposed a machine learning role-playing game to help K-12 stage students understand artificial intelligence with instructional design, indicating that task roles,

operation paths and feedback results in the education system can be designed as continuous decision-making processes, which is consistent with the state transition logic in reinforcement learning [16]. El Bahri et al. used convolutional neural network to identify learners' personality based on five-factor model, indicating that deep model can extract implicit individual features from learning behavior, but personality recognition is not directly equivalent to compliance judgment, and it still needs to be associated with terms, events and responsible subjects [17]. Hornberger et al. developed and verified the artificial intelligence literacy test for college students, which reflects the quantitative idea of artificial intelligence ability measurement in educational scenarios, and also provides index inspiration for the risk understanding ability and review consistency rate in this paper [18]. Ericsson and Johansson studied the long-term experience of young students practicing oral English with conversational AI, and showed that the continuous use of interactive systems will lead to behavioral adaptation and trust changes, which will affect the responsibility boundary and risk level judgment of the platform [19]. Andersen et al. proposed an automatic proficiency scoring method for early writing, transforming students' writing performance into computable scores, proving that the evaluation of educational texts can be automatically processed, and providing reference for the scoring of complaint statements and compliance reports [20].

In summary, the research of educational artificial intelligence covers the direction of dialogue system, text generation, visual recognition, individual feature extraction and automatic scoring, which provides a basis for data representation and model training for the dynamic assessment of educational legal risk. However, most studies still focus on learning effects or educational experience, and rarely incorporate legal provisions, behavior logs, disposal results and feedback signals into the same decision-making link. In this paper, the ERL-Risk model is constructed, and the educational legal risk is expressed as a continuous state sequence. The risk level identification, early warning and module contribution analysis are realized through the action space, reward function and strategy update mechanism, which makes the related research advance from static education data analysis to a trainable and traceable compliance evaluation process.

3 Construction of dynamic assessment model of educational legal risk

3.1 Data coding and feature construction of educational legal risk events

Educational legal risk events are composed of system texts, complaint records, contract terms, platform logs, teacher behavior records and historical disposal results. To make the risk material enter the reinforcement learning process, this section transforms the original records into structured event units and establishes a mapping relationship between event number, subject identity, timestamp, source system, evidence fragment and disposal result. The text field undergoes word segmentation, denoising, entity recognition and clause matching, and the log field undergoes time alignment, missing repair and anomaly filtering. Finally, the risk feature matrix that can be read by the policy network is formed.

In order to uniformly describe the text, subject, time, source, evidence and disposal attributes of educational legal risk events, the event semantic vector is constructed, as shown in the following equation:

$$e_i = \mathrm{LN}(W_e[B(t_i) \oplus P(u_i) \oplus \phi(\tau_i) \oplus S(o_i) \oplus D(r_i)] + b_e) \tag{1}$$

Here, e_i represents the semantic vector of the i risk event. $B(t_i)$ denotes the event text

encoding. $P(u_i)$ represents the subject identity embedding; Let $\phi(\tau_i)$ denote the temporal location encoding; $S(o_i)$ represents the source system code and $D(r_i)$ represents the historical disposal result code. W_e and b_e are trainable parameters. LN denotes the layer normalization. This formula compreszes the scattered evidence into the machine representation of the same scale, and reduces the influence of manual label differences on subsequent judgments.

Fig. 1 shows the processing path of educational legal risk events from raw data into the feature matrix. The process starts from multi-source data access, and after event cleaning, subject identification, clause mapping, behavior aggregation and feature fusion, the risk event vector is output for subsequent state space modeling.

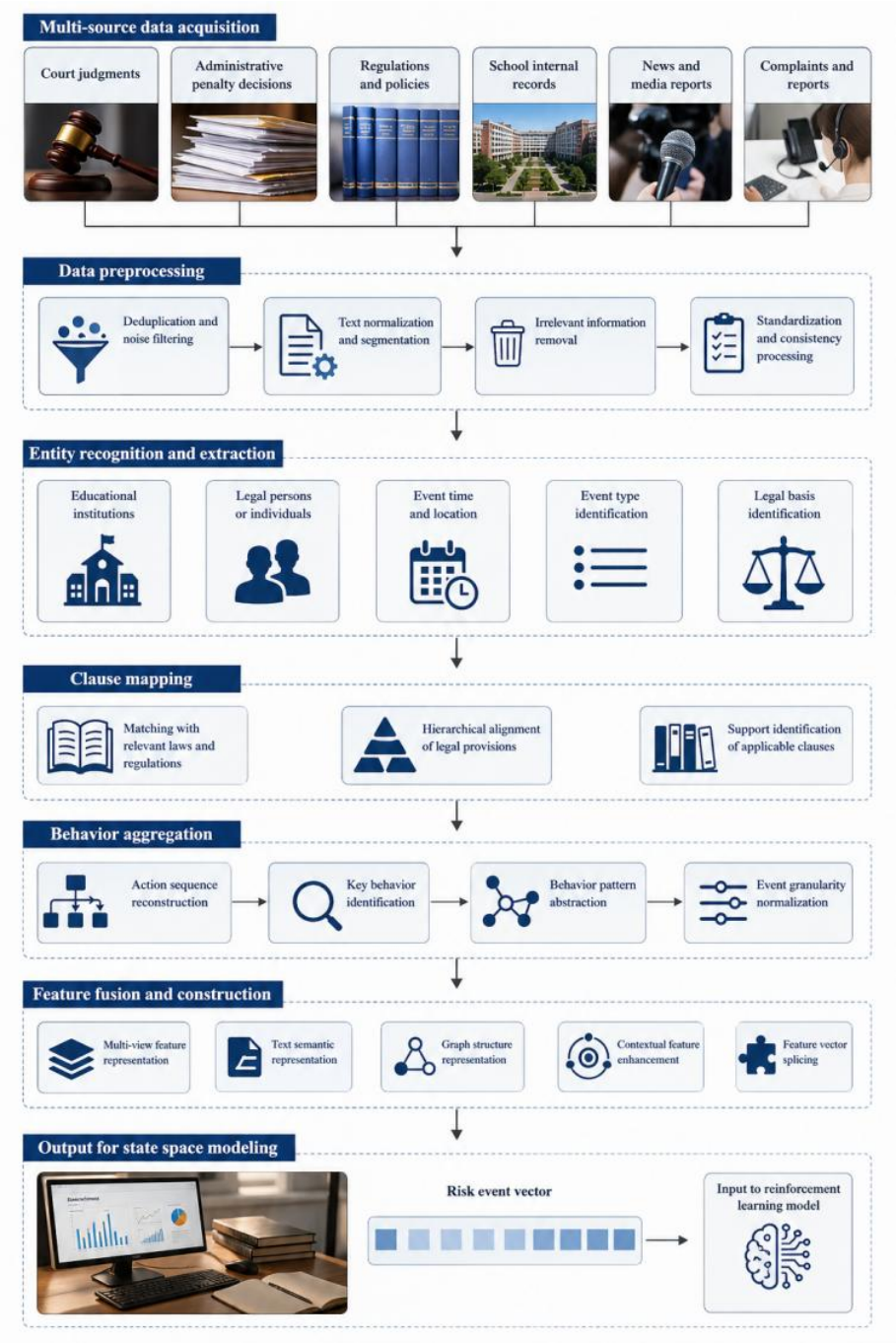

Figure 1: Data coding and feature construction process of educational legal risk events

In order to describe the semantic fit degree, constraint strength and evidence relationship between risk events and legal provisions, the association weight is introduced, as shown in the following equation:

$$a_{ij} = \frac{\exp(\,(W_q e_i)^T (W_k l_j)/\sqrt{d} + \beta m_{ij})}{\sum_{j=1}^{M} \exp(\,(W_q e_i)^T (W_k l_j)/\sqrt{d} + \beta m_{ij})} \tag{2}$$

Here, a_{ij} represents the association weight between event i and clause j; l_j denotes the clause vector; m_{ij} means artificial rule or keyword hit mark; β is the rule compensation coefficient; d denotes the hidden layer dimension. This formula makes text semantics and explicit legal clues participate in matching together, avoiding the model only relying on similar words and ignoring constraint relations.

In order to compress the access, commit, modify, appeal, withdraw and feedback traces in the platform behavior log, the behavior aggregation representation is constructed, as shown in the following equation:

$$b_i = \sum_{k=1}^{K} \alpha_{ik} \log(1 + n_{ik}) v_k, \quad \alpha_{ik} = \frac{\exp(q_i^T v_k)}{\sum_{k=1}^{K} \exp(q_i^T v_k)} \tag{3}$$

Here, b_i represents event correlation behavior vector; v_k represents the k action embedding; n_{ik} is the occurrence frequency of the behavior. α_{ik} is the action weight; q_i denotes the event query vector. This formula retains high-frequency behaviors and also highlights a small number of operation trajectories with strong risk pointing.

In order to form a unified feature that can be input into the reinforcement learning model, the semantic, clause, action and disposition codes are unified and fused as shown in the following equation:

$$x_i = \mathrm{Norm}(W_x[e_i \oplus \sum_{j=1}^{M} a_{ij} l_j \oplus b_i \oplus c_i] + b_x) \tag{4}$$

where x_i represents the final risk event characteristics; c_i stands for disposal context encoding; W_x and b_x denote the parameters of the fusion layer; Norm stands for the normalization operation. This formula completes the transformation from the original risk material to the feature matrix and provides a stable input for the state space modeling below.

The coding layer also retains the tags of event source, completeness of evidence chain and disposal stage, which facilitates the subsequent model to distinguish between ordinary compliance reminders, potential disputes and high-level legal risks. After the above processing, texts, logs, and rules no longer enter the model in isolated form, but are organized as computational samples with temporal order, subject relations, and clause constraints.

3.2 Risk state space Modeling based on reinforcement learning

Upon completion of event encoding, educational legal risks need to be represented as states in reinforcement learning. The state contains not only the current event characteristics, but also the historical disposition, the continuous behavior of the subject, the strength of the clause constraint and the feedback change. If only a single classification input is used, the model is difficult to identify the process of risk evolution from low level warning to high level warning. This section defines an educational legal risk state as a combination of observable features, implicit memory, and disposition context that enables the policy network to select an audit,

warning, or ordering action based on a continuous state.

In order to describe the observable state of educational legal risk at successive moments, a state vector containing subject, evidence and feedback is constructed, as shown in the following equation:

$$s_t = \mathrm{Concat}(x_t, p_t, g_t, z_t, r_{t-1}) \tag{5}$$

where s_t represents the risk state at time t; x_t represents the current event feature. p_t represents the subject portrait coding; g_t denotes clause constraint graph embedding; z_t represents the completeness of evidence chain; r_{t-1} represents the last round of disposal feedback. This formula puts the event itself in the same state as the external compliance context, so that the model can read the context of risk formation.

In order to retain the implied impact, cumulative effect and disposal trace of risk events over time, the gated state update relationship is designed, as shown in the following equation:

$$h_t = \lambda_t \odot \tanh(W_h[s_t, h_{t-1}] + b_h) + (1 - \lambda_t) \odot h_{t-1} \tag{6}$$

where h_t stands for implicit risk memory; Let λ_t denote the update gate; h_{t-1} represents the previous time memory; W_h and b_h denote the parameters; $\odot$ indicates element-wise multiplication. This formula enables the model to retain cumulative signals such as continuous complaints, repeated visits and multiple disposal failures.

Fig. 2 shows the formation process of the risk state space. After the risk event feature entered the timing buffer, it was co-encoded with the historical state, the subject portrait, the clause context and the feedback label, and then the current state was generated by the state update unit for the PPO policy network and the value network to call.

Figure 2: Reinforcement learning-based state space modeling process for educational legal risk

In order to calculate the impact degree of the disposal action on the next risk state, the transition probability formula with context and responsibility constraints is established, as shown in the following equation:

$$P(s_{t+1}|s_t, a_t) = \mathrm{Softmax}(W_p \tanh(W_s s_t + W_a a_t + W_c c_t)) \tag{7}$$

where a_t represents the disposal action; c_t represents responsibility and scene constraints; $P(s_{t+1}|s_t, a_t)$ represents the state transition probability after the action is performed. This equation is used to estimate the impact of different audit actions on risk diffusion or risk mitigation, which is the core basis for the update of reinforcement learning strategy.

To map discrete states to orderable risk intensities, a scoring function combining severity, confidence, and time decay is set as follows:

$$\rho_t = \sigma(w_\rho^T h_t + \gamma\xi_t + \delta\kappa_t - \eta\Delta t_t) \tag{8}$$

Here, ρ_t represents the risk intensity score; Let ξ_t denote the severity prior; κ_t is the model confidence. Δt_t represents the time interval from the most recent disposal; γ, δ and η are the weight coefficients. This formula transforms the hidden states into interpretable scores, which can directly serve the low, medium and high risk level outputs.

Compared with common classification models, the proposed state space can simultaneously contain textual evidence, behavior trends and disposal feedback, so that each risk assessment no longer relies on isolated samples. The continuously updated risk memory can retain the repeated boundary events, and even if the intensity of a single event is not high, the model can also combine the historical trajectory to judge its cumulative impact. Thus, the stability of the early warning results is enhanced, and the false alarm fluctuation and missed detection can be controlled.

3.3 Action Space and disposal strategy Design of educational legal risk

Educational legal risk action space is used to connect risk states and disposal strategies. ERL-Risk does not limit the output to a single risk level label, but converts the assessment results into executable actions, including low risk archiving, medium risk review, high risk warning, manual audit trigger, evidence acquisition and disposal prioritization. Action design directly affects whether the model can integrate the risk identification results into the education compliance process. Therefore, this section establishes the calculation relationship among risk intensity, audit cost, responsibility subject and disposal time limit, so that the model output has decision-making meaning and business executability.

Fig. 3 shows the generation process of action space and disposal strategy of educational legal risk. After receiving the current risk status, the model generates the actions of low risk archiving, medium risk review, high risk warning, manual review, evidence supplement and disposal sorting in the candidate action pool. The clause constraint verification module is used to exclude actions that do not meet the permission or compliance boundary, and the action score module calculates the priority value by integrating risk intensity, evidence integrity, audit cost and model confidence. The strategy probability allocation module outputs the action selection probability, and the manual review gating module intercepts the high risk or low confidence events, and finally forms an executable disposal strategy.

Figure 3: Action space and disposal strategy generation process of educational legal risk

In order to match the disposal action with the risk level, audit cost and subject responsibility, it is necessary to establish a unified scoring relationship for each candidate action and output the orderable action priority value, as shown in the following equation:

$$q_t(a) = \theta_a^T[s_t \oplus \rho_t \oplus c_t] - \lambda_1 \text{cost}(a) + \lambda_2 u(a) \tag{9}$$

Here, $q_t(a)$ represents the selection score of action a in state s_t. Let ρ_t denote the risk intensity; c_t represents the compliance scenario constraints; $\text{cost}(a)$ represents the audit cost required to perform the action; $u(a)$ represents the expected utility of the action for risk mitigation; The parameters θ_a, λ_1 and λ_2 are trainable parameters. This formula puts risk severity, compliance constraint and execution cost into the same calculation framework, so that the action output not only pursues high sensitivity, but also takes into account the audit load and disposal efficiency in the education management process.

In order to avoid the rigid action selection caused by the fixed threshold, the model uses a constrained probability assignment method to describe the trigger strength of the candidate action, and blocks the action that does not conform to the system boundary, as shown in the following equation:

$$\pi(a|s_t) = \frac{\exp(q_t(a)/T) \cdot I(a \in A_t^{\text{allow}})}{\sum_{a' \in A_t} \exp(q_t(a')/T) \cdot I(a' \in A_t^{\text{allow}})} \tag{10}$$

where $\pi(a|s_t)$ represents the probability of action selection. T represents the temperature coefficient, which is used to adjust the smoothness of the action distribution. A_t^{allow} denotes the set of actions allowed to execute in the current state. $I(\cdot)$ is the indicator function. This formula limits the exploration scope of the model through the constraint set, so that the policy network will not generate high-risk actions such as ultra vires disposal, insufficient evidence disposal or lack of manual review in the learning process.

In order to measure the comprehensive impact of different disposal actions on risk diffusion, disposal speed and audit load, it is necessary to construct an action utility function to synchronously incorporate risk mitigation benefits and process costs into the calculation, as shown in the following equation:

$$U_t(a) = \omega_1 \Delta risk_t(a) + \omega_2 \Delta time_t(a) + \omega_3 conf_t(a) - \omega_4 load_t(a) \quad (11)$$

Here, $U_t(a)$ represents the comprehensive utility of action. $\Delta risk_t(a)$ is the expected risk reduction after performing the action. $\Delta time_t(a)$ represents the early warning gain; $conf_t(a)$ is the model confidence. $load_t(a)$ indicates the manual review load; ω_1 to ω_4 are the weight coefficients. This formula is suitable for dealing with the tension between high-risk recall and low-cost disposal, avoiding the model to simply push all boundary events to manual review, and avoiding underestimating the urgency of handling high-risk events.

In order to suppress improper automatic disposal in high risk, low evidence integrity or low confidence scenarios, the model sets a manual review gating function to constrain the automatic output boundary, as shown in the following equation:

$$G_t = \sigma(\alpha_1 \rho_t + \alpha_2 (1 - z_t) + \alpha_3 sev_t - \alpha_4 conf_t) \quad (12)$$

Here, G_t represents the triggering probability of manual review; z_t represents the completeness of evidence chain; sev_t means severity prior; $conf_t$ represents the action confidence. α_1 to α_4 are the gating coefficients; Let $\sigma(\cdot)$ denote the Sigmoid map. This formula makes the events with high severity, insufficient evidence or insufficient model grasp enter the manual review queue, and reduces the unclear responsibility and result deviation caused by automatic disposal.

After the action space is completed, ERL-Risk can output multi-level disposal strategies according to different risk states. Low-risk events enter the archive and prompt, medium-risk events enter the review queue, high-risk events trigger early warning and evidence collection. Each action of the model is written into a log to form a traceable sample, which provides a basis for subsequent reward estimation and policy update. At the running level, the action set will also be verified with the role permissions. The ordinary management end can only trigger the prompt and review, and the legal end can confirm the high-level disposal. This restriction ensures that the output of the model is consistent with the real business authority, and makes the educational legal risk assessment shift from single identification to an executable, traceable and reviewable intelligent disposal process.

3.4 Multi-objective reward function and dynamic evaluation mechanism construction

The multi-objective reward function is used to evaluate the comprehensive effect after the execution of disposal actions. The dynamic assessment of educational legal risk cannot only pursue classification accuracy, but also need to take into account high-risk recall, early warning advance, false alarm penalty, disposal cost and manual review consistency. ERL-Risk converts the disposal feedback, review conclusion and subsequent event changes into reward signals, so that the policy network can form a stable trade-off between multiple objectives and gradually learn the disposal preferences in line with the educational compliance process.

Fig. 4 shows the closed-loop process of the multi-objective reward function and the dynamic evaluation mechanism. After the model outputs the disposal action, the system receives the feedback information such as the manual review conclusion, the subsequent change of the event, the disposal time, the false alarm record, the false alarm record, and the audit cost. The reward calculation module converts recognition accuracy, high-risk recall, warning advance, disposal cost and misjudgment penalty into a unified reward signal. The long-term value estimation module evaluates the impact of the current action on the subsequent risk change, and the policy network updates the action selection tendency

according to the reward value, and triggers the risk level reevaluation.

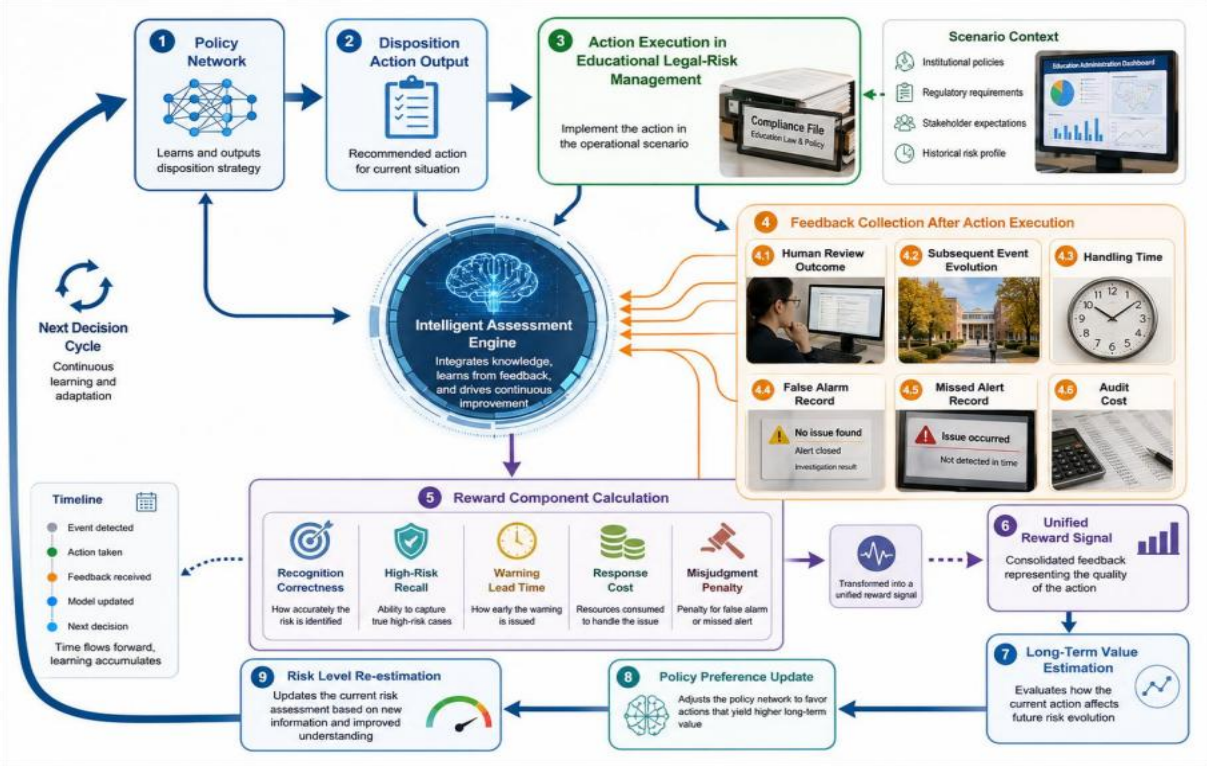

Figure 4: Multi-objective reward function and dynamic evaluation mechanism process

In order to take into account risk level recognition, high-risk recall, early warning and disposal cost, multiple evaluation objectives are compressed into a unified reward signal and dimensional alignment is completed, as shown in the following equation:

$$R_t = w_1 Acc_t + w_2 Rec_t + w_3 Lead_t - w_4 Cost_t - w_5 False_t \quad (13)$$

where R_t represents the immediate reward; Acc_t stands for risk level identification of correct benefits; Rec_t indicates high-risk recall gain; $Lead_t$ stands for early warning gain; $Cost_t$ stands for disposal cost; $False_t$ represents the false alarm penalty. w_1 to w_5 represent the reward weights. This formula unified the model performance index and the system operation cost as the optimization objective, so that the policy learning was not driven by a single accuracy, but could pay attention to the risk control effect and audit resource consumption at the same time.

In order to enhance the sensitivity of the reward function to high-risk missed detection, early warning delay and unsafe actions, the model introduces an asymmetric penalty term to constrain the underestimation behavior and the transboundary disposal behavior, as shown in the following equation:

$$P_t = \eta_1 I(y_t = H, \hat{y}_t \neq H) + \eta_2 I(delay_t > \tau) + \eta_3 I(a_t \notin A_t^{safe}) \quad (14)$$

where P_t stands for risk penalty; y_t represents the true risk level. $\hat{y}_t$ represents the model output level. H is the high risk category. $delay_t$ is the warning delay. Let τ denote the acceptable time limit; A_t^{safe} denotes the set of safe actions. $\eta 1$ to $\eta 3$ represent the penalty coefficients. This formula sets a higher penalty for high-risk missed detection, so that the model maintains a more prudent judgment boundary on serious events, and reduces the

subsequent diffusion caused by underestimating the risk.

In order to describe the long-term reward after the disposal action, the model defines the discount cumulative value estimation relationship, and incorporates future reward and future punishment into the current state evaluation, as shown in the following equation:

$$V(s_t) = \mathbb{E}_\pi \left[\sum_{k=0}^{L} \gamma^k \left(R_{t+k} - P_{t+k} \right) \mid s_t \right] \tag{15}$$

where $V(s_t)$ represents the state value; Let γ denote the discount factor; L represents the evaluation window length. R_{t+k} represents the future reward;P_{t+k} is the future penalty; Let $\mathbb{E}_\pi$ denote the expectation under policy π. This formula makes the model focus on the delayed impact of disposal actions, which is suitable for dealing with the event trajectory of "weak short-term impact and subsequent continuous diffusion" in educational legal risk.

In order to keep the policy update process stable, the model uses a trimmed objective function to constrain the parameter update range, and uses the policy distribution difference to suppress mutation, as shown in the following equation:

$$J(\theta) = \mathbb{E}[\min(r_t(\theta) A_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) A_t)] - \beta \mathrm{KL}(\pi_{old}, \pi_\theta) \tag{16}$$

where $J(\theta)$ represents the policy objective function; $r_t(\theta)$ is the probability ratio of the new strategy to the old strategy. A_t stands for dominance estimation; $\mathrm{clip}(\cdot)$ represents the clipping function; $\mathrm{KL}(\pi_{old}, \pi_\theta)$ is the difference between the distribution of the old and the new policy. Let β denote the constraint weight. The formula limits the single update range of the parameters, so that the model can still maintain the evaluation continuity after the entry of new samples, and avoid sharp fluctuations in the output of the risk level.

The dynamic evaluation mechanism completes the closed loop by reward transmission back. The false alarm confirmed by manual review will reduce the corresponding action revenue, the high-risk missed detection will significantly increase the punishment, and the action with timely warning and stable disposal results will obtain higher returns. The reward weights are not fixed constants, but are periodically calibrated according to the false positive, false negative, and review agreement rates in the validation set. If the high risk recall decreased, the missed detection penalty was increased. If the manual review pressure is too high, the system will reduce the automatic upgrade benefit of low confidence events. This design links model performance to actual audit resources. At the same time, the reward log retains the basis for each weight adjustment, which is convenient for subsequent audit, reproduction experiment and cross-batch performance comparison, and reduces the interpretation fault caused by model update.

3.5 Model training process and risk level output mechanism

The training process of ERL-Risk consists of four parts: offline pre-training, playback sample update, policy fine-tuning, and risk level output. The training data comes from the educational legal risk event feature matrix, human review labels, historical disposal feedback, and subsequent event status. The model learns the basic mapping between risk states and level labels in the offline stage, and modifies the action selection tendency according to the reward function in the strategy fine-tuning stage. The training process does not directly replace human compliance judgment, but converts complex events into computable, orderable, and reviewable graded output results.

Fig. 5 shows the model training process and the risk level output mechanism. After the

risk event samples entered the training queue, the system completed batch sampling and state reorganization, and then calculated synchronously by the policy network, value network and level output layer. The labels are manually reviewed into the calibration module to correct the level boundaries and action payoffs. After training, the model outputs low risk, medium risk and high risk to the warning interface, and synchronously saves the version number, confidence score and audit log.

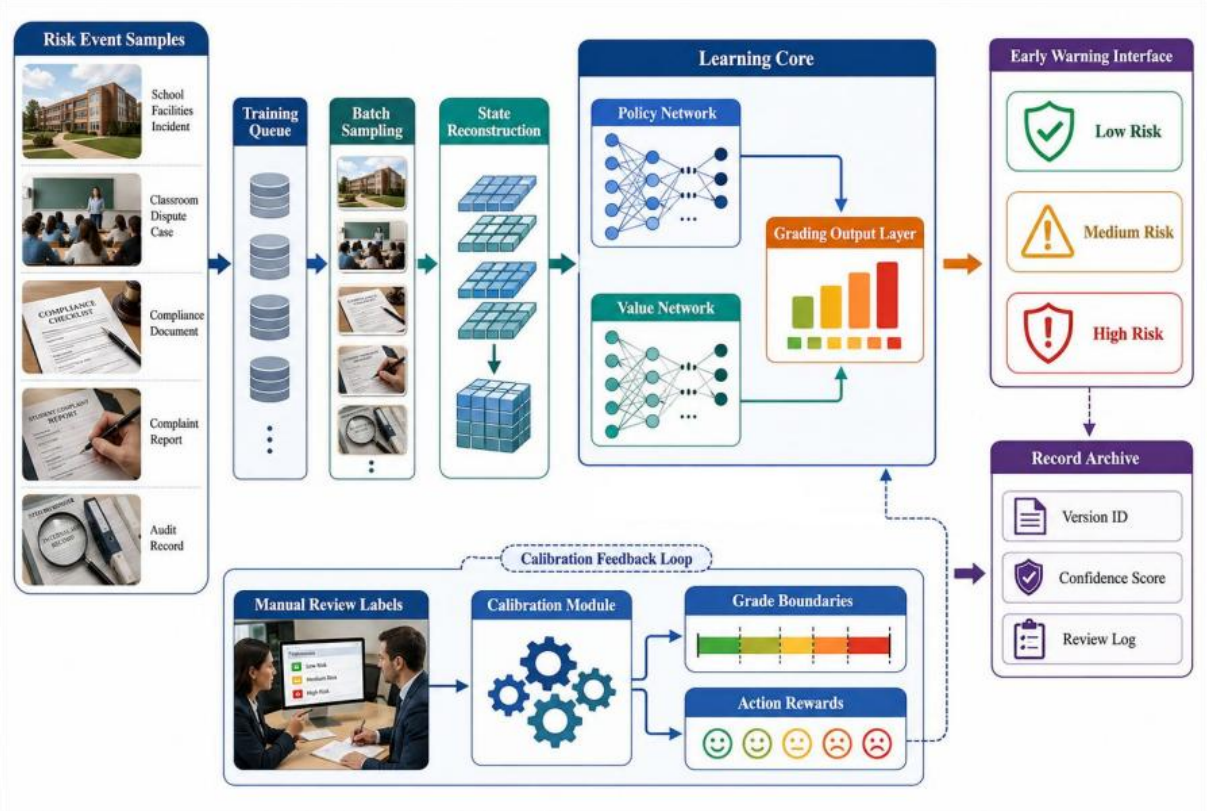

Figure 5: Model training process with risk level output mechanism

To keep the training process stable, the experimental parameters are set according to the risk sample size, state dimension, and online update frequency, as shown in Table 2. This setting ensures that the model can not only deal with multi-source features, but also complete the level output in the scenario with limited warning delay.

Table 2: ERL-Risk model training parameter Settings

Parameter Item	Setting Value	Function Description
Training Epochs	120	Controls the policy convergence process
Batch Size	64	Maintains stable gradient estimation
Learning Rate	0.0003	Controls the parameter update magnitude
Replay Capacity	20,000	Stores historical state transitions
Risk Categories	3 classes	Outputs low-risk, medium-risk, and high-risk levels

In order to make the policy learning, rank calibration and delay constraint converge under the same objective, the joint loss function is used in the training phase, as shown in the following equation:

$$\mathcal{L}_{\text{train}} = \lambda_1 \mathrm{CE}(y_t, \hat{y}_t) + \lambda_2 (\rho_t - \rho_t^*)^2 + \lambda_3 \max(0, d_t - \tau_d) + \lambda_4 \|\Theta\|_1 \qquad (17)$$

Here, $\mathcal{L}_{\text{train}}$ represents the model training loss; $\text{CE}(y_t, \hat{y}_t)$ is the cross-entropy between the true risk level and the predicted level. Let ρ_t^* denote the calibration strength after manual review; d_t represents the output delay; Let τ_d denote the allowable delay threshold; Θ denotes the set of model parameters; λ_1 to λ_4 represent the loss weights. This formula unifies the classification error, risk intensity deviation, response time delay and parameter sparsity constraints, and avoids the model only pursuing a single recognition accuracy while ignoring the system operating cost.

In order to map the continuous risk intensity into three levels of low, medium and high results, and retain the influence of double-check gating and value estimation on the level boundaries, the risk level is calculated as follows:

$$\hat{g}_t = \arg\max_{g \in \{L,M,H\}} \text{Softmax}\left(W_g[h_t \oplus \rho_t \oplus V(s_t) \oplus G_t] + b_g\right) \tag{18}$$

where $\hat{g}_t$ represents the final risk level; L, M, H represent low risk, medium risk and high risk respectively; $V(s_t)$ represents the state value estimate; W_g and b_g denote the rank output layer parameters. The formula takes temporal memory, risk score, long-term value and review boundary into the grade judgment, so that the output results have the basis of dynamic update and manual verification.

After model training, the system does not directly overwrite the old version, but performs grayscale evaluation in the validation set and small-scale online samples. If the high risk recall rate, false alarm rate, response delay and review consensus rate all meet the threshold requirements, the new version will enter the official interface. The risk level output layer will synchronously give the level label, confidence, trigger action and evidence index, which is convenient for the education management end and the legal end to view the judgment basis. The replay buffer continuously holds new samples, enabling the model to absorb recent compliance events and disposal feedback. This mechanism shifts educational legal risk assessment from static label judgment to a computational process of continuous training, continuous calibration, and continuous output.

4 Results

4.1 Dataset description and experimental environment setup

The experimental data were constructed around the dynamic assessment task of educational legal risk, and the sample sources were simulated school governance system, compliance review records and expert labeling cases. The dataset contains 8640 anonymized risk events, 42300 platform behavior logs, and 3120 policies and contract terms, covering scenarios such as student rights, teacher behavior boundaries, data authorization, contract performance, and platform content responsibility. All the texts were desensitized, and underwent word segmentation, entity recognition, clause matching and semantic coding. Behavior logs are aligned by timestamp, and actions such as access, commit, modify, appeal, and feedback are transformed into continuous behavior sequences. The disposal results are reviewed by experts and mapped into low, medium and high risk labels, which are used to support risk level recognition and reinforcement learning strategy training. In order to adapt to the ERL-Risk model, continuous events are sorted into state-action-feedback sequences, and the training set, validation set and test set are divided into 70%, 15% and 15% to avoid cross-set leakage of the same subject records.

In order to ensure that the experimental results can be reproduced, the dataset scale, label setting, operating environment and comparison model need to be explained in the same place.

As shown in Table 3, risk events, behavior logs and clause texts correspond to the main input sources of the model, and the training platform, GPU and comparison model are used to illustrate the experimental operating conditions and performance comparison basis.

Table 3: Educational legal risk dataset and experimental environment setup

Item	Setting
Risk Events	8,640 anonymized samples
Behavior Logs	42,300 platform operation records
Clause Texts	3,120 policy and contract clauses
Risk Categories	Low risk, medium risk, high risk
Data Split	70% training set, 15% validation set, 15% test set
Training Platform	Python 3.10, PyTorch 2.1
GPU Configuration	NVIDIA RTX 3090
Comparative Models	Rule Matching, XGBoost, BiLSTM, Transformer

The experimental environment is implemented with Python 3.10 and PyTorch 2.1, the GPU is NVIDIA RTX 3090, the batch size is set to 64, the initial learning rate is 0.0003, and the number of training rounds is 120. ERL-Risk adopts the structure of PPO and Actor-Critic, the capacity of replay buffer is 20000, and the reward item includes the accuracy of risk level recognition, the recall rate of high risk, the amount of early warning advance, the penalty of false alarm and the disposal cost. All comparison models use the same training set, validation set and test set to ensure that the experimental results are comparable.

4.2 Comparison of risk level identification performance

In this section, the risk level recognition performance of rule matching, XGBoost, BiLSTM, Transformer and ERL-Risk is compared under the same data partitioning and operating environment. The risk label was set according to low risk, medium risk and high risk, and the content focus level judgment ability, high risk identification ability and consistency of review results were evaluated. Since educational legal risk events usually contain text semantics, behavior trajectory and disposal feedback at the same time, a single classification model is easily affected by sample boundaries and label overlap. ERL-Risk introduces risk state memory and reward feedback mechanism in the testing phase, which is used to test the applicability of reinforcement learning structure in dynamic assessment tasks.

Fig. 6 shows the recognition results of ERL-Risk on the three-level risk labels using confusion matrix heat maps. The correct recognition rate of low risk samples was 96.2%, of which 2.8% were classified as medium risk. The correct recognition rate of the medium-risk samples is 91.5%, and the errors are mainly distributed in the adjacent grades. The recall rate of high-risk samples reaches 94.6%, and the proportion of samples misjudged as low risk is 1.3%. The results show that the model can better distinguish different levels of risks, especially maintain strong sensitivity in the identification of high-risk events, which meets the requirements of low missed detection of serious events in educational legal risk early warning.

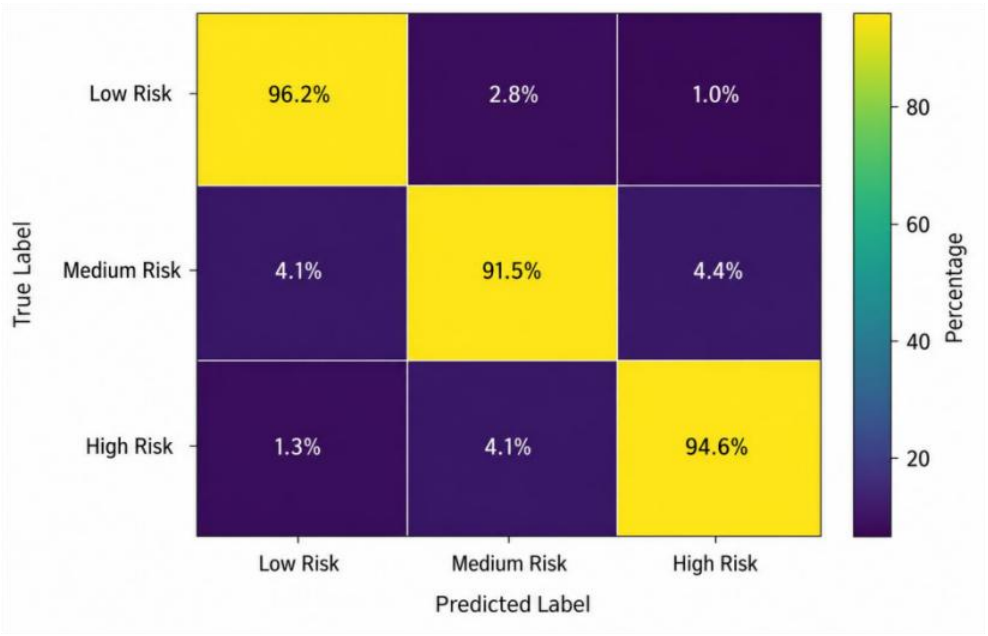

Figure 6: Heat map of ERL-Risk risk level confusion matrix

Fig. 7 compares the comprehensive performance of the five types of models in terms of accuracy, Macro-F1, high-risk recall and review agreement using radar chart. The four indexes of rule matching were 78.6%, 76.9%, 80.2% and 74.5%, respectively. XGBoost were 85.2%, 83.4%, 87.6% and 80.7%, respectively; BiLSTM were 88.7%, 86.1%, 89.9% and 84.3%, respectively. Transformers are 90.5%, 88.9%, 91.2% and 86.8%; ERL-Risk reached 93.4%, 91.8%, 94.6% and 90.7%, respectively. The coverage area of ERL-Risk in the radar chart is more complete, indicating that the reinforcement learning mechanism has more stable performance in risk level determination, high risk capture and manual review consistency.

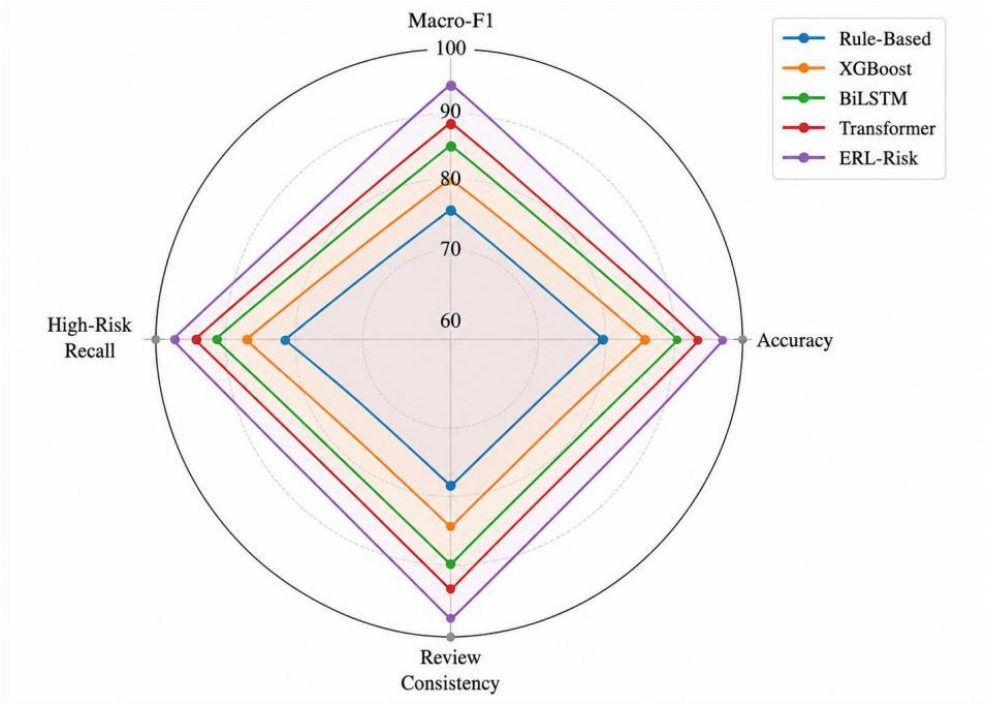

Figure 7: Radar chart of comprehensive performance of risk level identification

From the overall experimental performance, the advantages of ERL-Risk mainly come from the continuous linkage between state space, action feedback and reward constraints. After textual evidence, behavioral trajectories and historical disposal results are uniformly encoded, the model is able to correct the level boundary in the change of risk status. Compared with models that only rely on fixed rules or static features, this method is more suitable for the dynamic assessment task of educational legal risk, and also provides an experimental basis for subsequent high-risk event warning and strategy update analysis.

4.3 Analysis of warning effect of high-risk events

This section focuses on the analysis of the ability to detect high-risk events in advance, focusing on the identification sensitivity of the model, the amount of early warning and the level of false alarm control before risk escalation. Educational legal risks have the characteristics of stage evolution, and many high-risk events will show weak signals such as text semantic deviation, abnormal behavior frequency and unbalanced disposal feedback before they are officially triggered. Therefore, the effect of early warning cannot only look at the final classification results, but also the ability of the model to capture the rising process of risks.

To compare the distribution of the warning lead amount of different models for high-risk events, boxplots are used in Fig. 8 to show the warning time span of rule matching, XGBoost, BiLSTM, Transformer, and ERL-Risk on the test set. The median warning lead of ERL-Risk reaches 2.7 days, and the upper interquartile value is 3.4 days, which is significantly higher than Transformer's 2.1 days, BiLSTM's 1.8 days, XGBoost's 1.3 days, and rule matching's 0.9 days. At the same time, the discrete interval of ERL-Risk is more concentrated, which indicates that the warning output has better stability. This result shows that the reinforcement learning structure is able to use continuous state transition information to form earlier and more stable tips before the risk really escalates.

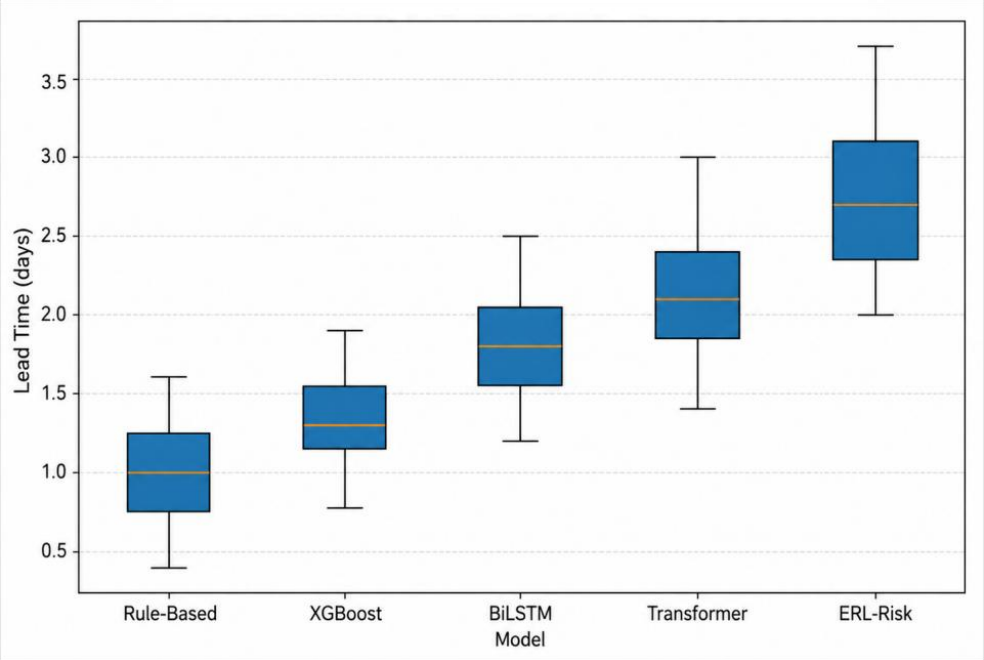

Figure 8: Box plots of the high risk event warning lead for different models

The difference reflected in Fig. 8 is directly related to the state memory and action feedback mechanisms. The static model relies on a single input to complete the judgment, and

the cumulative effect of continuous anomalies is underutilized. ERL-Risk incorporates clause evidence, behavior trajectory and historical disposal results into a unified state space, so that the model can form forward-looking judgments when the risk trend has not yet fully emerged. This ability has high adaptability for events such as complaint escalation, data authorization loss and contract execution disorder in education scenarios.

To further observe the relationship between warning hits and false alarms, Fig. 9 uses a heat map to show the high-risk hit rate and false alarm rate distribution of ERL-Risk under different warning Windows. When the warning window is set to one day, the hit rate is 88.9% and the false alarm rate is 5.4%. When the window is extended to two days, the hit rate rises to 92.7% and the false positive rate is 6.1%. When the window is 3 days, the hit rate reaches 94.3% and the false alarm rate is 6.8%. When the window continues to expand to 4 days, the hit rate only slightly increases to 94.8%, but the false positive rate increases to 8.2%. The thermal distribution illustrates that the 3-day window forms a more balanced configuration between hit rate and false alarm rate.

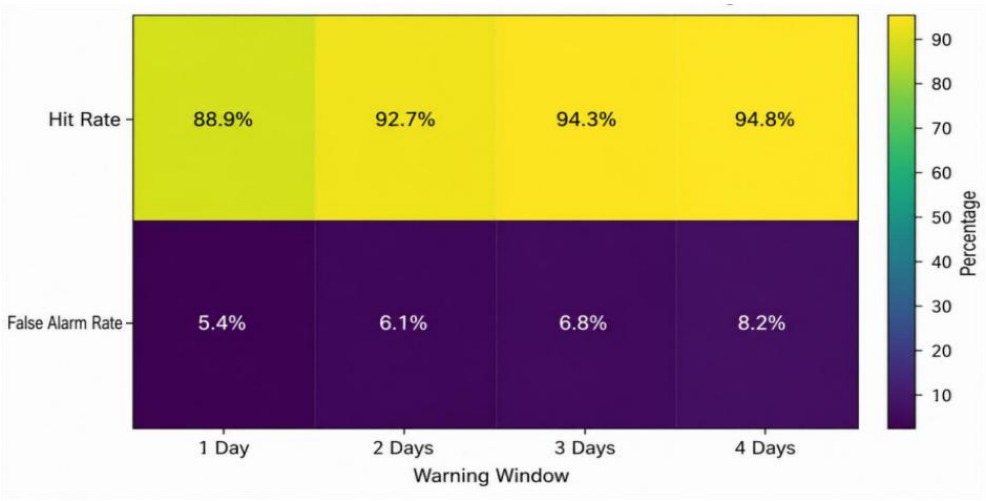

Figure 9: Heat plots of hit rate versus false alarm rate for different warning Windows of ERL-Risk

Together, the two results show that ERL-Risk forms a clearer time advantage and false alarm control capability in the early warning of high-risk events. The model uses continuous state memory to identify the signs of increasing risk, and then uses the reward function to constrain the early warning action, so that high-risk events can be captured in a more appropriate window.

4.4 Analysis of policy update stability and response delay

The stability of strategy update and the response delay are related to the online operation reliability of the dynamic assessment model of educational legal risk. ERL-Risk uses PPO clipping target, replay buffer and manual review feedback to control policy drift in the training phase, and records policy KL divergence, reward fluctuation and rank boundary change in each training phase. In order to avoid the level jump caused by the entry of new samples, the model sets the offline playback verification before updating, and the high-risk recall rate, review consensus rate and response delay are used as synchronization constraints.

Fig. 10 uses density plots to show the policy update convergence trajectory of ERL-Risk in the 30th, 60th, 90th and 120th training rounds. In the 30th round, the KL divergence of the strategy was mainly concentrated between 0.035 and 0.052, and the density peak was to the

right, indicating that the early strategy was still in a strong adjustment stage. After the 60th round, the peak value moved to 0.028, and the curve width significantly contracted. In the 90th round, the main density interval fell between 0.016 and 0.023. After the 120th round, the KL divergence stabilizes between 0.011 and 0.017. The density curve changes from wide to narrow, indicating that the policy network gradually converges in the feedback and the risk level boundary does not drift.

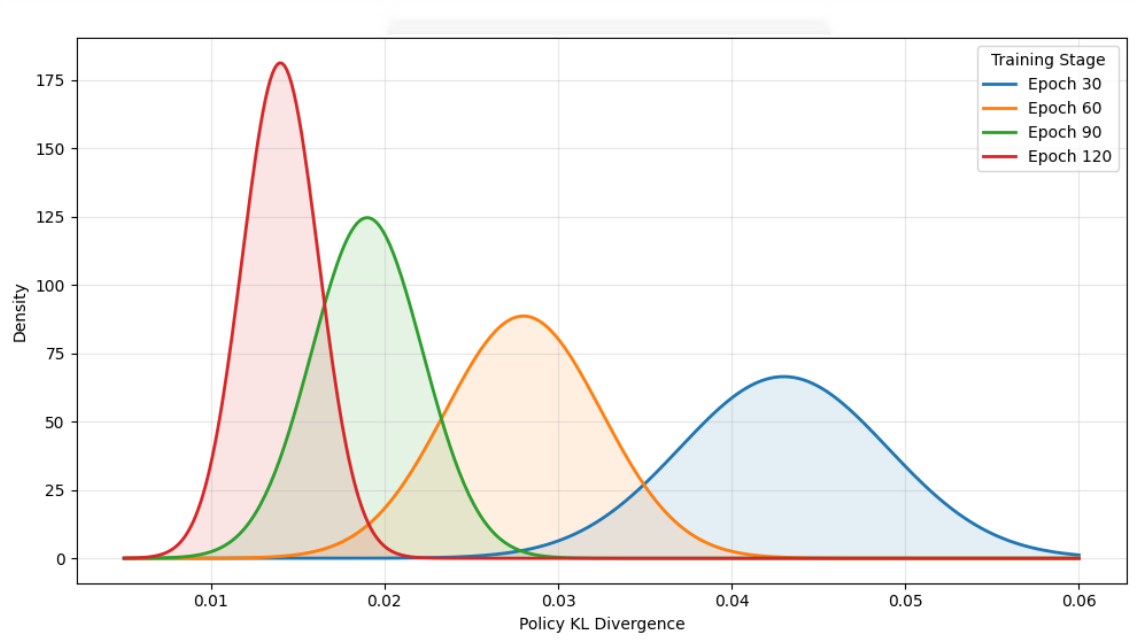

Figure 10: Density plot of convergence trajectories updated by the ERL-Risk policy

Fig. 11 uses contour plots to show the response delay distribution when the batch size and input feature dimension are changed together. When the batch size is 16 and the feature dimension is 128, the average response delay is about 0.18s. When the batch size is increased to 64 and the feature dimension is 256, the delay is 0.41s. When the batch size is 128 and the feature dimension is extended to 512, the delay increases to 0.73s. The contour line is relatively flat in the medium batch area, indicating that the model maintains a stable response in the school compliance review, complaint acceptance and data access alarm scenarios, and only when high-dimensional features are superimposed with large batch input, there is obvious computational pressure.

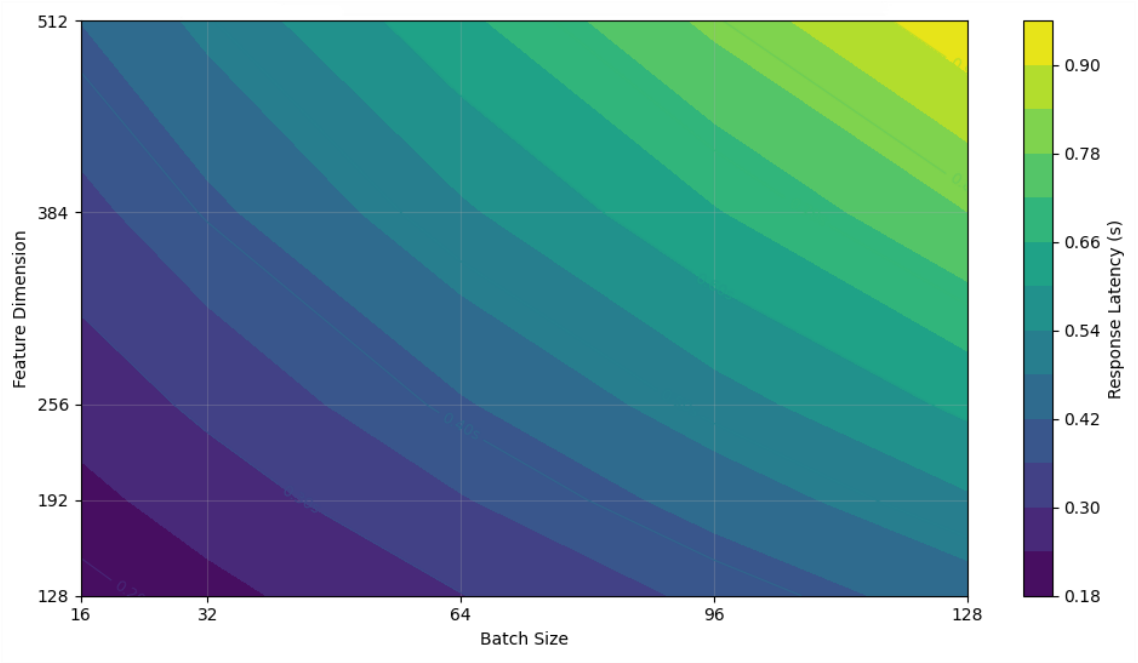

Figure 11: Contour plot of ERL-Risk response delay

Two results show that ERL-Risk maintains a balance between continuous learning and real-time inference. The distribution contraction in the process of policy update proves that the reward function and the review feedback have a constraint effect on the parameter update, and the response time delay results show that the model has the computing conditions to be embedded in the education management platform. Compared with the static classification model, ERL-Risk can maintain a stable judgment after new risk events continue to enter the system, and provide a reliable basis for the analysis of module contribution in ablation experiments.

4.5 Ablation experiment and module contribution analysis

This section performs module ablation of ERL-Risk on the same test set to evaluate the contribution of state memory, action constraint, reward penalty, and online update to risk level identification. The ablation method was to remove the key units item by item on the basis of the complete model and observe the changes of the three indicators.

As shown in Table 4, the precision of full ERL-Risk is 93.4%, Macro-F1 is 91.8%, and the recall of high risk is 94.6%. After removing the state memory, the precision rate was reduced to 89.1%, and the high-risk recall rate was reduced to 90.4%, indicating that the risk accumulation judgment was weakened after the lack of temporal connection between textual evidence and behavior trajectory. After removing action constraints, Macro-F1 decreases to 88.6%. After removing the reward penalty, the high risk recall rate decreases to 87.9%, indicating that the missed detection penalty has a constraint effect on serious event recognition.

Table 4: Results of ERL-Risk ablation experiments

Model Setting	Accuracy/%	Macro-F1/%	Recall/%
Full ERL-Risk	93.4	91.8	94.6
Without State Memory	89.1	87.5	90.4
Without Action Constraint	90.2	88.6	91.1
Without Reward Penalty	87.6	85.9	87.9
Without Online Update	88.8	86.7	89.6

To observe the effect of each module on error suppression, Table 5 shows the module contribution and error sources. State memory compresses boundary misjudgment caused by repeated complaints and contract deviation. Automatic escalation when there is insufficient evidence of action constraint blocking; Reward and punishment strengthen the cost of high risk missed detection; The absorbing review sample is updated online so that the risk boundary is calibrated with recent events.

Table 5: Analysis of module contributions versus error sources

Module	Computational Function	Error Suppression Target
State Memory	Preserves historical trajectories	Confusion in boundary samples
Action Constraint	Restricts out-of-bound actions	False alarms caused by insufficient evidence
Reward Penalty	Amplifies the cost of missed detection	Underestimation of high-risk events
Online Update	Calibrates recent samples	Sample distribution drift

The results in both tables show that the performance of ERL-Risk comes from the linkage between the state, action, reward, and update mechanisms. State memory provides a continuous evidence base, action constraints ensure that the output conforms to the educational compliance process, rewards and punishments put high-risk events into more stringent learning objectives, and online updates reduce the adaptation bias of old samples to new events. After ablation, each index decreased synchronously, indicating that the complete model was more suitable for the dynamic assessment scenario of educational legal risk.

5 Discussion

This paper establishes a computational link from data encoding, state modeling to action output and reward update around the dynamic assessment of educational legal risk. Compared with the rule matching and static classifier, the model organizes the complaint text, contract terms, platform logs and historical disposal results into continuous states, so that the risk judgment does not stop at a single label prediction. In the experiment, the model maintains good performance in accuracy, Macro-F1, high-risk recall and warning advance, indicating that the reinforcement learning structure can use subsequent feedback to correct the level boundary. Action space and manual review gating limit the scope of automatic disposal and prevent high-risk and low-evidence events from being directly judged. Ablation results also show that state memory, reward punishment and online update all contribute to high-risk identification. The value of the model does not lie in replacing legal personnel, but in transforming scattered evidence into traceable, orderable and recheckable risk clues, and providing stable compliance calculation support for education management platform. In a real deployment, model output would also need to be run in conjunction with permission control,

log auditing, and accountability retention. The risk score, trigger action, and evidence index should be saved synchronously for subsequent review. Educational legal risks are affected by changes in system versions, platform rules and subject relationships, and the sample distribution is easy to shift with business scenarios. The continuous learning mechanism can incorporate recent events and review results into the update process, and reduce the error caused by the dominant judgment of old samples. The PPO clipping update and the replay buffer jointly constrain the parameter change, so that the model can still maintain a stable response speed and risk level boundary after the new event enters. Based on this characteristic, the model is more suitable to undertake the tasks of auxiliary screening, early warning ranking and evidence aggregation.

6 Conclusions and future work

This paper constructs a dynamic assessment model of educational legal risk based on reinforcement learning, which converts educational management texts, platform behavior logs, policies and contract terms, and historical disposal results into trainable state sequences, and uses action space, reward function and strategy update mechanism to complete risk level recognition and early warning output. The experimental results show that the model has stable performance in risk level judgment, high risk recall, early warning advance and response delay, which can provide auxiliary decision support for digital campus compliance review. There are still some limitations in the research. The data mainly came from simulated school governance scenarios and expert annotation cases, and the coverage could not represent all educational institutions. Risk labels rely on manual review consistency, and complex dispute events may still have fuzzy boundaries. The model is sensitive to the version changes of legal rules, and the ability to transfer across regions and learning segments still needs to be verified. The follow-up research can be carried out from four directions: introducing the knowledge graph of education law, strengthening the clause relationship and the reasoning of the responsible subject; Federated learning and privacy computing were used to reduce the risk of cross-institutional data sharing. The fine-grained interpretation of complaint text and contract semantics was completed by combining the large language model. We develop interpretable reinforcement learning to enable clearer chains of evidence for each warning action, reward change, and risk level output. In the future, online testing can also be carried out in the real school affairs platform to verify the continuous operation ability of the model under the cooperation of multi-source data concurrency, rule update and manual review. On this basis, the model can be extended from a single proof sample to a regional education governance scenario, forming a more robust compliance early warning service, and retaining a complete basis for audit review.

About the Author

Zhang Jianfeng obtained his Doctor of Philosophy in Educational Administration from Beijing Normal University in June 2018. He is currently an Associate Professor at the School of International Rule of Law, where his research focuses on Chinese and international educational law and policy. He works at the School of International Rule of Law, Gansu University of Political Science and Law.

References

[1] Franceschelli G, Musolesi M. Reinforcement learning for generative AI: State of the art, opportunities and open research challenges[J]. Journal of Artificial Intelligence Research, 2024, 79: 417-446.

[2] Fritz-Morgenthal S, Hein B, Papenbrock J. Financial risk management and explainable, trustworthy, responsible AI[J]. Frontiers in artificial intelligence, 2022, 5: 779799.

[3] Memarian B, Doleck T. Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100152.

[4] Ismail N, Yusof U K. A systematic literature review: Recent techniques of predicting STEM stream students[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100141.

[5] Marx E, Leonhardt T, Bergner N. Secondary school students' mental models and attitudes regarding artificial intelligence-A scoping review[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100169.

[6] Farhi F, Jeljeli R, Aburezeq I, et al. Analyzing the students' views, concerns, and perceived ethics about chat GPT usage[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100180.

[7] Ocak C, Kopcha T J, Dey R. An AI-enhanced pattern recognition approach to temporal and spatial analysis of children's embodied interactions[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100146.

[8] Rodway P, Schepman A. The impact of adopting AI educational technologies on projected course satisfaction in university students[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100150.

[9] Khosravi H, Denny P, Moore S, et al. Learnersourcing in the age of AI: Student, educator and machine partnerships for content creation[J]. Computers and education: Artificial intelligence, 2023, 5: 100151.

[10] Priyambada S A, Usagawa T. Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100149.

[11] Annamalai N, Ab Rashid R, Hashmi U M, et al. Using chatbots for English language learning in higher education[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100153.

[12] Sanusi I T, Oyelere S S, Vartiainen H, et al. Developing middle school students' understanding of machine learning in an African school[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100155.

[13] Chomphooyod P, Suchato A, Tuaycharoen N, et al. English grammar multiple-choice question generation using Text-to-Text Transfer Transformer[J]. Computers and

Education: Artificial Intelligence, 2023, 5: 100158.

[14] Abichandani P, Iaboni C, Lobo D, et al. Artificial intelligence and computer vision education: Codifying student learning gains and attitudes[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100159.

[15] Bezirhan U, von Davier M. Automated reading passage generation with OpenAI's large language model[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100161.

[16] Kajiwara Y, Matsuoka A, Shinbo F. Machine learning role playing game: Instructional design of AI education for age-appropriate in K-12 and beyond[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100162.

[17] El Bahri N, Itahriouan Z, Abtoy A, et al. Using convolutional neural networks to detect learner's personality based on the Five Factor Model[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100163.

[18] Hornberger M, Bewersdorff A, Nerdel C. What do university students know about Artificial Intelligence? Development and validation of an AI literacy test[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100165.

[19] Ericsson E, Johansson S. English speaking practice with conversational AI: Lower secondary students' educational experiences over time[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100164.

[20] Andersen M R, Kabel K, Bremholm J, et al. Automatic proficiency scoring for early-stage writing[J]. Computers and Education: Artificial Intelligence, 2023, 5: 100168.