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

Design and Effect Evaluation of Immersive Teaching Mode of Virtual Simulation Technology in Legal Training Scene

Bingcui Jian[1,*]

[1] School of Law, Jianghan University, Wuhan, 430056, Hubei, China

SUMMARY: *Legal training based on virtual simulation requires accurate scene reconstruction, character interaction tracking, and computable immersion effect evaluation. Integrating VR scene modeling, multimodal interactive capture and intelligent feedback, this paper designs an immersive teaching mode for legal training scenarios. The virtual court, mediation room and legal consultation space are constructed with procedural nodes, evidence objects, role tasks and case materials. Speech, eye movement, head movement, operation log, task duration, and interaction frequency were collected from 96 learners in 12 rounds of training to form 2840 labeled samples. The multi-modal fusion network is used to identify the character behavior and evaluate the immersion state, and the edge assistance module is used to reduce the feedback delay. The experimental results show that the behavior recognition accuracy is 92.4%, the weighted F1 value is 0.887, the task adaptation accuracy is 88.6%, and the average response delay is 46 ms. The results show that the system can support data-driven legal training in immersive practice environment, more reliable process evaluation, task feedback and real-time interactive feedback.*

KEYWORDS: *Virtual simulation; Legal training; Multimodal interaction; Immersion effect evaluation*

1 Introduction

Virtual simulation technology promotes legal training from static case teaching to immersive training form. Legal training includes tasks such as trial procedure, evidence cross-examination, mediation and negotiation, which not only needs to restore spatial order, role relationships and procedural nodes, but also needs to capture learners 'process data in speech, gaze, evidence selection and task submission. 3D modeling, human-computer interaction and multimodal recognition technologies provide a deployable computing framework for such scenarios, so that case facts and character actions form a continuous mapping in the simulation space.

Hidayah et al. [1] studied the application of virtual reality moot court in the study of constitutional court procedures, and showed that legal procedures can be transformed into role tasks and operation paths in the virtual scene. Alshaer[2] proposed an immersive VR case for prisoner training to show the adaptation value of VR environment in behavior training and situational feedback. Barbe et al. [3] constructed an open source VR training framework for the criminal justice system, which provided a technical reference for the scene reuse of legal training. King et al. [4] studied teacher-student interaction in an automated VR training system, and showed that interaction data could support training feedback and effect

*janewuhan@jhun.edu.cn
https://doi.org/10.65102/is2026977
Published: 30 April, 2026

verification. Calandra et al. [5] proposed to combine immersive VR with passive haptic interface for procedural learning tasks, so that operation sequence, action feedback and task completion can be recorded.

Boel et al. [6] studied low-cost mobile immersive VR serious games and showed that lightweight terminals can expand the scope of use of virtual training. Strojny and Dużmańska-Misiarczyk[7] summarized the measurement methods of virtual training effect and emphasized that the evaluation indicators should cover performance, experience and transfer results. Conrad et al. [8] studied the learning effect of immersive virtual reality in education and training, and pointed out that interaction intensity, immersion depth and task structure would affect the training results. Jongbloed et al. [9] conducted a review on immersive program training, indicating that procedural tasks are suitable for verification through VR process modeling, behavior recording and result analysis.

This paper constructs an immersive virtual simulation teaching mode around the legal training scene, uses three-dimensional scene modeling to express the spatial relationship of virtual court, mediation room and consultation room, and converts case materials, evidence objects, program nodes and role tasks into callable data resources. The system records the learner's multimodal behavior through the VR headset, eye tracking module, voice acquisition device and interaction handle, and uses the timestamp to synchronize the voice, gaze, action trajectory and operation log. At the model end, a multimodal fusion network is introduced to identify court speeches, evidence call, program advancement, role collaboration and invalid operations. Combined with the immersion state evaluation module, the scene engagement, interaction continuity, attention stability and task adaptation results are output. The simulation engine, data cache module, model reasoning interface and feedback generation module were used to run in cooperation on the platform, forming a closed-loop process of role execution, data collection, model judgment and result feedback. The model retains the whole process data of legal training, reduces the subjective fluctuation of manual evaluation, and provides stable samples for model training, scene iteration and quality analysis.

2 Related work

2.1 Research on Virtual Simulation Technology and Legal Training Scene

The core of virtual simulation technology used in legal training is to transform court space, procedure sequence, evidence relationship and role tasks into interactive data structures. Simulation trial, mediation negotiation and legal consultation are not simple three-dimensional display tasks, but a process chain composed of court instructions, identity confirmation, evidence cross-examination, opinion expression and result feedback. 3D modeling, collision detection, event monitoring and log tracking can solidify these processes in the virtual space, so that students 'speech, movement, gaze and evidence call can be recorded by the system. This kind of design enables legal training to have the computational basis of scene reproduction, process trace and result verification, and also provides data sources for behavior recognition and effect evaluation.

Thomann et al. [10] studied the knowledge acquisition and motivation change of immersive VR in vocational education, and showed that the effect of virtual training was related to task structure, feedback form and interaction density. Legal training also depends on the division of task nodes. The judge's bench, the original defendant's bench, the evidence display area and the auditing area in the virtual court need to correspond to the procedural tasks. Khorasani et al. [11] analyzed the influence of hands-on interaction on embodied

learning in VR, and pointed out that manipulation intervention would change learners' understanding of spatial relationships and task goals. Evidence dragging, dossier leafing, and character stance adjustment can be set as computable events instead of staying at the visual presentation level.

Wheeler et al. [12] designed and evaluated a virtual reality learning environment for firefighters, breaking down high-risk tasks into scene objects, action paths, and outcome indicators. This kind of path can be used for reference in legal training to integrate case facts, evidence materials and procedural actions into a continuous interactive chain. Clay et al. [13] evaluated the role of immersive VR in medical procedure training and pointed out that procedure training requires recording action sequences, dwell times, and error nodes. The same idea can be used to collect the timing of speech, the sequence of evidence presentation and procedural omission in the trial. Hjellvik and Mallam[14] studied training transfer validity in VR simulator evaluation, emphasizing that virtual training results should establish a correspondence with real task performance. From this point of view, the research of legal training scene needs to deal with three types of mapping: space mapping to restore court layout, program mapping to limit role behavior, and data mapping to save voice, sight, action and log. In this paper, the virtual training scene is organized as an object base, a program base and an interaction rule base. The object library saves seats, evidence and role actions, the program library saves court, inquiry, cross-examination and debate nodes, and the interaction rule library limits permissions and feedback conditions. The three types of resources jointly support real-time rendering, event archiving and model input, so that legal training turns to a simulation system that can be recorded, analyzed and fed back.

2.2 Research on multimodal interaction recognition and immersion effect evaluation

Multimodal interaction recognition focuses on learners' speech, gaze, motion trajectory, and operation log in the virtual space, while immersion evaluation focuses on the relationship between these signals and character tasks, program completion, and scene engagement. Legal training has a clear division of labor and procedural constraints, and a single click record is difficult to explain whether students complete legal reasoning, evidence judgment and collaborative response. Laine et al. [15] analyzed the challenges of immersive VR in complex skill training and pointed out that task complexity, feedback mode and user load would affect training continuity. This conclusion suggests that the behavior process and task results should be included into the model input synchronously in the legal training evaluation. Cabrera-Duffaut et al. [16] studied the contribution of immersive learning platforms to the development of higher education capabilities, emphasizing that platform evaluation needs to combine the ability structure, interaction path and learning evidence. Coban et al. [17] used meta-analysis to discuss the learning potential of immersive VR, indicating that the immersive experience needs to be transformed into a verifiable indicator system.

In terms of computational implementation, the legal training system can align voice, eye movement, movement and log according to a unified timestamp, and slice events according to court speech, evidence viewing, material switching, role response and so on. The speech duration and term hit were extracted from the speech stream, the eye movement stream calculated the fixation duration and gaze shift in the evidence area, the action stream recorded the head orientation, the handle trajectory and the body displacement, and the log stream saved the evidence call, program advancement and task submission. There is a complementary relationship between different modalities, and there is also redundant information, which needs to be fused by attention weight or gating mechanism. To make the

correspondence between related methods and the task in this paper clearer, existing multimodal recognition ideas can be organized into the following types, as shown in Table 1.

Table 1: Comparison of multimodal interaction recognition and immersion evaluation methods

Method Type	Main Input Data	Computational Processing Method	Evaluation Metrics	Adaptation Direction for Legal Training
Speech Semantic Recognition	Speech content, pauses, legal terms	ASR transcription, keyword matching, semantic encoding	Speech completeness, expression accuracy	Court statement, cross-examination response
Gaze and Posture Analysis	Gaze areas, head orientation, body posture	Eye-tracking heat-zone statistics, posture estimation	Attention stability, role engagement	Evidence review, courtroom interaction
Operation Log Modeling	Clicks, dragging, evidence invocation, task submission	Event slicing, sequence encoding, node archiving	Procedural completion rate, response time	Evidence submission, process advancement
Multimodal Fusion Evaluation	Speech, gaze, motion, logs	Attention fusion, gated weighting, result calibration	Immersion state, task adaptability	Comprehensive practical training effect evaluation

Petersen et al. [18] studied the driving effect of immersion and interactivity on VR learning, and pointed out that spatial presence, action participation and task engagement would jointly affect learning results. This view is applicable to the evaluation of role substitution in legal training. The system can generate immersion scores according to whether students pay attention to the evidence area, whether they respond according to the procedure, and whether they maintain the consistency of role behavior. Makransky and Mayer[19] proposed the evidence of immersion principle in virtual visit learning, indicating that visual presence can change the information processing path. Court layout, evidence display and role distance in legal training can also affect students' understanding of procedural relations. Mayer et al. [20] discussed the potential and limitations of immersive VR learning, emphasizing that realism, cognitive load and task objectives need to be kept in harmony. Based on this, this paper designs the immersion effect evaluation as a multimodal fusion task, and uses speech semantics, eye hotspots, action trajectories and operation logs to generate training portraits, which are aligned with the teacher review results. This path can extend the immersion experience from subjective questionnaire to traceable calculation indicators, and provide data basis for real-time feedback of the platform, case difficulty adjustment and effect verification.

3 Design of immersive virtual simulation teaching mode for legal training

3.1 System architecture design of virtual simulation scene for legal training

3.1.1 Mapping between legal training task types and virtual scenes

After the legal training task enters the virtual simulation system, it needs to complete the task type identification and space unit binding. In this paper, five basic tasks are set up, including simulated trial, evidence cross-examination, mediation and negotiation, legal consultation and clerk recording. Each task corresponds to different spatial regions, role permissions and event trigger conditions. In the virtual court, the judge's bench, the original defendant's bench, the evidence display area, the auditing area and the material retrieval area no longer exist as static models, but are encoded as interactive nodes with procedural attributes. After the student enters the scene, the system assigns the operable objects according to the role identity, and records its moving path, evidence call, speech response and node completion status.

In order to quantify the correspondence strength between legal tasks and virtual space objects, the system needs to consider the relationship between procedures, evidence and role at the same time, and the mapping score is as follows:

$$M_{r,s} = \sigma(\alpha C_{r,s} + \beta E_{r,s} + \gamma P_{r,s} - \delta L_{r,s}) \tag{1}$$

Here, $M_{r,s}$ represents the mapping score between the r legal task and the s simulation space object. $C_{r,s}$ denotes the fit between the role and the spatial location; $E_{r,s}$ denotes the correlation degree between the evidence object and the task content; $P_{r,s}$denotes the matching degree of program nodes; $L_{r,s}$ denotes the interactive load generated by spatial handoff; $\alpha,\beta,\gamma,\delta$ are weight parameters, and $\sigma(\cdot)$ is used to compress the results to a unified evaluation interval. This formula is used to decide the preferential binding relationship of different tasks after they enter the virtual space.

In order to ensure that the task path is consistent with the simulation space nodes, the system needs to constrain the node order, role permissions and trigger relationships. The path function is shown in the following equation:

$$Q_i = \frac{\sum_{k=1}^{K_i} \rho_k \, I(v_k \to v_{k+1})}{K_i + \eta_i + 1} \tag{2}$$

where Q_i represents the program consistency of the i training path; K_i denotes the number of nodes in the path; Let ρ_k denote the weight of the k program node; $I(v_k \to v_{k+1})$ indicates whether the sequence of nodes satisfies the legal process constraints. Let η_i denote the number of invalid jumps occurring in the path. The formula is able to determine whether the student advances the task in the order of court session, evidence, cross-examination, debate, and feedback.

In order to depict students' space access quality in different legal tasks, the system jointly models stay, operation and evidence hit degree, and the evaluation results are shown in the following equation:

$$U_i = \tanh(\mu_1 T_i + \mu_2 O_i + \mu_3 H_i - \mu_4 D_i) \tag{3}$$

Here, U_i represents the spatial entry quality of the student in the i task; T_i is the effective dwell time; O_i represents the number of operations completed. H_i indicates the degree of evidence hit. D_i represents the interference value formed by irrelevant region stay and error operation. μ_1 to μ_4 are the tuning parameters. The formula makes the virtual space not only assume the display function, but also assume the training behavior judgment function.

The above mapping relationship enables legal tasks, scene regions and student behaviors to form a unified data link. The system completes spatial assembly before training starts, event listening during training, and path retracing after training. The design can avoid the virtual scene staying at the display level, so that the tasks such as simulated trial, cross-examination and mediation have clear computing entrances, which provides a stable basis for subsequent role interaction modeling and immersion effect evaluation.

3.1.2 Role interaction process and program node design

The role interaction in legal training is obviously procedural. The judge is responsible for process control, the lawyer is responsible for evidence cross-examination, the party is responsible for fact statement, the clerk is responsible for record filing, and the mediator is responsible for dispute coordination. The virtual simulation system needs to decompose these actions into events that can be listened to, including speech initiation, evidence submission, objection raising, node confirmation, task transfer and feedback generation. Each role only has the operation permission matching with the identity, and the system restricts the interaction sequence through the program node to avoid students from directly entering the subsequent link before completing the pre-task.

In order to describe the state transition relationship of multiple roles in the process of program advancement, the system needs to record the identity permission, speech sequence and interactive response, and the transition probability is shown in the following equation:

$$P(s_{t+1} = j | s_t = i, a_t, r) = \frac{\exp(\phi_{ij} + \psi(a_t, r) + \chi_t)}{\sum_{m=1}^{S} \exp(\phi_{im} + \psi(a_t, m) + \chi_t)} \tag{4}$$

where s_t represents the current program state, s_{t+1} represents the next program state, a_t represents the current interaction action,r represents the role identity, ϕ_{ij} represents the base transition strength between program states, $\psi(a_t, r)$ represents the matching degree between actions and role permissions, and χ_t represents the time constraint of the current node. This formula is used to calculate whether the role action can legally push the program to the next node.

In order to determine whether the role interaction meets the requirements of the program node, the system takes the speech object, evidence action and process stage into the calculation together, and the matching degree is shown in the following equation:

$$G_t = \text{sigmoid}(\lambda_1 A_t + \lambda_2 B_t + \lambda_3 C_t - \lambda_4 R_t) \tag{5}$$

Here, G_t represents the matching degree between the t interaction event and the program node. A_t indicates whether the speaker is correct or not; B indicates whether the evidence call meets the task requirements; C_t indicates whether the process phase is in the allowable range; R_t represents deductions for duplicate operations, ultra vires, and invalid responses. λ_1 to λ_4 are the parameter weights. This formula can transform the role interaction from manual observation to systematic judgment.

In order to evaluate the continuity degree of role collaborative behavior in virtual space,

the system integrates response delay, round cohesion and task feedback, and the collaboration index is shown as follows:

$$H = \frac{1}{T}\sum_{t=1}^{T} \omega_t \left(1 - \frac{d_t}{d_{max}}\right) \frac{q_t + p_t}{2} \tag{6}$$

where H represents the role synergy index;T is the number of interaction rounds; Let ω_t denote the interaction weight in round t; d_t is the response delay, and d_{max} is the maximum delay allowed by the system. q_t represents the quality of task cohesion. p_t represents the feedback completion degree. This formula can reflect whether a continuous, compliant and traceable training process is formed between different roles.

The design of role interaction process makes the virtual legal training have a clear program boundary. The system no longer only records whether the student completes the task, but further identifies whether the role action conforms to the identity authority, the program state and the collaborative relationship. The judge's flow control, the attorney's evidence response, the parties' statements of fact, and the clerk's node records can all be transformed into structured events. This processing method provides a clear basis for subsequent training behavior collection, immersion state analysis and teacher review.

3.1.3 Data processing of case materials, evidence objects and scene resources

Case materials, evidence objects and scene resources are the basic data of virtual legal training. In this paper, the case facts, legal provisions, evidence pictures, audio and video materials, role descriptions, program nodes and scene components are integrated into the resource library, and coded according to the link of "material - evidence - node - role - feedback". Case materials are used to generate the task background, evidence objects are used to support cross-examination and judgment, scene resources are used to restore the court and mediation space, and program nodes are used to constrain the task sequence. All resources are configured with a unique number, type label, access permission and call status, which ensures that the model can identify the source of resources and interaction results.

In order to realize the unified coding of case materials, evidence objects and scene resources, the system needs to build a cross-resource correlation vector and index structure, which is expressed as follows:

$$Z_r = \mathrm{Norm}(B_m A_m + B_e A_e + B_s A_s + \kappa U_r) \tag{7}$$

Here, Z_r represents the uniform encoding vector of the r resource object. A_m represents the case material feature, A_e represents the evidence object feature, A_s represents the scene component feature, and U_r represents the resource usage state. B_m,B_e and B_s are mapping matrices of different resource types, κ represents the adjustment coefficient of resource status on coding results, and $\mathrm{Norm}(\cdot)$ represents the standardization function. The formula is used to transform text materials, evidence objects and 3D resources into input representations in the same computational space, which is convenient for subsequent action recognition and effect evaluation.

In order to ensure that the resource processing flow can support the platform call, the system establishes the resource structure according to the data source, coding field, call mode and evaluation purpose, as shown in Table 2.

Table 2: Datafication structure of case materials, evidence objects and scene resources

Resource Type	Data Object	Encoding Field	Invocation Method	Evaluation Purpose
Case Materials	Case description, litigation claims, factual summary	Case cause ID, disputed issues, text labels	Task loading, background prompts	Evaluates the foundation of legal understanding
Evidence Objects	Images, recordings, contracts, transcripts	Evidence ID, source type, associated nodes	Cross-examination invocation, evidence display	Evaluates the quality of evidence application
Procedural Nodes	Court opening, evidence presentation, cross-examination, debate	Node sequence number, trigger conditions, completion marks	Process control, event monitoring	Evaluates the accuracy of procedural advancement
Scene Resources	Seats, evidence desk, display screen, file cabinet	Spatial coordinates, object type, interaction permission	3D rendering, operation response	Evaluates the rationality of spatial interaction
Role Resources	Judge, lawyer, litigant, clerk	Identity labels, permission sets, task scope	Role assignment, behavior constraint	Evaluates role performance

The resource structure in Table 2 enables the virtual simulation platform to complete case assembly before training, resource invocation during training, and behavior backtracking after training. Case materials and evidence objects enter the task chain through the node relationship, scene resources enter the rendering engine through spatial coordinates, and role resources enter the interactive control module through the permission set. This processing method ensures that each material viewing, evidence calling, node advancement and role response in the process of legal training can be recorded by the system, and provides a stable data basis for subsequent immersion effect evaluation.

3.2 Construction of immersive scene of legal training based on 3D interaction

The construction of immersive scene based on 3D interaction needs to map space, roles, evidence and procedural actions in legal training to a virtual environment. In this paper, the virtual court, mediation room and legal consultation room are used as the basic scene, and the judicial bench, evidence table, speaking table, filing cabinet and multimedia display screen are modeled as interactive objects. Each object contains spatial coordinates, collision boundaries, interaction permissions, and event callback fields, and students complete movement, gaze, evidence viewing, material dragging, and speech confirmation with the support of VR headsets and gamepads. The system deals with the visual space in the rendering layer, monitors the operation event in the interaction layer, and saves the behavior track in the data layer, so that the scene is not only a display model, but also a training environment that can be recorded, calculated and fed back.

In order to ensure that 3D objects, program nodes and role actions are bound in the same

coordinate frame, the system uses the spatial semantic matching function to calculate the scene assembly strength, as shown in the following equation:

$$S_o = \frac{\sum_{a=1}^{A} \theta_a R_{o,a} + \sum_{b=1}^{B} \vartheta_b C_{o,b}}{1 + |p_o - p_t| + \xi_o} \tag{8}$$

Here, S_o represents the scene assembly strength of the o 3D object. $R_{o,a}$ denotes the semantic association of the object with the legal task of type a; $C_{o,b}$ denotes the firing relationship between the object and the b program node. p_o represents the space coordinates of the object, p_t represents the coordinates of the target task area. Let ξ_o denote the penalty term generated by object permission conflict. The θ_a and ϑ_b are the weight coefficients. This formula is used to judge whether objects such as evidence tables, seats, and file cabinets should be loaded into the current training scene.

To illustrate the constitutive relationship of 3D scenes from resource assembly to interactive feedback, Fig. 1 shows the immersive scene construction process. The left side shows case resources and scene objects, the middle side shows 3D rendering, interactive monitoring and event binding, and the right side shows behavior data caching and feedback output. This process ensures that every evidence call, role movement and node confirmation in the virtual space can enter the subsequent evaluation module.

Figure 1. Immersive Legal Training Scene Construction Flow Based on 3D Interaction

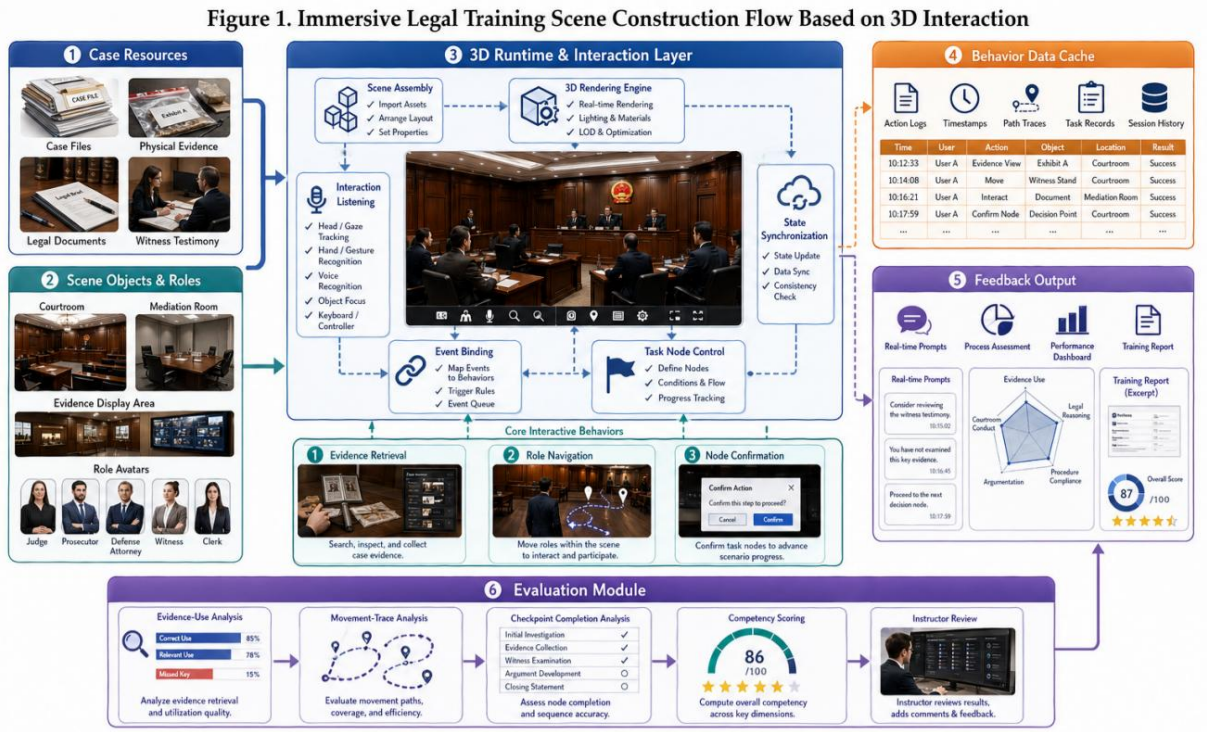

Figure 1: Construction process of immersive legal training scenario based on 3D interaction

In order to calculate the continuous degree of learners' interaction in three-dimensional space, the system fuses three kinds of signals, such as moving path, fixation stay and operation response, for dynamic estimation, as shown in the following equation:

$$I_t = \mathrm{ReLU}(\alpha_1 \Delta x_t + \alpha_2 g_t + \alpha_3 e_t - \alpha_4 n_t) \tag{9}$$

Here, I_t represents the interaction strength at time t; Δx_t represents the change of

spatial displacement of students at adjacent moments. g_t represents the fixation stay value for the evidence area, the speaker's table or the procedure cue area; e_t is the effective event firing number. n_t represents the noise term formed by invalid clicks, false triggers, and repeated movements. α_1 to α_4 are the tuning parameters. The formula can reflect whether students are continuously inside the legal training task, rather than staying in irrelevant areas.

To control the delay fluctuation between scene rendering and interactive response, the system constructs a real-time stability index based on the terminal frame rate and event return time, as shown in the following equation:

$$V = \exp\left[-\left(\frac{\tau_r - \tau_0}{\tau_0}\right)^2\right] \cdot \frac{f_r}{f_r + |f_r - f_0|} \tag{10}$$

where V represents the real-time stability of the virtual scene. τ_r is the actual interaction response time; τ_0 is the target response time set by the system. f_r stands for real-time rendering frame rate. f_0 represents the target frame rate. This formula simultaneously conconstricts the delay and frame rate, so that the evaluation results of the model are not disturbed by terminal stall, picture lag or event return delay.

The above 3D interactive construction method makes the legal training scene have a technical foundation that can be rendered, operated, recorded and evaluated. The scene objects are assembled by spatial semantic matching, the student behavior is recorded by the interaction intensity index, and the system running state is constrained by the real-time stability index. As a result, virtual courts, mediation rooms, and legal consulting rooms are no longer just visual scenes, but computational environments that can host program training, evidence manipulation, character interaction, and feedback generation. The design provides a stable entrance for the subsequent training behavior collection process, and also provides a continuous and traceable data source for the immersion effect evaluation module.

3.3 Training behavior collection and process modeling for role tasks

The training behavior collection process oriented to role tasks needs to be carried out around the task boundaries of judges, lawyers, parties, clerks and mediators. Different roles have different operation permissions, speaking objects and program responsibilities in the virtual scene. The system can not only record a unified click log, but bind the role identity to the behavior event. In this paper, the training behaviors are divided into six categories: voice expression, evidence manipulation, eye attention, spatial movement, node confirmation and collaborative response. Each type of behavior has a timestamp, role label, scene location and program state. After entering the behavior cache, the model encodes the sequence and determines the validity.

In order to present the flow of training behavior from collection to model input, Fig. 2 shows the collection process of role task behavior. The upper part of the figure is the multimodal acquisition end, including voice, sight, action and log. The middle part is role permission verification and event slicing. The lower part is the behavior sequence cache and evaluation interface. This structure enables the system to track the specific performance of students in different roles, and provides a source of evidence for teachers to review.

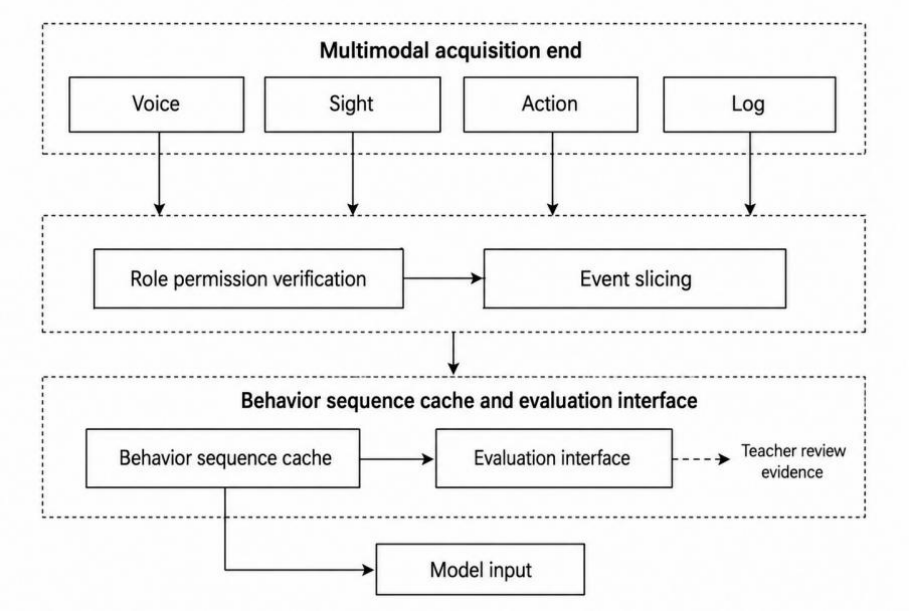

Figure 2: Behavior collection process of legal training oriented to role tasks

In order to unify the behavior events of different roles, the system converts the multi-source signals into event vectors with role constraints, and calculates the event validity score, as shown in the following equation:

$$E_t^{(r)} = \text{sigmoid}(\beta_1 L_t^{(r)} + \beta_2 Y_t^{(r)} + \beta_3 K_t^{(r)} - \beta_4 \Omega_t) \tag{11}$$

where $E_t^{(r)}$ represents the event validity score of role r at time t. $L_t^{(r)}$ denotes the legal operation mark of the role; $Y_t^{(r)}$ represents the matching degree between the semantic response and the current task; $K_t^{(r)}$ represents the evidence or node operation hit degree; Ω_t represents the interference term formed by ultra vires, false contacts and repeated responses. β_1 to β_4 are the weight parameters. The formula is used to filter invalid behaviors and retain the training data that can reflect the performance of the role.

In order to depict the stability of students 'behavior sequence in the same role task, the system applies joint constraints on event interval, program sequence and behavior integrity, as shown in the following equation:

$$B_r = \frac{\sum_{t=1}^{T_r} \zeta_t E_t^{(r)} \cdot \Gamma_t}{T_r + \sum_{t=1}^{T_r} |\Delta t_t - \bar{\Delta t}|} \tag{12}$$

Here, B_r denotes the behavior sequence stability of role r; T_r denotes the number of valid events generated by the role; Let ζ_t denote the task weight of the t event; Γ_t denotes whether the sequence of events meets the program requirements; $\Delta t_t t$ represents the interval between adjacent events. The $\bar{\Delta t}$ represents the standard task interval. The formula can identify situations such as too slow speech, disjointed evidence call, and missing node confirmation, so that the process collection has time continuity.

In order to evaluate the task completion quality under multi-role collaboration, the system combines role contribution, interaction response and program closure state into the same calculation equation, as shown in the following equation:

$$C_q = \frac{1}{R}\sum_{r=1}^{R}(\omega_r B_r + \pi_r J_r + \varphi_r F_r) - \varepsilon D_q \tag{13}$$

where C_q represents the collaborative completion quality of the q virtual training task. R represents the number of participating roles; J_r indicates the interaction response quality between the character and other objects; F_r indicates that the role is responsible for the completion status of the node; D_q indicates the degree of program deviation occurring in the task; ω_r,π_r,φ_r,ε are the regulation coefficients. The formula can unify individual behavior, role collaboration and process completion into the same evaluation result.

The collection process of role task behavior makes the virtual training data transform from scattered records to structured samples. Voice records are no longer just speech texts, but can correspond to role identities and program nodes. Gaze trajectories are no longer just stopping points, but can point to evidence attention and scene engagement. The action log is no longer just the number of clicks, but can reflect the progress of the task and the sequence of the program. After event slicing, role constraints and sequence coding, the system can form a trainable data set to support the operation of subsequent legal training behavior recognition, immersion state evaluation and feedback modules. This process also ensures that teachers can trace back the evaluation basis and view the speech, operation and node records corresponding to each scoring result.

3.4 Evaluation of immersion effect and feedback generation based on multimodal data

The immersion evaluation and feedback module based on multimodal data takes the character task process as input, fuses speech expression, gaze stay, spatial action, evidence operation and program node log, and outputs the immersion state score, feedback type and review mark. The system divides event segments such as opening statements, evidence cross-examination, mediation response and consultation reply according to a uniform timestamp, so that the evaluation results can correspond to specific legal tasks.

In order to make the immersion evaluation absorb semantic, visual, action and log evidence at the same time, the system constructs a fusion scoring function with modal belief correction to avoid single modal fluctuations affecting the overall judgment, as shown in the following equation:

$$R_i = \mathrm{sigmoid}\left(\sum_{m=1}^{4}\omega_m \rho_{i,m} F_{i,m} + \lambda A_i - \xi N_i\right) \tag{14}$$

where, R_i represents the immersion effect score of the i training segment. $F_{i,m}$ represent modal features such as speech semantics, eye attention, spatial action and operation log. Let $\rho_{i,m}$ denote the confidence of each modality. ω_m denotes the modal weights; A_i represents the inter-modal consistency compensation term; N_i denotes the abnormal fragment noise. This formulation compreses multi-source sensing data into a unified evaluation space, which can simultaneously reflect role engagement, program continuity and evidence manipulation quality.

The feedback generation needs to be based on the gap between the current performance of students and the standard task, and the system takes the evaluation score, procedure deviation, evidence hit and response delay into the feedback strength calculation process, as shown in the following equation:

$$G_j = \exp(-\|T_j - X_i\|_2) \cdot (1 + \eta E_i) - \tau P_i - \nu L_i \tag{15}$$

Here, G_j represents the generation strength of the j feedback node. T_j denotes standard task requirements; X_i represents the student's current performance vector; P_i stands for program deviation. E_i is the evidence hit rate. L_i is the feedback delay; η,τ,ν are the regulation coefficients. The formula is used to judge the system push program reminder, evidence supplement, role expression correction or teacher review mark, so that the feedback content maintains a corresponding relationship with the specific task link.

In order to ensure that the model output can enter the subsequent training update, the system further calculates the evaluation credibility, and synchronously writes the teacher review results and the model distribution into the sample cache, as shown in the following equation:

$$K_i = 1 - \frac{D_{KL}(P_i^{model} \| P_i^{teacher})}{1 + \theta H_i + \zeta O_i} \tag{16}$$

Here, K_i represents the credibility of the evaluation results. P_i^{model} represents the model output probability distribution. $P_i^{teacher}$ represents the distribution of teacher review labels; H_i is the historical consistency coefficient. O_i represents the number of abnormal interaction fragments. This formula is used to determine whether the evaluation results are directly entered into the training file or whether they are transferred to the manual review queue.

After processing by this module, the speech, gaze, operation and node advancement in legal training can form a continuous evidence chain. The evaluation of immersion effect is no longer a subjective experience description, but a process data that can be calculated, backtracked and corrected, which provides a basis for the next round of virtual scene adjustment and task feedback generation, and enhances the stability of feedback closed-loop.

4 System deployment and test environment

4.1 Legal Training Environment and terminal deployment scheme

The virtual simulation environment of legal training is deployed around four tasks: trial, evidence examination, mediation and consultation. The system adopts a hierarchical structure of "student end-edge node-platform server-teacher review end". The student side is responsible for 3D scene entry, role identity loading and interactive action collection. The edge nodes are responsible for voice noise reduction, eye movement slicing, handle trajectory caching and log compression. The platform server completes scene scheduling, data archiving and evaluation interface calling. When the terminal was deployed, the virtual court, mediation room and consultation room shared the same resource library, and only role permissions, evidence objects and program nodes were switched for different tasks to ensure the uniform format of training data. The main deployment parameters are shown in Table 3. And keep version consistency, data consistency and interface consistency between different scenarios.

Table 3: Terminal deployment parameters of virtual simulation environment for legal training

Configuration Item	Specification Parameters	Quantity	Functional Role
VR Headset Terminal	4K display, 90 Hz refresh rate, 6DoF positioning	24	Supports scene access and immersive display
Eye-Tracking Module	120 Hz sampling rate, gaze-point error ≤ 0.8°	24	Records gaze data in evidence areas and role areas
Interactive Controller	Millimeter-level positioning with haptic feedback	48	Supports evidence invocation, node confirmation, and role movement
Edge Computing Node	RTX A2000, 32 GB memory	6	Performs data preprocessing and event caching
Platform Server	64-core CPU, 128 GB memory, dual GPUs	1	Supports scene scheduling, model inference, and data archiving
Teacher Review Terminal	Web-based visualization interface with trajectory replay	4	Supports score verification and abnormal segment inspection

The device configuration in the table ensures the synchronous operation of virtual space rendering, behavior acquisition, and feedback display. The VR head display provides spatial perspective and immersive screen, the eye movement module records the gaze data of the evidence area, the speaker's table and the program prompt area, and the handle and the positioning base station record evidence call, role movement and node confirmation. Edge computing nodes convert raw data into structured events and upload them to the platform server. The teacher review end receives model scores, abnormal segments and student task trajectories for verifying the training process. The deployment scheme enables the legal training environment to have low-latency acquisition, continuous interaction and traceable evaluation capabilities.

4.2 Model training and platform integration scheme

The configuration of virtual simulation platform integration and model training revolves around scene engine, data interface, fusion model and feedback module. The platform uses Unity 3D engine to load resources of virtual court, mediation room and legal consultation room, and the back-end receives voice, eye movement, handle trajectory and operation log with Python service. The training samples were collected from 12 rounds of training records of 96 learners, and a total of 2840 labeled samples were formed, covering court speech, evidence call, program advancement, role collaboration and invalid operation. The model training and platform interface configuration are shown in Table 4. Anonymization, time alignment and abnormal segment removal were completed before data entered the training, and the model version was saved by rounds to facilitate the comparison of recognition stability, feedback delay and task adaptation changes under different parameters.

Table 4: Virtual simulation platform integration and model training configuration

Module Name	Configuration Content	Parameter Settings	Function Description
Scene Engine	Unity 3D simulation platform	90 Hz rendering, HDRP pipeline	Loads the virtual courtroom, mediation room, and legal consultation room scenes
Data Interface	REST API + WebSocket	20 ms polling interval	Receives behavioral data and returns feedback results
Training Dataset	Multimodal legal training samples	2,840 samples with five behavior labels	Supports behavior recognition and immersion evaluation
Fusion Model	Multimodal attention network	Four-channel input of speech, eye movement, motion, and logs	Outputs behavior categories and immersion states
Loss Function	Classification loss + regression loss + consistency constraint	Weight ratio of 1:0.4:0.2	Jointly optimizes recognition and scoring results
Training Resources	Dual-GPU server	Batch size 32, 80 training epochs	Supports offline training and model updating

The configuration in the table keeps the model training on the same data link as the virtual simulation platform. The multimodal fusion network receives four types of input, and outputs the behavior category, immersion state and feedback label after temporal coding and attention weighting. At the training end, cross-entropy loss, scoring regression loss and consistency constraint are jointly updated to keep the recognition results corresponding to the teacher review records. The platform uses REST interface to receive inference requests, and WebSocket returns real-time feedback.

5 Analysis of results

5.1 Accuracy analysis of legal training behavior recognition

The legal training behavior recognition experiment selected five types of behaviors as test objects: trial statement, evidence submission, cross-examination response, procedure confirmation and collaborative discussion. The model input consists of voice transcription, eye movement stay, handle trajectory and operation log, and the training samples are from 2840 valid training segments of 96 learners. In order to observe the differences of different legal acts in multi-modal recognition, this paper uses accuracy, recall and F1 value as evaluation indicators.

As shown in Fig. 3, the recognition results of program confirmation, evidence submission and trial statement were relatively high, in which the accuracy of program confirmation reached 95.2%, the recall rate was 94.1%, and the F1 value was 0.946. The accuracy of evidence submission is 94.6%, and the F1 value is 0.941. The accuracy of the trial statement is 93.8%, and the F1 value is 0.933. The accuracy of collaborative discussion is 87.9%, and the F1 score is 0.872, which is lower than other behavior categories.

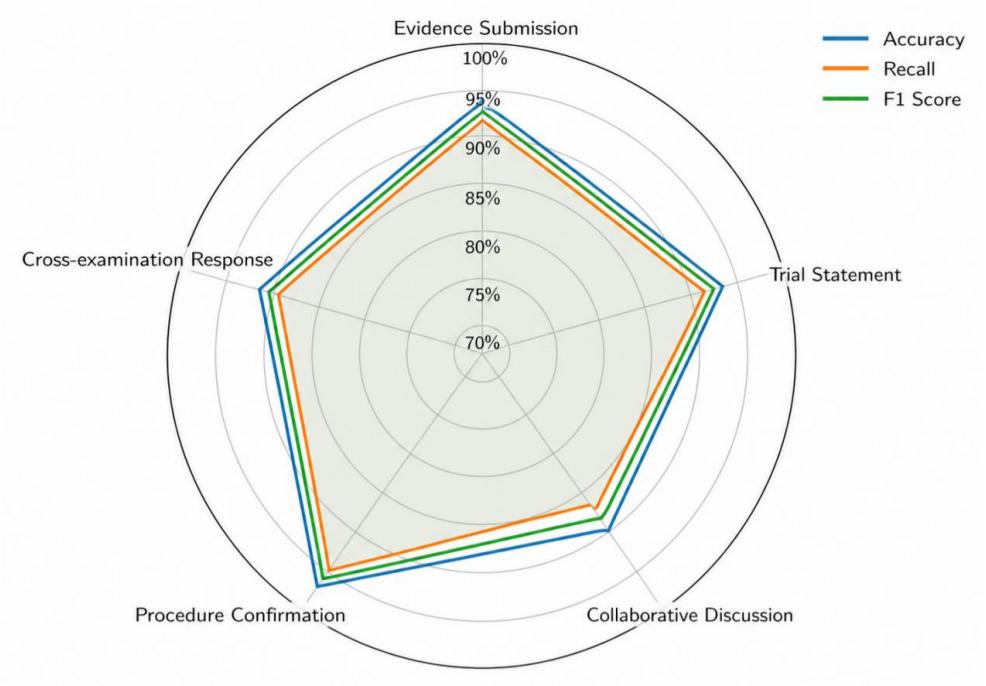

Figure 3: Radar chart of behavior recognition performance in legal training

The results show that the more structured the legal training behavior is, the easier it is to be stably identified by the system. Based on the prediction results of five types of legal training behaviors in the test set, the overall behavior recognition accuracy of the model is 92.4%and the weighted F1 value is 0.887. The program confirmation has clear node marks, and the evidence submission depends on the evidence number, call time and role permissions, which can be jointly supported by operation logs and action traces. Collaborative discussion involves multiple people switching rounds, speech overlap and eye shift, and the boundary is relatively fuzzy. This result shows that the multimodal fusion model is suitable for dealing with clear legal behaviors of program nodes, and it still needs to enhance voice separation and role relationship modeling for open collaborative behaviors.

5.2 Modeling and analysis of immersion state

The immersive interaction status assessment was used to judge the learner's role engagement in the virtual courtroom, mediation room, and counseling room. In this paper, the immersion states are divided into five categories: high immersion, stable immersion, fluctuating immersion, low immersion and disengagement. The input features include gaze focus time, character speech frequency, spatial movement continuity, evidence area stay time and program response delay. The model output is compared with the teacher review label to form a state classification matrix.

As shown in Fig. 4, 82 samples were correctly identified in the high immersion state, 79 in the stable immersion state, 73 in the fluctuating immersion state, 76 in the low immersion state, and 81 in the disengaging state. The number of samples where stable immersion was misclassified as fluctuating immersion was 8 and fluctuating immersion was misclassified as stable immersion was 9.

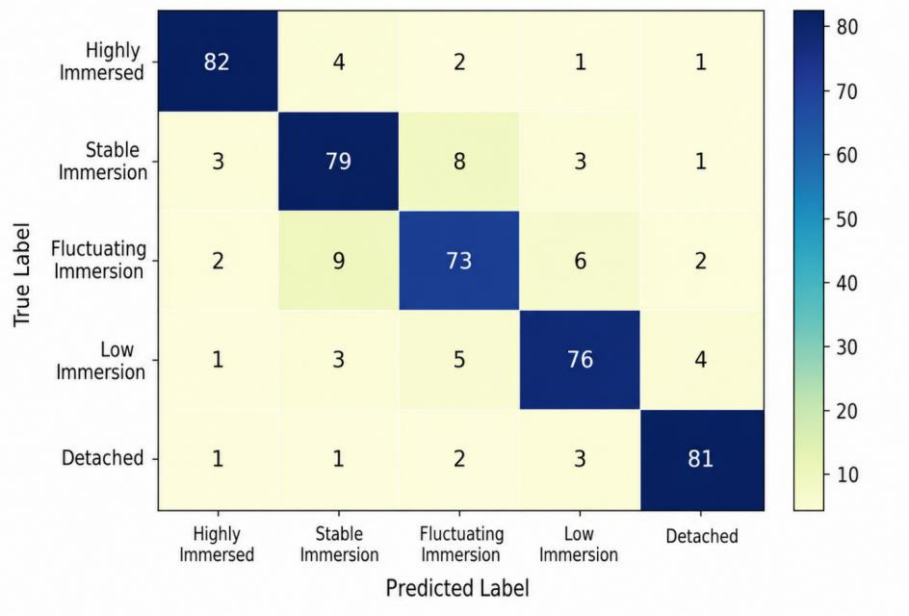

Figure 4: Heat map of confusion matrix for immersive interaction state recognition

The results show that the classification boundary of extreme immersion state is clear. The correct recognition samples of the highly immersed and disengaged states reached 82 and 81, respectively, indicating that the system was able to capture typical signals such as continuous gaze, timely response, continuous action, and obvious departure. There is a crossover between stable and fluctuating immersion. The main reason is that both types of states may have phasic gaze shifts and response delays, and it is difficult to completely distinguish between decreased engagement and task switching in a single episode. The recognition results of low immersion samples are relatively stable, which indicates that discrete gaze, long operation interval and insufficient program response can form a clearer basis for discrimination. On the whole, the multimodal immersion state modeling can better reflect students 'investment degree in virtual legal training, but the critical state still needs to be corrected by combining more fine-grained semantic response and role task progress.

5.3 Virtual training task completion and adaptation effect analysis

The virtual training task completion and adaptation experiment is used to analyze whether the system push task conforms to the current role state and program progress of students. This paper sets up five types of tasks: simulated prosecution, evidence cross-examination, court debate, mediation and consultation, and legal advice. For each type of task, the matching accuracy, completion rate, program compliance rate and response time are recorded. Considering the discrete distribution of the performance of different students on the same task, box plots are used to present the median, upper and lower quartiles, and outlier ranges of task completion rates.

As shown in Fig. 5, the median completion rate of legal consultation tasks is 90.2%, the upper quartile reaches 93.5%, the median completion rate of evidence cross-examination tasks is 88.5%, the simulated prosecution is 86.8%, the mediation and consultation is 84.7%, and the trial debate is 82.9%. The lower quartile of court argument was 78.6%, which was the most discrete, indicating that this task was significantly affected by immediate expression, opinion organization and role response.

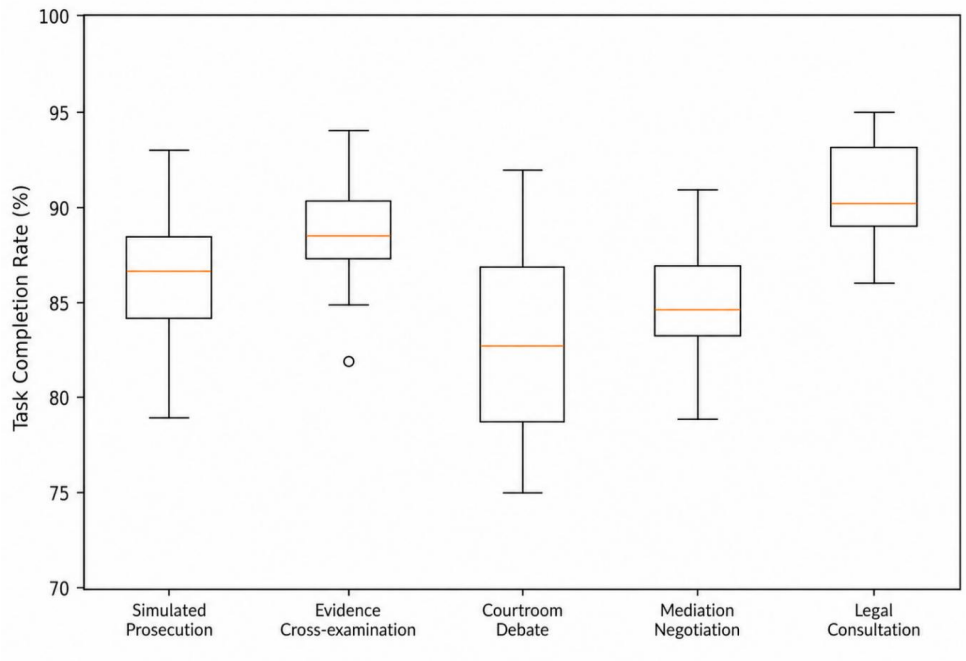

Figure 5: Box plot of virtual training task completion rate

The results show that tasks with clear structure and stable interactive objects are more likely to achieve a high completion rate. The median completion rates of legal advice and evidence cross-examination tasks were both over 88%, indicating that the system was able to push moderately difficult training content based on student historical performance. The completion rate of trial debate is the lowest, and the distribution range is wider, reflecting that multi-round attack-defense expression has higher requirements for language organization and instant invocation of evidence. This result proves that the task adaptation module needs to be dynamically adjusted according to the complexity of the cause of action, the role ability and the program load.

5.4 Effect analysis of multimodal interactive data fusion

The effect analysis of multimodal interactive data fusion was used to test the support ability of different input combinations for legal training evaluation results. Four combinations of speech + log, sight + action, speech + sight + action, and speech + sight + action + log were set up in the experiment, and the behavior recognition accuracy, immersion evaluation agreement rate and feedback response score were recorded in 10 repeated rounds of experiments. In order to show the dispersion degree and stability range of the comprehensive performance distribution under different modal combinations, the three indicators are normalized to calculate the comprehensive fusion score, and the violin plot is used to show the distribution form of each combination.

As shown in Fig. 6, the four-modal fusion group had the most concentrated distribution, and the overall position was the highest, with a median of 91.8 and a main distribution interval of 90.4-93.2. The median of the speech + sight + movement group was 88.7, and the distribution interval was 86.9-90.6. The median of the gaze + action group was 84.9, and the distribution range was 82.8-86.4. The median of voice + log group was the lowest, 82.6, and the distribution interval was 80.7-84.1. The narrow violin width of the four-modal fusion group indicates that the fluctuation between repeated experiments is smaller and the system stability is higher.

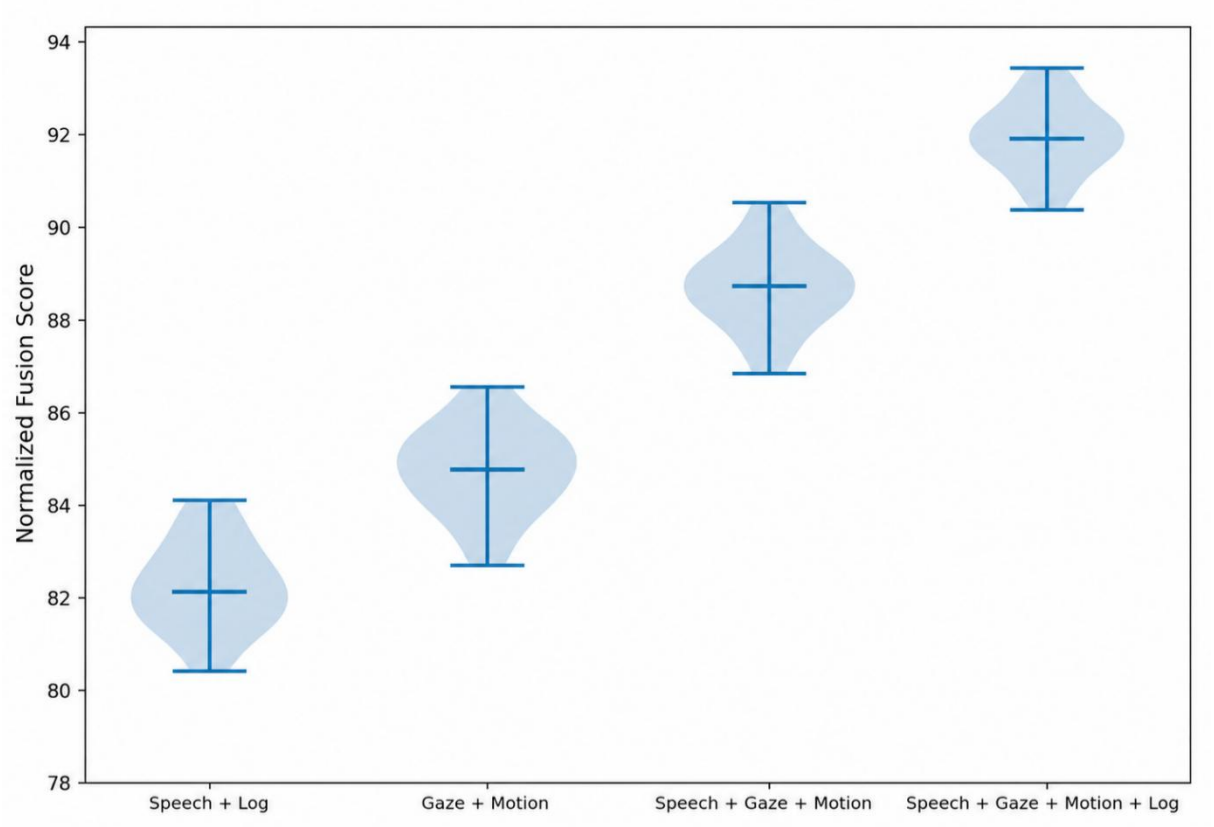

Figure 6: Violin plot of multimodal interaction data fusion effect

The results show that multi-modal information is not simply superimposed, but forms a more complete representation of legal training behavior through temporal alignment and attention weight allocation. Voice data can support the judgment of statement quality, gaze data can reflect the position of evidence attention, action trajectories can describe the spatial interaction process, and log data can record program progress and node confirmation. After the four-modal fusion, the comprehensive score was increased by more than 7 points compared with the dual-modal combination, and the distribution was more concentrated, indicating that the system obtained a more stable input basis in the three levels of behavior recognition, immersion evaluation and feedback generation. This result indicates that the complete multimodal link is more suitable for complex interaction analysis in legal training scenarios.

5.5 Influence analysis of virtual simulation perception accuracy on evaluation results

The influence of perception parameters on the training evaluation results is mainly reflected in the detail retention of multimodal data, the continuity of time series and the stability of model discrimination. In the experiment, three parameters including speech sampling rate, eye movement sampling frequency and action update rate were selected, five sets of terminal configurations were set, and the behavior recognition accuracy, immersion evaluation consistency and task adaptation accuracy were calculated respectively. In order to avoid purely presenting point-like changes, this paper uses ladder partition plots to show the performance hierarchy under different parameter configurations, so that the differences between perceptual parameters and evaluation results are clearer.

As shown in Figure7, the low-configuration group achieved 86.2% behavior recognition

accuracy, 84.7% immersion evaluation agreement, and 83.9% task adaptation accuracy using 16kHz speech sampling rate, 60Hz eye movement sampling rate, and 30Hz action update rate. When the speech sampling rate is increased to 22.05kHz, the three indicators rise to 88.5%, 86.8% and 85.7%, respectively, indicating that the retention of speech details can improve the recognition effect of court statements, cross-examination responses and other behaviors. When the parameters are further increased to 22.05kHz, 120Hz and 60Hz, the behavior recognition accuracy reaches 90.1%, the immersion evaluation consistency rate reaches 88.9%, and the task adaptation accuracy reaches 87.6%. The enhancement of eye movement trajectory and action continuity makes the evidence attention and program following state easier to be captured by the model.

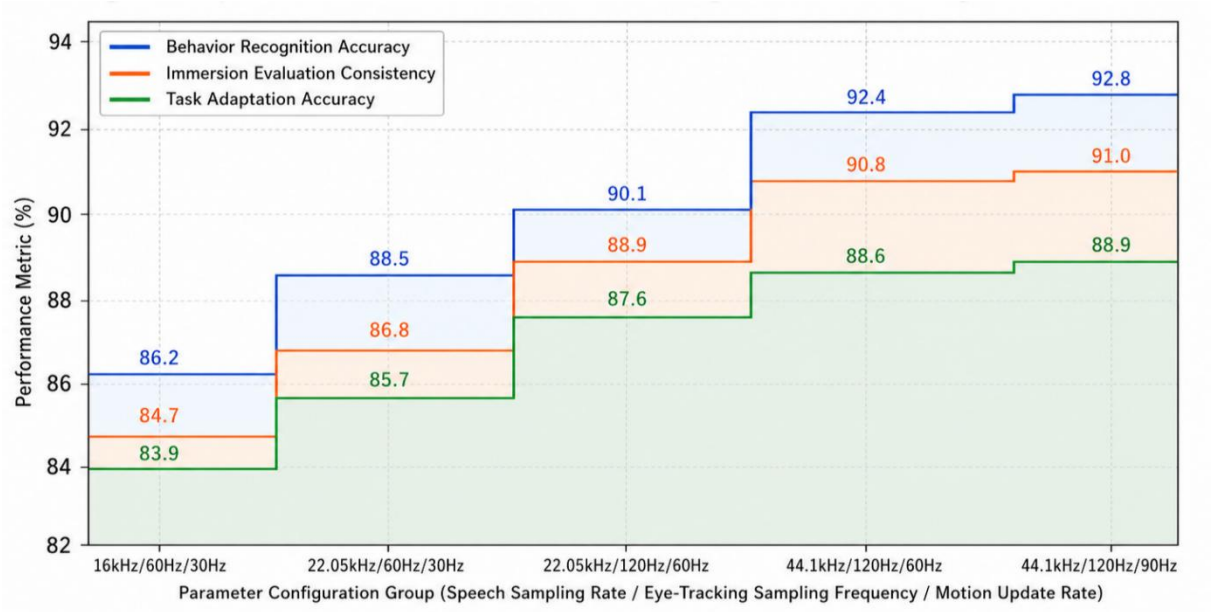

Figure 7: Ladder partition diagram of the influence of perceptual parameters on the results of the training evaluation

Under the configuration of 44.1kHz, 120Hz and 60Hz, the three indicators reach 92.4%, 90.8% and 88.6%, respectively, and the system enters a relatively stable evaluation interval. After increasing the action update rate to 90Hz, the behavior recognition accuracy was only increased to 92.8%, the immersion evaluation consistency rate was 91.0%, and the task adaptation accuracy rate was 88.9%. The increase was significantly reduced. This result shows that the perceptual performance of the legal training virtual simulation platform does not depend on a single hardware parameter stack, but depends on the matching degree between speech, eye movement and action signals. Reasonable configuration of terminal parameters can ensure the quality of input data, and control the calculation load and feedback delay, so that the training evaluation results maintain good stability and reusability.

6 Discussion

Experimental results show that the immersive virtual simulation platform can stably support behavior recognition and process evaluation in legal training. The accuracy rate of program confirmation is 95.2%, the accuracy rate of evidence submission is 94.6%, and the accuracy rate of trial statement is 93.8%, which indicates that the behavior with clear nodes, role permissions and operation logs is easier to form stable characteristics. The accuracy of

collaborative discussion is 87.9%, which is mainly affected by the overlapping of multiple speeches, gaze switching and response sequence change. In the multimodal fusion experiment, the comprehensive score of voice, sight, action and log after joint input reached 91.8, which was higher than that of the voice + log group (82.6), indicating that the complete perception link could make up for the lack of understanding of legal behavior by a single modality. In the evaluation of immersion state, the boundary between high immersion and out of state is clear, and there is a cross between stable immersion and fluctuating immersion, which reflects that the input state in virtual training has the characteristics of continuous change. Perceptual parameter analysis shows that 44.1kHz speech, 120Hz eye movement and 60Hz action update can maintain stable output, and the increase is limited after increasing the action update rate. The system is suitable for legal training tasks with clear procedure nodes and clear evidence operation. Open-ended debate, complex mediation, and emergent question answering still need to introduce mechanisms for semantic understanding, role relationship modeling, and cross-scenario transfer.

7 Conclusions

This paper constructs an immersive virtual simulation teaching model around the legal training scene, which integrates the virtual court, mediation room and legal consultation room into a unified platform, forming a complete link of scene modeling, role interaction, behavior collection, immersion evaluation and feedback generation. Based on three-dimensional interaction, the system converts case materials, evidence objects, program nodes and role permissions into callable data resources, and records the specific performance of students in the training process through voice, eye movement, action track and operation log. The multimodal data fusion mechanism makes legal training no longer stop at outcome scoring, but can present the association between process behavior, program execution and role engagement. There are still some limitations in this paper. The scene types mainly focus on typical legal training tasks, and the coverage of complex cases, multi-role conflicts and open debate scenes is insufficient. Semantic evaluation still relies on rule labeling and teacher review, and the model's understanding of implicit legal reasoning and complex language expressions needs to be enhanced. In the future, the scale of the case base can be expanded, the legal knowledge graph and large language model are introduced to assist semantic parsing, and the lightweight edge reasoning and privacy protection mechanism are combined to further enhance the cross-scenario generalization ability, real-time feedback stability and data security level of the system.

Funding

This work was supported by Ministry of Education Industry-Academia Cooperation and Collaborative Education Project (231100874090918) "Research on Virtual Simulation Experimental Course of Civil Litigation Evidence Collection and Application

Author's Profile

Bingcui Jian was born in Fushun，liaoning, P.R. China, in 1978. I obtained a master's degree from Jilin University in China. I am currently working at the School of Law, Jianghan University. My main research direction is law, legal education and community governance. janewuhan@jhun. edu.cn

References

[1] Hidayah N P, Wicaksono G W, Perdana M I, et al. Implementation of Virtual Reality Moot Court for Simulation and Procedural Law Learning of the Constitutional Court[J]. JOIV: International Journal on Informatics Visualization, 2024, 8(4): 2444-2451.

[2] Alshaer A. Virtual reality in training: a case study on investigating immersive training for prisoners[J]. International Journal of Advanced Computer Science and Applications, 2023, 14(10).

[3] Barbe H, Müller J L, Siegel B, et al. An open source virtual reality training framework for the criminal justice system[J]. Criminal Justice and Behavior, 2023, 50(2): 294-303.

[4] King S, Boyer J, Bell T, et al. An automated virtual reality training system for teacher-student interaction: A randomized controlled trial[J]. JMIR serious games, 2022, 10(4): e41097.

[5] Calandra D, De Lorenzis F, Cannavò A, et al. Immersive virtual reality and passive haptic interfaces to improve procedural learning in a formal training course for first responders[J]. Virtual Reality, 2023, 27(2): 985-1012.

[6] Boel C, Rotsaert T, Valcke M, et al. Applying educational design research to develop a low-cost, mobile immersive virtual reality serious game teaching safety in secondary vocational education[J]. Education and Information Technologies, 2024, 29(7): 8609-8646.

[7] Strojny P, Dużmańska-Misiarczyk N. Measuring the effectiveness of virtual training: A systematic review[J]. Computers & Education: X Reality, 2023, 2: 100006.

[8] Conrad M, Kablitz D, Schumann S. Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings[J]. Computers & Education: X Reality, 2024, 4: 100053.

[9] Jongbloed J, Chaker R, Lavoue E. Immersive procedural training in virtual reality: A systematic literature review[J]. Computers & Education, 2024, 221: 105124.

[10] Thomann H, Zimmermann J, Deutscher V. How effective is immersive VR for vocational education? Analyzing knowledge gains and motivational effects[J]. Computers & Education, 2024, 220: 105127.

[11] Khorasani S, Syiem B V, Nawaz S, et al. Hands-on or hands-off: Deciphering the impact of interactivity on embodied learning in VR[J]. Computers & Education: X Reality, 2023, 3: 100037.

[12] Wheeler S G, Hoermann S, Lukosch S, et al. Design and assessment of a virtual reality learning environment for firefighters[J]. Frontiers in Computer Science, 2024, 6: 1274828.

[13] Clay C J, Budde J R, Hoang A Q, et al. An evaluation of the effectiveness of immersive virtual reality training in non-specialized medical procedures for caregivers and students: a brief literature review[J]. Frontiers in Virtual Reality, 2024, 5: 1402093.

[14] Hjellvik S, Mallam S. Training transfer validity of virtual reality simulator assessment[J]. Virtual Reality, 2024, 28(4): 165.

[15] Laine J, Rastas E, Seitamaa A, et al. Immersive virtual reality for complex skills training: content analysis of experienced challenges[J]. Virtual Reality, 2024, 28(1): 61.

[16] Cabrera-Duffaut A, Pinto-Llorente A M, Iglesias-Rodríguez A. Immersive learning platforms: analyzing virtual reality contribution to competence development in higher education—a systematic literature review[C]//Frontiers in Education. Frontiers Media SA, 2024, 9: 1391560.

[17] Coban M, Bolat Y I, Goksu I. The potential of immersive virtual reality to enhance learning: A meta-analysis[J]. Educational Research Review, 2022, 36: 100452.

[18] Petersen G B, Petkakis G, Makransky G. A study of how immersion and interactivity drive VR learning[J]. Computers & Education, 2022, 179: 104429.

[19] Makransky G, Mayer R E. Benefits of taking a virtual field trip in immersive virtual reality: Evidence for the immersion principle in multimedia learning[J]. Educational psychology review, 2022, 34(3): 1771-1798.

[20] Mayer R E, Makransky G, Parong J. The promise and pitfalls of learning in immersive virtual reality[J]. International Journal of Human–Computer Interaction, 2023, 39(11): 2229-2238.