

Design and Preliminary Application Study of an AI Driven Personalized VR Rehabilitation Support System for Children with Autism

Guangke Qi ¹, Zhiliang Xia ^{1,2,*}, Yinhui Yin ³

¹Wenzhou Polytechnic University, Wenzhou 325035, China

²Blch Pneumatic Science & Technology Co., Ltd., Wenzhou 325000, China

³Zhejiang Economic & Technical Development Zone, Zhejiang P.R.China 325025, China

Abstract

Objective To address insufficient task adaptation, fragmented process information, and a lack of transparency in the basis for training adjustments for children with autism, this study designs an AI driven personalized VR rehabilitation support system and establishes a preliminary application evaluation framework. **Methods** Drawing on TEACCH structured teaching to organize tasks, the system integrates the DeepSeek large language model, Unity virtual environments, and MQTT communication to support ability profiling, strategy recommendation, training interaction, and coordination across multiple terminals. Candidate strategy screening, output validation, and therapist confirmation form a closed loop for task adjustment. Technical system tests and a small scale feasibility study are used to evaluate operational performance, training acceptability, and task performance. **Results** A system design covering perceptual, communication, and social tasks was developed, together with four functional modules—Explore at Your Own Pace, Fun Practice Zone, Everyday Scenarios Town, and Peer Play Community—and corresponding evaluation indicators. **Conclusion** This study proposes a VR rehabilitation support system that combines structured tasks, AI recommendations, and professional supervision. Actual system feasibility and training related outcomes must be assessed using application data; no definitive conclusion about rehabilitation effectiveness can yet be drawn.

Keywords

Autism; virtual reality; large language model; personalized training; structured teaching; coordination across multiple terminals.

1. Introduction

Children with autism differ in their communication methods, activity preferences, sensory experiences, and understanding of tasks. The same training objective may therefore require different prompting methods, task pacing, and levels of interaction difficulty. For a rehabilitation support system, the central challenge is to translate these differences into specific tasks while retaining an understandable and traceable basis for professional adjustments. Simply increasing the number of environments does not ensure personalization. Likewise, data generated during training cannot readily support ongoing adjustments unless they feed back into task planning.

Virtual reality can provide repeatable interactive situations and a technological platform for investigating cognitive and social tasks. Didehbani et al. investigated VR training for social cognition in children [1]. A review of adaptive VR systems by Maddalon et al. showed that existing studies have used adaptation strategies such as switching task levels, adjusting

feedback, and responding to behavioral or physiological signals [2]. These studies suggest that evaluations of VR systems should describe training objectives, adaptation signals, and adjustment mechanisms together. Findings obtained with different device types and study samples should not be generalized directly to all children.

TEACCH structured teaching provides a reference for organizing environments, scheduling activities, and presenting tasks. Drawing on its approach to task organization, this study represents training as activities with explicit goals, visible steps, and clear completion criteria. A relevant meta-analysis reported varying degrees of change across outcomes while emphasizing the limited amount of evidence [3]. Accordingly, the inclusion of TEACCH elements is not treated here as evidence that the system's therapeutic effectiveness has already been established.

Supported by a rehabilitation support system development project in Wenzhou, this study addresses three questions: how to translate children's abilities and support needs into executable VR task parameters; how to constrain large language model recommendations through existing strategies and human supervision; and how to retain task performance and the basis for adjustments during operation across multiple devices. The paper presents the system architecture, key modules, and methods for preliminary application evaluation. Technical usability and practical feasibility are the primary evaluation goals, while changes in task performance are treated as exploratory outcomes.

2. Requirements Analysis and Overall System Design

2.1. User Requirements and Scope of Application

The system serves three user roles: children, therapists, and parents. The child interface focuses on simple tasks, selectable activities, and clear feedback. The therapist interface supports ability assessment, task confirmation, real time supervision, and report review. The parent interface allows access to authorized training summaries and feedback on everyday performance. Children's support needs are described by task dimension rather than assigning each child a fixed overall category of low, medium, or high ability.

The system is positioned as an aid to rehabilitation training under professional guidance. It does not perform diagnosis or automatically determine clinical status from a single behavior. Institutions and families may share task definitions and record structures, but equipment, supervision arrangements, and operating permissions for home use require separate confirmation. A headset is not the default entry point for every child. Actual eligibility should be determined according to device suitability requirements and professional assessment, and the use of alternative interaction methods, such as tablets, should be documented.

2.2. Mapping Teaching Requirements to Task Functions

Structured tasks should help children understand the current activity, the sequence of actions, the completion criteria, and what happens next. The system therefore links activity schedules, step prompts, and outcome feedback to objects in the virtual environment, preventing decorative content from obscuring training objectives. The same objective may be supported by different contextual resources, while task success criteria, prompting levels, and data fields remain consistent to enable comparisons across environments.

Table 1 Mapping Structured Teaching Requirements to System Functions

Design requirement	System function	Recorded information
Predictable activity schedule	Visual schedule and preview of the next activity	Activity sequence and transitions

Design requirement	System function	Recorded information
Understandable task steps	Stepwise interaction and visual prompts	Step outcomes and prompting levels
Recognizable completion criteria	Completion indicators and brief feedback	Completion status and time taken
Adjustable support	Configurable prompt frequency and difficulty	Parameters before and after adjustment and reasons

2.3. System Architecture and Data Flow

The overall architecture comprises terminal interaction, application services, intelligent support, and data management. The VR terminal presents situations and collects interaction events; the tablet terminal supports supervision and task management; and the robot terminal delivers controlled prompts or interactions. Application services manage training sessions and permissions, the intelligent support module produces candidate plans, and the data layer stores task resources, strategy versions, ability profiles, and event records. Figure 1 illustrates the relationships between modules at the design level and does not imply that all modules have been deployed.

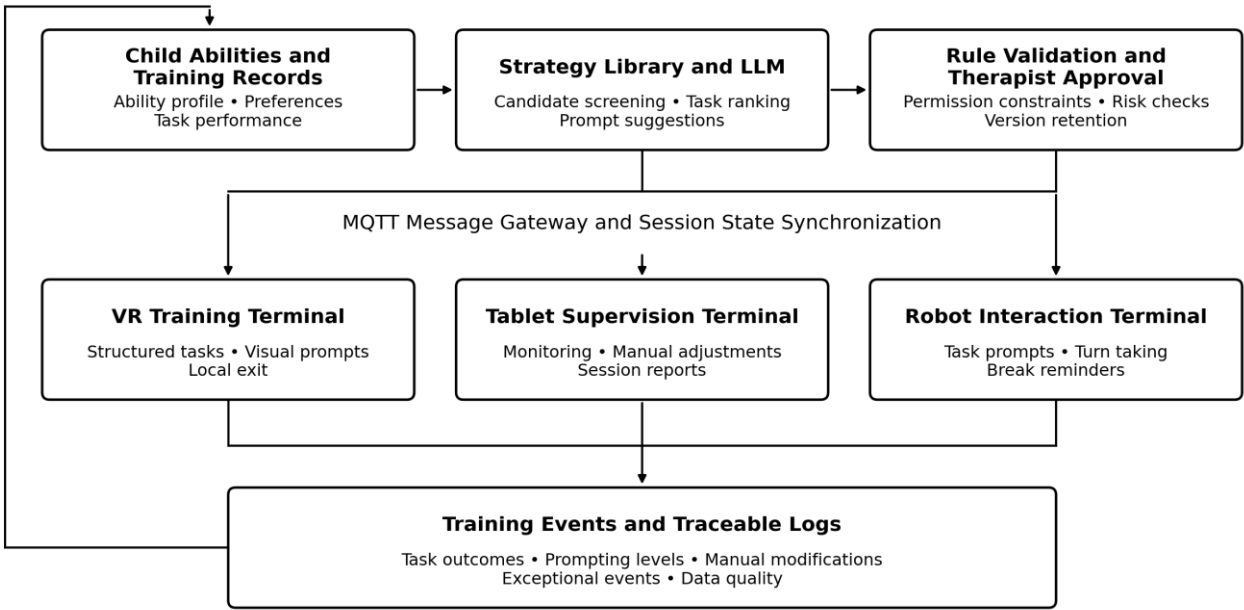


Figure 1 Overall System Architecture and Training Feedback Process

Data are linked through anonymized child identifiers, session identifiers, and task identifiers. Before training, the therapist confirms the ability profile and task list. During training, terminals report step outcomes, prompt use, and exceptional events. After training, the system summarizes performance and generates suggestions for the next round. Each task adjustment records the model recommendation, rule validation results, and the final human decision, allowing the recommendation process to be reviewed.

3. Key Module Design and Implementation Plan

3.1. Ability Profiles and Task Resource Model

Ability profiles include task related abilities, communication methods, support needs, interests and preferences, and historical performance. Perceptual dimensions may record color or shape matching performance; communication dimensions may record requests and instruction comprehension; and social dimensions may record turn taking, waiting, and collaborative task completion. Profile fields should identify their information sources and update times, distinguishing professional assessments, parent feedback, and system observations. When sources conflict, the therapist verifies the information before updating the profile.

Task resources are described using a unified data structure. Core fields include the task identifier, target ability, prerequisites, difficulty level, prompting method, environmental stimulus parameters, completion criteria, and exit conditions. Training performance cannot be assessed solely by completion speed: children who complete the same task may require different degrees of assistance. Independent completion and prompting level must therefore both be retained. Task version numbers distinguish content changes, preventing scores obtained at different difficulty levels from being pooled directly.

3.2. Personalized Strategy Recommendation Supported by a Large Language Model

Strategy recommendation follows a sequence of candidate screening, model recommendation, rule validation, and human confirmation. First, a candidate set is generated from a professionally reviewed strategy library according to prerequisites, mastered skills, and constraints. A de-identified ability summary, recent task performance, and candidate tasks are then submitted to the large language model, which is required to return task identifiers, suggested parameters, selection rationales, and identifiers for relevant records. The model neither directly generates unreviewed new training content nor directly controls terminal actions.

Model outputs use structured fields. DeepSeek provides JSON output capabilities [4], but syntactically valid JSON does not guarantee correct content. Application services must therefore verify whether the task belongs to the candidate set, whether parameters exceed permitted ranges, whether the rationale corresponds to the input information, and whether required fields are missing. If a timeout, empty output, or validation failure occurs, the system reverts to the most recently confirmed human approved plan and retains a failure log.

Within a task, stopping and exiting are handled by local deterministic logic. Between tasks, the large language model assists in selecting the next activity or prompting method. This timing avoids dependence on network inference for immediate responses. Therapists can accept, modify, or reject recommendations. The system separately stores the model version, prompt version, input summary, and final configuration. The recommendation acceptance rate reflects only the degree of agreement between recommendations and professional judgment; it cannot independently be interpreted as a training outcome.

3.3. State Recognition and Feedback from Multiple Sources

The state recognition module must distinguish observable behaviors from inferences about psychological states. Frowning, turning away, smiling, and pausing interaction may be used as behavioral labels, but no single behavior can be equated directly with anxiety, enjoyment, or refusal. The labeling scheme should specify operational definitions, time windows, annotation sources, and procedures for resolving disagreements. Segments that cannot be judged should be marked as uncertain rather than forced into predefined categories.

Head mounted devices obscure parts of the face. This design therefore limits analysis of full facial expressions to phases in which valid images are available, such as before or after training or during tablet interaction. Feedback during headset training comes primarily from task outcomes, prompt counts, pauses and exits, and therapist observations. Eye tracking or other sensor data are included only when the device actually provides those capabilities. Head orientation must not be described as eye tracking, and the absence of camera visibility must not be misrecorded as a child's failure to respond.

If the implementation uses a convolutional neural network, the number of children in the dataset, label distribution, acquisition conditions, model architecture, and training parameters should be documented in full. Training and test sets should be separated by child to prevent adjacent video frames from the same child appearing in both sets. Evaluation should report macro averaged F1, recall for each class, and a confusion matrix, together with the proportion of indeterminate cases. Model outputs serve only as supplementary cues; rules and professional observation jointly determine whether an adjustment should be recommended.

3.4. VR Environments and Tiered Task Design

Unity is used to organize interactive objects, event triggers, and step feedback. Environments are grouped into four functional modules: Explore at Your Own Pace, Fun Practice Zone, Everyday Scenarios Town, and Peer Play Community. These modules are functional categories; the numbers of specific environments and executable tasks must be counted separately. Environments use simplified models, clear targets, and configurable audiovisual stimuli, with complexity adjusted according to task requirements. Low saturation or low polygon styles are design choices whose comfort still requires evaluation through actual user feedback.

Table 2 Functional Modules and Training Task Design

Module	Example tasks	Main recorded indicators
Explore at Your Own Pace	Choosing activities and becoming familiar with the environment	Choices, time spent, and exits
Fun Practice Zone	Color matching and sequential construction	Accuracy, prompt counts, and proportion of independent steps
Everyday Scenarios Town	Shopping requests and street crossing steps	Spontaneous expression, step accuracy, and assistance level
Peer Play Community	Taking turns choosing and completing tasks together	Successful turns, waiting behavior, and completion status

Localized design draws on everyday situations familiar to children in Wenzhou, translating objects, activity sequences, and social conventions into repeatable training tasks. For example, a shopping task can be divided into checking a list, finding items, expressing needs, and ending an interaction. Whether regional adaptation facilitates task comprehension requires comparison with generic environments or observation in everyday life. Rehabilitation outcomes cannot be inferred directly from familiarity with an environment.

Prompts are documented using explicit levels. For each task, records retain independent completion, visual prompts, verbal prompts, or any other support actually used. The system provides preview, pause, and exit options and stops the current activity immediately when an exit is requested. Environments involving movement use assessed interaction methods, while training space requirements and supervision arrangements are included in operating procedures. Completion of VR tasks cannot replace assessment of competence in high risk activities, such as navigating real roads.

3.5. Coordination across Multiple Terminals and Traceable Records

Devices exchange status and task messages through the MQTT publish–subscribe mechanism [5]. Messages contain session identifiers, task versions, sequence numbers, timestamps, event types, and payloads. Real time status messages, task results, and control commands use delivery configurations appropriate to their purposes. Action commands additionally require acknowledgment, expiry times, and deduplication to prevent repeated delivery from causing repeated execution. The specific protocol version and quality of service levels should be reported from actual deployment records.

The robot is designed to provide prompts, turn taking interactions, and break reminders; it does not freely execute actions based on model generated text. During disconnection, the VR terminal retains its local exit function and suspends receipt of unconfirmed new tasks. The tablet displays device status. After reconnection, logs are retransmitted according to session identifiers and sequence numbers. Raw audio and video are managed separately from structured training records. External model calls transmit only the de-identified information needed for task recommendations, and the data processing conditions of the actual deployment must be verified.

4. Methods for the Preliminary Application Study

4.1. Study Design and Participants

The preliminary application is planned as a small scale feasibility study based on technical system testing and training process records, focusing on participation, completion, acceptability, and the feasibility of data collection. This paper provisionally adopts a single group pretest–posttest design; if the actual study uses another design, the description should be revised to reflect the real procedure. Pretest–posttest comparisons can describe change but cannot rule out natural development, repeated practice, or concurrent services, nor can they independently identify the contribution of the AI module.

Recruitment is planned through partner institutions in Wenzhou. Eligible children will have a diagnosis confirmed by a professional institution, be assessed as able to participate in the relevant training, and have guardian consent. The actual age range must reflect headset suitability requirements. The study will document communication methods, existing training arrangements, and relevant support needs, and specify procedures for children who cannot or do not wish to use a headset. Sample size should be justified by feasibility objectives, the available population, and estimation precision, rather than solely by an anticipated statistically significant difference.

4.2. Training Procedure and Measurement Schedule

Training comprises four stages: baseline assessment, device familiarization, formal tasks, and an end of session review. At baseline, professionals confirm target abilities and task levels. During familiarization, they observe acceptance of the device and interaction method. Formal tasks follow the confirmed plan, recording assistance, pauses, and exits. During review, automatically generated records are checked and reasons for manual changes are retained. Training frequency and duration are reported according to the actual implementation plan; sessions are not forcibly extended to meet a task completion quota.

Pretests and posttests use tasks of equivalent difficulty, with prompting methods and assessment environments controlled. If transfer to real life is assessed, observational activities should be established in everyday settings with the same goals as the virtual tasks but different materials. The assessor and whether the assessor knows the training phase should be specified. Observed real life performance and VR scores are reported separately. Concurrent routine

rehabilitation, missed sessions, and device failures are included in process records to help explain individual differences.

4.3. Technical Testing and Evaluation Indicators

Technical evaluation covers interactive operation, strategy recommendation, and communication coordination. Latency is defined as the time from generation of a relevant event to completion of the specified processing at the receiving end, with explicit clarification of whether network transmission and model generation are included. In addition to the mean, the median and 95th percentile are reported to reflect slower responses. Tests also specify terminal models, network conditions, the number of concurrent devices, and the number of repetitions. Recommendation tests use professionally reviewed cases and report the format pass rate, rule pass rate, and types of recommendations requiring manual modification.

Table 3 Preliminary Application Evaluation Indicators and Definitions

Evaluation level	Indicator	Definition or data source
Technical operation	Response latency and frame rate	Unified event timestamps and device performance logs
Recommendation quality	Rule pass rate and manual modification rate	Number passing or modified divided by the corresponding total recommendations
Feasibility of use	Training completion rate	Completed scheduled sessions divided by planned sessions
Feasibility of use	Exits and discomfort events	Child counts, event counts, and reasons reported separately
Task performance	Proportion of independently completed steps	Independently completed steps divided by applicable steps
Task performance	Prompt use	Standardized by task or training time and reported by level
Transfer to real life	Performance in everyday situations	Independent observation records and agreement between ratings
Implementation burden	Time spent organizing records	Actual time for an equivalent scope of records

If rating scales are used as exploratory outcomes, their names, versions, applicable age ranges, authorization requirements, assessors, and scoring directions should be specified, together with their relationship to target abilities. Parent satisfaction assesses user experience, while therapist opinions assess task suitability and workflow. Neither can replace objective task performance or clinical outcomes. For self developed questionnaires, item development and scoring procedures should be described, avoiding the characterization of an unvalidated total score as a standardized rehabilitation scale.

4.4. Data Analysis

Continuous variables are described using means and standard deviations or medians and interquartile ranges according to their distributions. Categorical variables are reported as counts and proportions, with confidence intervals for key feasibility proportions. Pretest–posttest analyses use the child as the unit of analysis and prioritize individual changes and within group differences. Paired tests are used when their assumptions are met, with effect sizes and confidence intervals reported. Inference from small samples remains exploratory, and statistical significance is not equated with practical benefit.

Multiple sessions and video frames from the same child are not treated as independent participants. Reasons for missing data are recorded, distinguishing nonattendance, voluntary withdrawal, missing sensor data, and technical failures. The actual denominator for each analysis is reported. Comparisons across tasks must account for differences in difficulty and prompting. Feasibility studies emphasize reporting recruitment, acceptability, and data completeness, a principle informed by the pilot study methods literature [6]. That reference concerns randomized pilot trials; if this study uses a single group design, it does not claim compliance with all of those standards.

5. System Implementation and Preliminary Application Results

5.1. Implementation Scope and Technical Performance

System performance results must correspond to the test conditions in Section 4. If state recognition has not been tested on an independent sample of children, only functional integration testing of the module is reported, without a recognition accuracy value. Device coordination should be reported separately under normal and abnormal network conditions, avoiding the equation of message delivery with correct execution of a terminal action.

5.2. Participation and Feasibility of Use

Acceptability results should cover all children who start using the system, rather than only those who complete training. If no discomfort is reported during the observation period, the observation duration, monitoring method, and sample size must also be stated. This finding cannot establish that the system is risk free for all children. Therapist and parent feedback should combine quantitative counts with thematic summaries and retain negative opinions, including dissatisfaction and operating difficulties.

5.3. Illustrative Training Processes

Illustrative cases are presented in the sequence of baseline performance, initial task, records triggering adjustment, human decision, and subsequent performance. Cases use anonymized identifiers, and reasons for selection are stated. At least one case showing no clear improvement or requiring frequent adjustments should be retained to reflect the limits of system adaptation. Cases illustrate processes and cannot replace aggregate statistical results.

6. Discussion

6.1. Technical Significance of the Personalization Mechanism

This paper operationalizes personalization through task selection, support levels, and feedback pacing, linking each adjustment to the child's specific performance. The large language model organizes semantic information and recommends candidate plans; the structured strategy library restricts the executable range; and therapists make contextual judgments. This division of responsibilities allows model recommendation quality, rule processing capability, and final use to be evaluated separately, rather than conflating the effects of different modules.

The system's value should be tested against existing working practices. To identify the incremental contribution of the large language model in future studies, rule based recommendations could be compared with model assisted recommendations under identical environments and training durations, controlling for additional human attention. Pretest–posttest score comparisons alone cannot determine whether changes arise from AI, VR, professional support, or repeated practice. The current technical contribution is a traceable pathway for task adjustment; actual benefits still require experimental evidence.

6.2. Local Environments and Transfer to Real Life

Localized resources may align task content more closely with everyday experiences, but similarity between environments does not establish skill transfer. Effective transfer assessment requires children to use the relevant skills with different people, objects, or settings. The system can link the same objective to multiple environments to reduce reliance on cues from a single scene. Whether changes occur in everyday life requires independent observation and follow-up.

The rationale for coordination across multiple terminals likewise depends on task needs. Whether a robot creates additional opportunities for interaction and whether a tablet reduces documentation burden must be evaluated separately. If multiple terminals increase operating complexity or distract attention, a simpler device configuration should be permitted. The degree of technical integration should not itself be taken as evidence of system superiority.

6.3. Study Limitations and Future Directions

Evidence from actual application still needs to be supplied through real operational records and participant data. The planned small scale single group study has limitations including restricted sample representativeness, no concurrent control group, and a limited observation period. State recognition is affected by occlusion, individual differences, and label reliability. Findings based on facial data cannot be generalized unconditionally to headset training. Changes in large language model versions may also affect recommendation consistency; the test version should therefore be fixed and its configuration recorded.

Following feasibility evaluation, future work could conduct prospectively designed controlled studies, include samples from multiple institutions, and add follow-up in real life settings. Strategy recommendation should be compared against a rule based baseline, and localized environments should be compared under task equivalent conditions. System optimization should also address children who cannot tolerate a headset or gain limited benefit from training, providing alternative interaction pathways and clear limits of applicability.

7. Conclusion

Addressing personalized training needs among children with autism, this paper proposes a rehabilitation support system that integrates structured tasks, large language model assisted recommendations, VR situations, and coordination across multiple terminals. It establishes a feedback process from ability profiling to task planning, process recording, and human confirmation, and provides methods for evaluating technical performance, feasibility of use, and exploratory task performance. The design emphasizes executable tasks, traceable adjustments, and professional intervention. At present, the system design and evaluation framework can be confirmed. Actual operational feasibility, changes in children's training performance, and transfer to real life must each be reported on the basis of relevant data; rehabilitation effectiveness cannot be inferred directly from the technical proposal.

Funding

Wenzhou Science and Technology Project, "Development and Application of a Personalized Rehabilitation System for Children with Autism in Wenzhou Using AI Driven Immersive VR Environments" (Project No. RC20250348).

References

- [1] Didehbani, N., Allen, T., Kandalaft, M., et al. (2016). Virtual reality social cognition training for children with high functioning autism. *Computers in Human Behavior*, 62, 703–711. <https://doi.org/10.1016/j.chb.2016.04.033>
- [2] Maddalon, L., Minissi, M. E., Parsons, T., et al. (2024). Exploring adaptive virtual reality systems used in interventions for children with autism spectrum disorder: Systematic review. *Journal of Medical Internet Research*, 26, e57093. <https://doi.org/10.2196/57093>
- [3] Virues-Ortega, J., Julio, F. M., & Pastor-Barriuso, R. (2013). The TEACCH program for children and adults with autism: A meta-analysis of intervention studies. *Clinical Psychology Review*, 33(8), 940–953. <https://doi.org/10.1016/j.cpr.2013.07.005>
- [4] Deepseek. (n.d.). JSON output. https://api-docs.deepseek.com/guides/json_mode/
- [5] OASIS. (2019). MQTT version 5.0. <https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html>
- [6] Eldridge, S. M., Chan, C. L., Campbell, M. J., et al. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *BMJ*, 355, i5239. <https://doi.org/10.1136/bmj.i5239>