Computing, Performance and Communication Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/cpcs.2026.100106
ISSN 2371-8889 Vol. 10 Num. 1

Design and Implementation of an Express Warehousing Cloud Platform Based on Multimodal AI

Ji Botong, Liang Can, Wu Ruihao, Zhu Yining, Peng Yihua

School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, Liaoning, China

Keywords: Intelligent Warehouse; AI Large Models; AGV; Spring Cloud; Internet of Things

Abstract: To address issues such as low efficiency and poor real-time performance in traditional warehouse management models, this paper designs a cloud-based warehouse platform solution utilizing AI large models and Internet of Things technology. The system achieves comprehensive intelligent management of warehouse inbound and outbound processes through multi-modal AI shared decision-making (DeepSeek-R1 language model and GLM-4V vision model), AGV high-precision navigation, and 3D visualization technology, forming a closed-loop management system of "perception-decision-action-monitoring." The solution employs modular structural design, supports easy scalability, and features storage and computing design combining edge and cloud modes. It integrates dual-factor authentication and distributed transaction design to ensure the reliability and security of the warehouse management system. This solution enables multi-modal data extraction, dynamic path planning, and real-time data display, providing a relatively complete and extensible technical approach to intelligent warehouse logistics. Future work on this solution may explore the integration of 5G technology and digital twin technology for further applications.

1. Introduction

With the continuous expansion of e-commerce scale and increasing consumer demands for logistics timeliness, traditional warehouse management models centered on manual labor are facing severe challenges. These models generally suffer from inherent shortcomings such as operational efficiency bottlenecks, high error rates in manual inventory counting and sorting, untimely inventory status updates, and slow on-site anomaly response. They cannot meet the requirements of modern supply chains for precision, real-time operation, and reliability. To address the above pain points and integrate the intelligent analytical capabilities of Artificial Intelligence (AI) with the comprehensive sensing capabilities of the Internet of Things (IoT), creating a warehouse system capable of full-process automation and intelligent decision-making has become a key means of achieving industry upgrading [1].

Against this background, this paper designs and implements an intelligent express warehouse management platform. The system leverages IoT technology to accurately track the entire lifecycle of goods from inbound to outbound, and introduces AI large models (DeepSeek-R1 and GLM-4V) to enhance data analysis and decision-making capabilities in complex scenarios. The system integrates Automated Guided Vehicles (AGVs) to optimize internal paths, aiming to create an efficient, secure, and scalable intelligent warehouse management framework [2]. The modular architecture proposed in this study effectively integrates core functions such as automated handling, real-time 3D monitoring, and intelligent early warning, intending to provide a reusable technical solution for the industry.

2. System Overview

2.1 Overall Architecture Composition

The overall architecture of the intelligent express warehouse cloud platform is built using IoT, big data, and AI technologies, adopting a layered design pattern of "perception-network-analysis-application." The perception layer achieves full-domain data acquisition through multi-modal devices: RFID electronic tags (recognition accuracy ≥99.9%) support second-level identification and tracking of goods [3]; temperature, humidity, weight, and vibration sensors monitor environmental status in real time; AI cameras combined with the GLM-4V model detect abnormal cargo stacking, realizing digital mapping of physical scenes. The network layer employs a hybrid communication scheme to ensure data transmission efficiency and reliability, including dual-channel Wi-Fi 6 and 5G private network for transmitting real-time AGV status data, and long-distance fiber optic backbone network (10Gbps bandwidth) for multi-warehouse data synchronization, ensuring overall consistency using timestamps and collision detection algorithms.

The data layer is based on a Hadoop+Spark distributed architecture to achieve efficient processing of massive data: the Kafka stream processing engine achieves second-level inventory status updates (throughput ≥50,000 records/second). Cold data is archived to MinIO object storage with fast retrieval capabilities. Analytical models use TensorFlow for demand forecasting (MAE=3.2%) and the Apriori algorithm to mine product association rules. The application layer provides core service functions through modular design: dynamic slot allocation optimizes warehouse space utilization; intelligent order splitting and path planning improve order processing efficiency; AGV scheduling uses multi-objective optimization algorithms to balance task priorities; the interactive interface achieves 3D digital twin panoramic display based on the Three.js engine, combined with ECharts to create inventory heatmaps and task efficiency dashboards, and supports seamless integration with third-party systems such as ERP and WMS via RESTful API interfaces (Figure 1).

Figure 1 System Layered Architecture Interaction Diagram (Perception Layer → Network Layer → Data Layer → Application Layer)

2.2 Warehouse Main System Overview

The warehouse main system is the core of intelligent warehouse management, integrating functions such as user authentication, full lifecycle cargo management, intelligent scheduling, and decision support. It uses modular design to achieve efficient collaboration. The system is built on a microservices architecture (Spring Cloud Alibaba), containing seven core modules: login authentication, inbound/outbound management, logistics audit process, message center, inventory counting process, AGV scheduling, and AI intelligent warehousing. The login module uses dual-factor authentication (JWT token + dynamic verification code) and distributed session management to ensure the security of multi-terminal access [4]. The inbound/outbound management module utilizes RFID technology to achieve fast cargo identification and automatic data collection, combined with an AI large model (DeepSeek-R1) to dynamically optimize inventory distribution strategies, significantly reducing the need for manual intervention.

The AI intelligent warehousing module is an innovative highlight of the system. This intelligent assistant interface (as shown in Figure 2) supports natural language interaction and data visualization analysis. Users can input commands like "query today's outbound records in Zone A" in the chat window. The system uses the DeepSeek-R1 model to retrieve the database in real time and provide structured feedback, while simultaneously presenting inventory trends and turnover rates using line charts, heatmaps, etc. This interface integrates anomaly warning functions. For example, when the GLM-4V vision model detects cargo slippage or AGV path deviation, the interface automatically pops up a warning window and recommends solutions, forming a circular management of "perception-analysis-response."

The system uses distributed technology to ensure stability in high-concurrency scenarios. For example, it uses RocketMQ transactional messages to ensure the atomicity of AGV task assignment and inventory updates, and uses a Redis cluster to cache frequently accessed data. The connection design between the 3D digital twin interface and the multi-terminal visualization dashboard allows managers to grasp the overall warehouse dynamics in real time, quickly identify bottlenecks, and provide comprehensive support for the intelligent upgrade of warehouse operations.

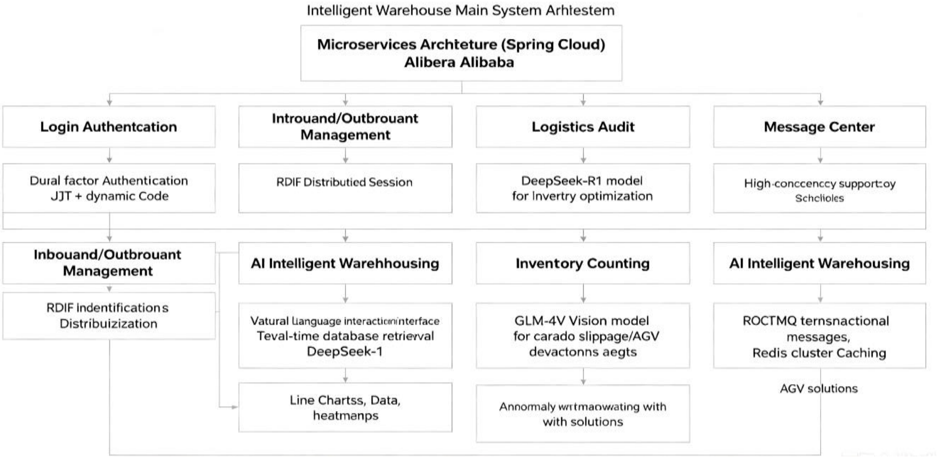
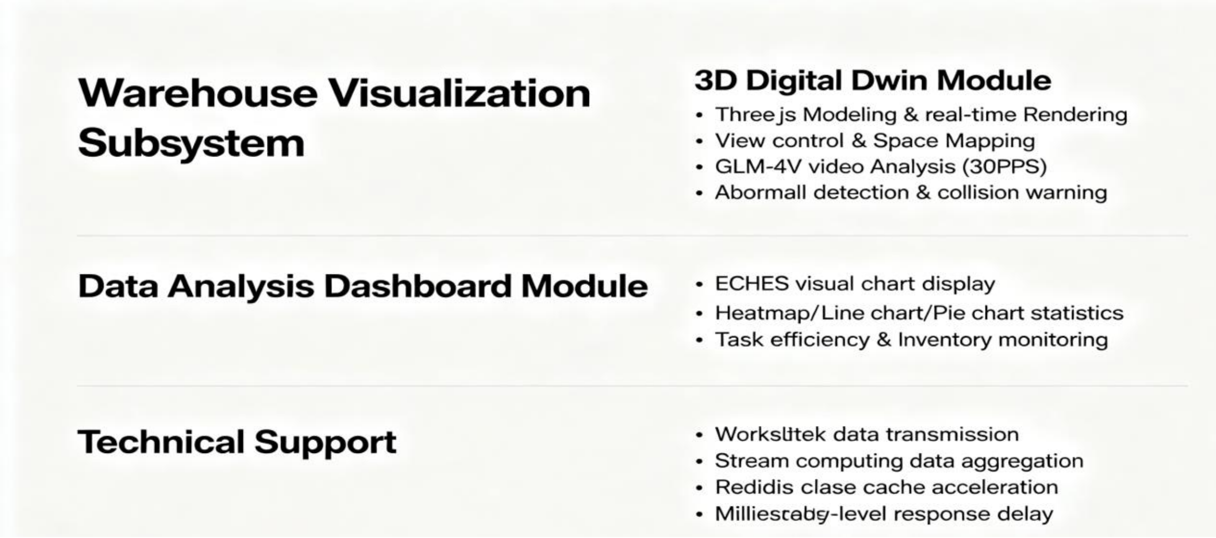


Figure 2 AI Intelligent Assistant Interface

2.3 Warehouse Visualization Subsystem Overview

The warehouse visualization subsystem, through the collaborative design of 3D digital twin and data analysis dashboards, provides intuitive, real-time global monitoring and decision support for warehouse management. The 3D digital twin module, built based on Three.js technology, renders the warehouse 3D model and AGV trajectories in real time. It supports functions such as view zooming, rotation, and shelf status perspective, achieving precise mapping between physical and virtual spaces [5]. The system integrates multi-sensor data streams with the GLM-4V vision understanding model. It uses the WebSocket protocol to analyze 30 FPS camera feeds in real time, detect abnormal cargo postures, and mark them on the 3D interface. Combined with collision detection algorithms, it updates abnormal areas (e.g., cargo, equipment failure) synchronously, providing managers with immersive monitoring and intelligent early warning capabilities.

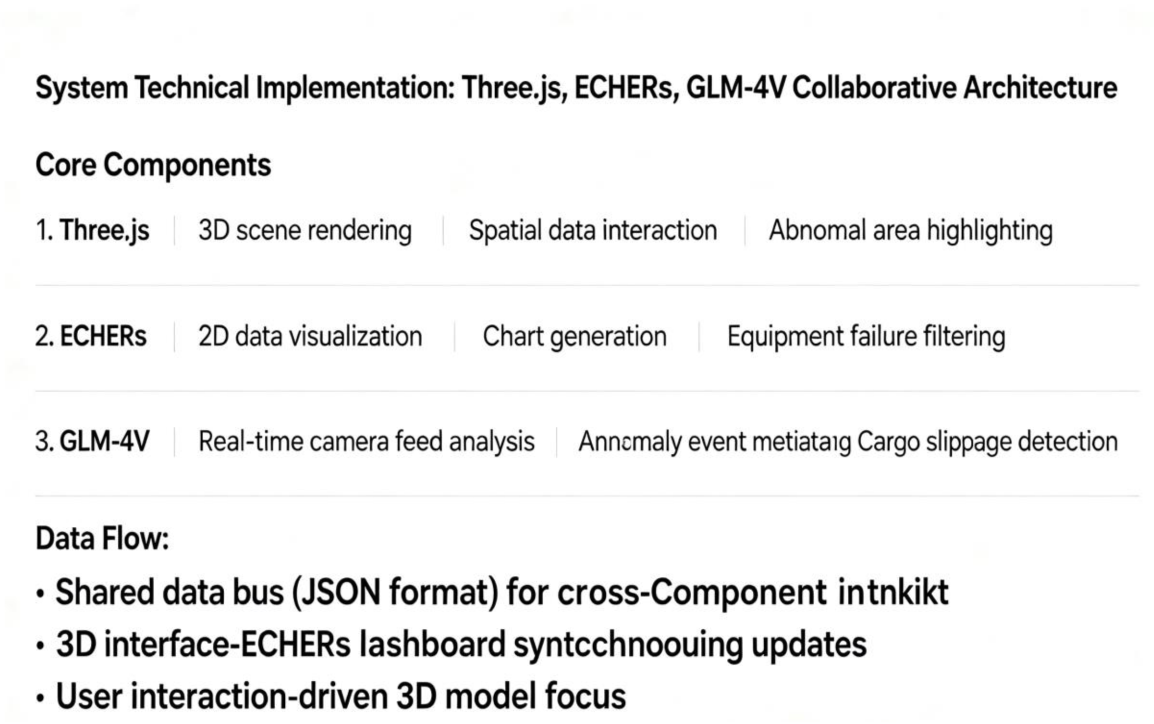
The data analysis dashboard module, developed using the ECharts framework, builds a 2D visualization interface for large screens. It dynamically displays core warehouse indicators using multi-dimensional charts such as heatmaps, line charts, and pie charts. The inventory trend heatmap uses color gradients to identify high/low turnover areas based on shelf zoning and cargo turnover frequency. The task completion rate statistics show AGV task execution efficiency over time and integrate the results of GLM-4V anomaly detection, supporting filtering by different anomaly types. The system aggregates data reported by edge nodes using a stream computing engine, combined with caching strategies (Redis cluster), to reduce the response latency of the visualization interface to the millisecond level, ensuring the real-time and interactive nature of data analysis (Figure 3).



Warehouse Visualization Subsystem	**3D Digital Dwin Module** • Threejs Modeling & real-time Rendering • View control & Space Mapping • GLM-4V video Analysis (30PPS) • Abormall detection & collision warning
Data Analysis Dashboard Module	• ECHES visual chart display • Heatmap/Line chart/Pie chart statistics • Task efficiency & Inventory monitoring
Technical Support	• Worksütek data transmission • Stream computing data aggregation • Rededis clase cache acceleration • Milliestabg-level response delay

Figure 3 Visualization Interface Rendering

In terms of technical implementation, the system uses a collaborative architecture of Three.js, ECharts, and GLM-4V: Three.js is responsible for 3D scene rendering and spatial data interaction, ECharts handles 2D data visualization and chart generation, and the GLM-4V model analyzes camera feeds in real time to generate anomaly event metadata. The three are linked using a shared data bus (JSON format). When GLM-4V detects cargo slippage on a certain shelf, the 3D interface automatically highlights the corresponding area, and the dashboard synchronously pops up a solution. When users filter by "equipment failure" category on the dashboard, the 3D model focuses on the abnormal AGV trajectory. This design balances functional completeness and user experience of the visualization system, providing technical support for efficient warehouse management decision-making (Figure 4).



System Technical Implementation: Three.js, ECHERs, GLM-4V Collaborative Architecture

Core Components

1. **Three.js**	3D scene rendering	Spatial data interaction	Abnomal area highlighting
2. **ECHERs**	2D data visualization	Chart generation	Equipment failure filtering
3. **GLM-4V**	Real-time camera feed analysis	Anncmaly event metiataıg Cargo slippage detection	

Data Flow:

- **Shared data bus (JSON format) for cross-Component intnkikt**
- **3D interface-ECHERs lashboard syntcchnoouing updates**
- **User interaction-driven 3D model focus**

Figure 4 Three.js and ECharts Technology Implementation Flow

3. Hardware System Design

3.1 IoT Technology Selection

The AGV uses a fusion scheme of magnetic navigation and visual SLAM. Its technology selection is based on a comprehensive evaluation of cost, accuracy, and adaptability (Table 1):

Table 1 Comparison of AGV Navigation Technologies

Navigation Method	Cost (CNY 10k/unit)	Positioning Accuracy	Environmental Adaptability
Magnetic Navigation	8-12	±1 cm	Fixed Path
Laser Navigation	15-25	±0.5 cm	Dynamic Environment
Visual SLAM	10-18	±2 cm	Dynamic Environment

The fusion scheme of magnetic navigation and visual SLAM improves AGV navigation efficiency through complementary advantages: Magnetic navigation, based on embedded magnetic tracks and Hall sensors (±0.1 mT), achieves trajectory tracking accuracy of ±1 cm. A PID controller dynamically adjusts motor steering, reducing the cost per unit by 40%, suitable for large-scale deployment. Visual SLAM, using the ORB-SLAM3 algorithm, processes 30 frames per second to extract environmental feature points. Combined with IMU data, it achieves real-time localization and mapping. It not only compensates for the path rigidity of magnetic navigation with ±2 cm accuracy but also compensates for the poor flexibility of magnetic navigation through dynamic obstacle avoidance capabilities (e.g., handling temporary shelf adjustments), ultimately achieving an optimized comprehensive positioning accuracy of ±1 cm, greatly enhancing the system's environmental adaptability and cost-effectiveness in complex scenarios.

3.2 AGV Hardware Implementation

The AGV hardware system is based on a modular design concept, utilizing high-precision sensors and efficient drive components to achieve comprehensive optimization of navigation, positioning, and energy management. The drive module uses the TB6612 motor driver chip combined with Mecanum wheels. TB6612 uses PWM signals to precisely control the speed of four DC motors, featuring bidirectional current output and overcurrent protection, ensuring stable operation under a 200kg load. Mecanum wheels, with their omnidirectional movement capability, reduce the minimum turning radius to 0.5m, greatly improving maneuverability in narrow shelf areas. The positioning module integrates multi-source perception technologies: the magnetic navigation system, relying on embedded magnetic tracks and Hall sensors (sensitivity ±0.1 mT), achieves ±1 cm level path tracking. The visual SLAM module is equipped with an OpenMV camera and a 16-line LiDAR, using the ORB-SLAM3 algorithm to instantly extract environmental feature points. It uses AprilTag recognition (TAG36H11 family) for auxiliary positioning, dynamically compensating for magnetic navigation path deviations, achieving a comprehensive positioning accuracy of ±1.5 cm. The energy module uses a high-density lithium battery pack (48V/20Ah) to supply power, featuring intelligent power monitoring circuits and fast charging protocols, combined with dynamic power management, achieving a single charge endurance of 12 hours, meeting full-day operational requirements.

4. Software System Design

4.1 Multi-modal AI Collaborative Decision-Making

In the smart express warehouse management system, multi-modal AI collaborative decision-making technology integrates text, vision, and sensor data to build domain-wide perception and intelligent response capabilities, driving the optimization of the entire warehouse management process. Based on the collaborative architecture of the DeepSeek-R1 large language model and the GLM-4V vision model, the system achieves cross-modal information complementarity.

DeepSeek-R1 is based on the Transformer architecture. During pre-training, it uses a logistics corpus containing 100,000 work order texts. Knowledge distillation technology (hierarchical feature matching and adaptive temperature adjustment) reduces the model size by 30%.GLM-4V adopts a cascaded architecture of YOLOv8 object detection and Mask R-CNN instance segmentation, using a Dynamic Weight Assignment (DWA) mechanism to balance detection and segmentation tasks. This multi-modal feature fusion method references cutting-edge results in deep learning-based RGB-D image salient object detection [6]. The formula is as follows:

Here, IoU_det represents the Intersection over Union for object detection, and mAP_seg is the mean Average Precision for instance segmentation. This formula, inspired by the multi-modal feature fusion method of "FusionRCNN" from CVPR 2023, dynamically adjusts the weights of detection and segmentation tasks to optimize model performance. During training, text data undergoes synonym replacement and sentence restructuring to enhance robustness, while visual data uses random cropping and illumination perturbations to optimize feature extraction capabilities [7,8].

4.2 Microservices Architecture

The system is built on a distributed microservices architecture, achieving functional decoupling and elastic scaling using the Spring Cloud Alibaba framework. In terms of service decomposition, core business modules are separated into three major services:

Inventory Service: Stores cargo information using MySQL sharding (horizontal partitioning by shelf zone), ensuring high-frequency read/write operations (TPS ≥ 5000), and combines pessimistic locking mechanisms to prevent overselling. AGV Scheduling Service: Uses RocketMQ transactional message queues to achieve asynchronous task sending and status synchronization, ensuring reliable transmission of AGV commands. It also dynamically adjusts task order using priority queues based on path congestion index and AGV battery level.AI Analysis Service: Uses a Redis cluster to cache hot data, employing an LRU elimination strategy to maintain a cache hit rate of no less than 85%. It integrates lightweight AI models (DeepSeek-R1, GLM-4V) to provide edge inference support (Figure 5).

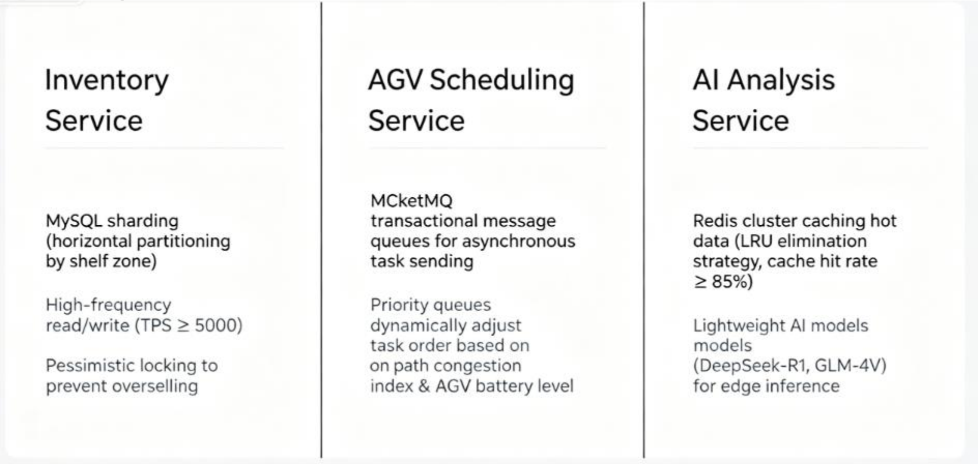


Figure 5 Microservices Architecture Interaction Flowchart

5. Conclusion

This paper proposes a design method for an intelligent express warehouse cloud platform based on multi-modal AI. By integrating multi-modal AI collaborative decision-making (DeepSeek-R1 and GLM-4V), an edge-cloud layered architecture, and modular hardware design, the system achieves intelligence throughout the entire logistics process of warehouse management. The system integrates high-precision AGV navigation (magnetic navigation + visual SLAM) with comprehensive positioning accuracy of ±1 cm. The main system decision-making is optimized using a microservices architecture (throughput ≥50,000 TPS) along with the dynamic monitoring functions of the warehouse visualization subsystem. It combines dual-factor authentication (JWT + dynamic verification code) and distributed transactions (Seata) to ensure data security and consistency. Future research could further reference security enhancement methods for the integrated convergence of sensing, computing, control, and intelligence in industrial internet to improve the system's protection level. Current research needs to further improve generalization capability in complex scenarios. Future work will integrate 5G and digital twin technology, deepen the edge-cloud collaboration mechanism, and support the dynamic growth of ultra-large-scale logistics networks.

Acknowledgements

This research was supported by the 2026 Innovation Project of the University of Science and Technology Liaoning.

References

[1] Ai Lebo Robot. The Future of WMS Warehouse Management Systems: Trends in AI, Big Data, and IoT [R/OL]. 2023-11-02. (Chinese)

[2] Xiao Guangwei, Chen Hao, Shao Shizhou, et al. Research on intelligent warehouse management model based on IoT technology [J]. Logistics Technology and Application, 2021, 45(4): 58-63. (Chinese)

[3] Gao Ning, Yang Yongfeng. Security protection strategy for IoT warehouse systems [J]. Journal of Information Security, 2024, 11(3): 45-52. (Chinese)

[4] Zhang Zezhi, Lei Jie. Design of warehouse visualization system based on Three.js and ECharts [J]. Application Research of Computers, 2023, 40(5): 123-127. (Chinese)

[5] Liu Pang. Design and optimization of AGV system based on magnetic navigation and visual SLAM [J]. Automation Technology and Application, 2024, 43(2): 12-18. (Chinese)

[6] Li Yanan. Research on path planning algorithm for intelligent warehouse system [D]. Beijing: Tsinghua University, 2023. (Chinese)

[7] Huang Nianchang, Yang Yang, Zhang Qiang, et al. Frontiers in Deep Learning-Based RGB-D Image Salient Object Detection [J]. Chinese Journal of Computers, 2025, 48(2): 284-316. DOI:10.11897/SP.J.1016.2025.00284. (Chinese)

[8] Wang Yazhe, Ren Lei, Feng Dengguo, et al. Security Enhancement for Novel Industrial Internet PLCs Towards the Integration of Sensing, Computing, Control, and Intelligence: Trends and Prospects [J]. Chinese Journal of Computers, 2025, 48(3): 738-762. DOI:10.11897/SP.J.1016.2025.00738. (Chinese)