

## PAPER

# Architecture and Implementation of a Mobile Internet of Things-Enabled Logistics Management System

Yong Chen()Shangqiu Polytechnic,  
Shangqiu, China[13507699675@163.com](mailto:13507699675@163.com)**ABSTRACT**

The deep integration of the Mobile Internet of Things (MIoT) and edge intelligence enables intelligent logistics to enter a new phase of development. However, traditional end-cloud architectures are constrained in highly dynamic logistics scenarios, impeding large-scale industrial upgrading. Therefore, a three-level innovation framework encompassing paradigm, methodology, and enabling technologies was established, with proactive collaborative edge intelligence agents as the core paradigm to overcome the limitations. A hierarchical architecture consisting of an intelligent terminal layer, a collaborative edge layer, and a cognitive cloud brain layer was constructed, with three key methodological components integrated: dynamic federated task offloading, large language model (LLM)-driven decision generation, and privacy-enhanced cross-domain knowledge transfer. Correspondingly, enabling technologies—including a mobility-aware scheduling algorithm and an LLM-agent tool invocation mechanism—were designed. The study advances the deep integration of edge intelligence and MIoT technologies in vertical industries and exhibits substantial academic significance and industrial application potential.

**KEYWORDS**

Mobile Internet of Things (MIoT), proactive collaborative edge intelligence agents, dynamic federated learning, large language model (LLM), LLM-driven decision-making, logistics management systems

## 1 INTRODUCTION

In recent years, the integration of Mobile Internet of Things (MIoT) [1, 2] with edge intelligence [3] has emerged as a global research frontier in the field of intelligent logistics. A growing body of studies has demonstrated that the deep deployment of fifth-generation vehicle-to-everything (5G-V2X) and edge computing technologies can substantially enhance logistics resource scheduling efficiency and service quality, thereby accelerating the transition of the logistics industry toward intelligent and high-efficiency paradigms [4, 5]. Nevertheless, traditional end-cloud

Chen, Y. (2026). Architecture and Implementation of a Mobile Internet of Things-Enabled Logistics Management System. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(5), pp. 87–101. <https://doi.org/10.3991/ijim.v20i05.60725>

Article submitted 2025-11-03. Revision uploaded 2025-12-20. Final acceptance 2026-01-13.

© 2026 by the authors of this article. Published under CC-BY.

architectures continue to encounter three fundamental bottlenecks in highly dynamic logistics environments. First, continuous changes in network topology induced by terminal mobility render static architectures incapable of maintaining stable connectivity [6]. Second, centralized cloud-based decision-making mechanisms incur significant latency, which is incompatible with the real-time collaborative requirements of vehicle platooning, emergency rerouting, and so on [7, 8]. Third, pronounced risks of data privacy leakage arise during cross-domain logistics data sharing, severely constraining the improvement of collaborative efficiency [9, 10]. At present, no internationally recognized dynamic collaborative architectural paradigm has been established that is fully compatible with highly dynamic scenarios. Substantial research gaps persist in areas such as joint optimization mechanisms across mobility, computational resources, and energy consumption [11, 12], as well as the engineering-level integration of large language models (LLMs) [13, 14] into industrial environments [15, 16]. Addressing these technological bottlenecks is therefore associated with both significant academic value and pressing practical urgency.

In response to these challenges, three core scientific questions should be identified. First, how can a proactive collaborative architectural paradigm be constructed to accommodate highly dynamic logistics scenarios, enabling adaptive networking among terminals and edge nodes as well as efficient task offloading? Second, how can an LLM-driven semantic decision-making mechanism be designed to achieve accurate and efficient transformation from natural-language business instructions to executable workflows while ensuring decision robustness? Third, how can efficient cross-domain knowledge transfer be realized under stringent data privacy constraints while achieving an optimal balance between privacy protection and system performance?

The central objective of this study is to establish a proactive collaborative edge intelligence agent architecture, together with its supporting technologies, to enable low-latency, high-reliability, and privacy-enhanced intelligent management for MIIoT-driven logistics systems. On this basis, a prototype system is developed and validated through multidimensional evaluation. In alignment with this objective, the primary academic contributions of the study are summarized in four aspects. First, a proactive collaborative edge intelligence paradigm is introduced, and a three-level innovation framework encompassing paradigm, methodology, and enabling technologies is constructed, systematically addressing the core challenges of mobility, real-time performance, and privacy in logistics scenarios. Second, a mobility-aware dynamic federated task offloading and scheduling algorithm is designed. By formulating a multi-objective optimization model, a coordinated balance among latency, energy consumption, and computational resources is achieved, thereby enabling precise adaptation to the demands of highly dynamic logistics environments. Third, an LLM-agent-driven digital twin decision-making framework is proposed. Targeted prompt engineering and tool invocation mechanisms are developed to effectively address the challenges of accurate semantic parsing and robust execution of natural-language instructions in industrial scenarios. Finally, a full-stack prototype system is developed, and a comprehensive experimental validation framework is established, encompassing simulation-based experiments, real-world deployment scenarios, and extreme failure conditions. A reproducible engineering implementation paradigm is thereby provided, with experimental results demonstrating that the proposed system consistently outperforms state-of-the-art solutions.

## 2 SYSTEM ARCHITECTURE DESIGN

The architecture of the proposed system is developed in accordance with five fundamental principles to ensure effective adaptation to highly dynamic logistics scenarios and to satisfy practical deployment requirements. The mobility-first adaptation principle is adopted to ensure network connectivity stability and continuity of task collaboration during high-speed terminal movement, thereby establishing a robust foundation for system operation in dynamic environments. The principle of balanced real-time performance and reliability is realized through proximity-based computation at the edge side combined with cross-layer collaborative resource scheduling. In this manner, decision latency is reduced while multipath redundancy mechanisms are established to guarantee the reliability of data transmission and task execution. The privacy enhancement principle is enforced across all architectural layers. Through distributed data processing and secure communication mechanisms, the risks associated with centralized storage of sensitive logistics data are effectively mitigated. The semantic-driven decision-making principle is enabled by the semantic understanding capabilities of LLM agents, allowing accurate mapping between business requirements and system-level execution, thereby enhancing the intelligent adaptability of the overall architecture. The principle of scalability and compatibility is incorporated to ensure that the architecture remains compatible with heterogeneous logistics terminal devices and communication protocols. Dynamic expansion of terminal scale and service capacity is thereby supported, enabling adaptation to the diverse and evolving requirements of the logistics industry.

Figure 1 illustrates the architecture of the MIIoT-driven logistics management system. A three-level hierarchical architecture—comprising the intelligent terminal layer, collaborative edge layer, and cognitive cloud brain layer—is adopted. Through coordinated interactions across these layers, a closed-loop intelligent management process encompassing perception, decision-making, and execution is established. Both the physical deployment and the interlayer interaction logic are explicitly designed to accommodate the highly dynamic characteristics of logistics scenarios. The intelligent terminal layer functions as the active perception and execution unit. Multi-source fused positioning modules and 5G-V2X/reduced capability (RedCap) communication modules are integrated, while lightweight artificial intelligence models are deployed to enable real-time local perception and data preprocessing. An active collaborative protocol stack, extended from Message Queuing Telemetry Transport for Sensor Networks (MQTT-SN), is designed to allow terminals to proactively initiate networking negotiation and task collaboration, thereby enhancing the responsiveness and efficiency of distributed coordination. The collaborative edge layer adopts a hybrid coordination model that combines fixed edge servers with dynamic mobile edge clusters. The dynamic mobile edge clusters are autonomously formed by geographically adjacent terminals. A dynamic federated scheduler is responsible for core orchestration functions, including cluster formation, task decomposition, efficient offloading, and task migration during mobility, thereby enabling dynamic and balanced allocation of computational resources. The cognitive cloud brain layer employs multi-granularity digital twin modeling at the device, link, and network levels to accurately reflect the operational states of the physical logistics network. An LLM-agent-driven decision engine, consisting of instruction parsing, simulation-based evaluation, and workflow generation modules, enables fully automated processing from business instructions to executable strategies.

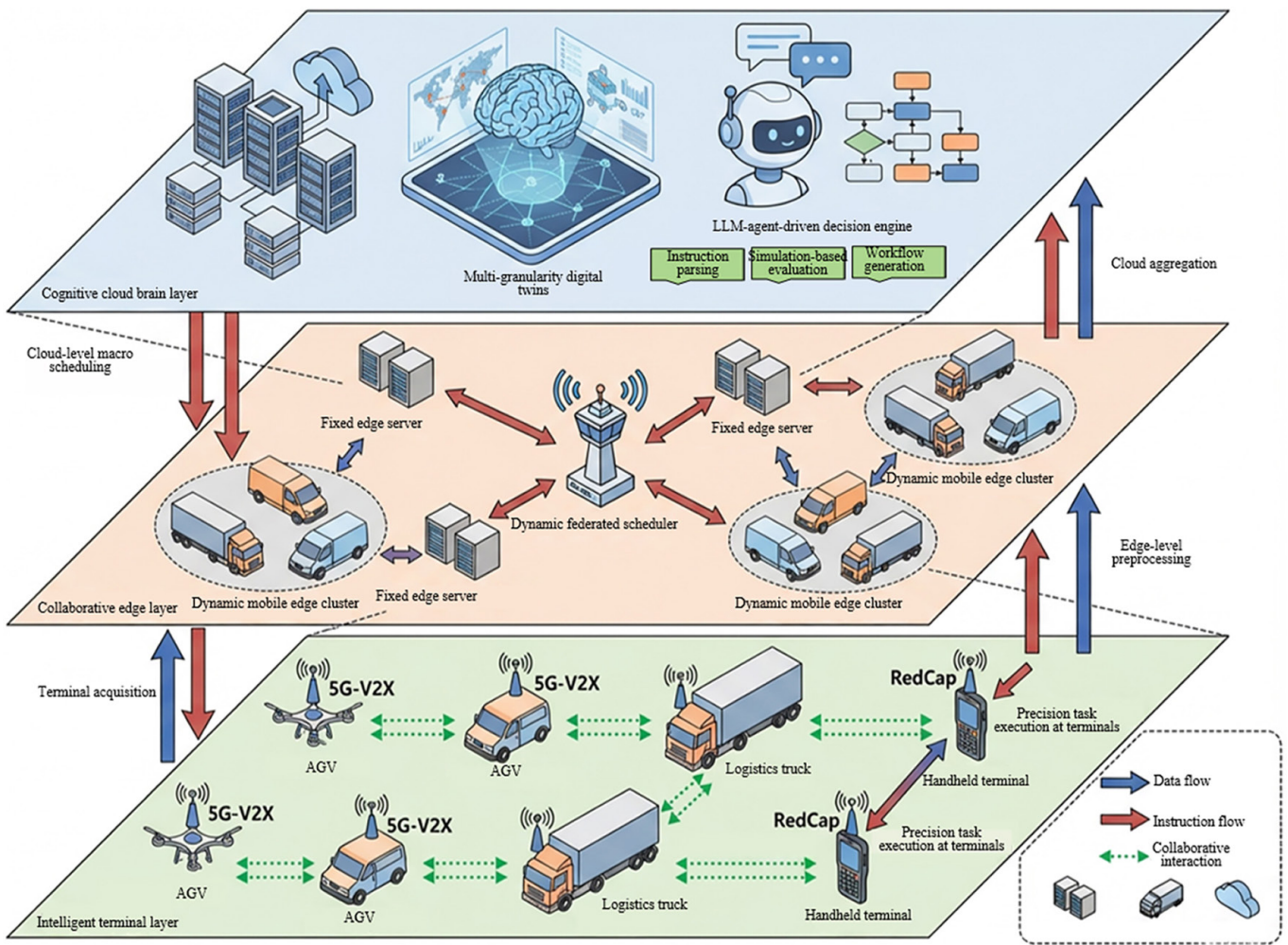


Fig. 1. Architecture of the MII-driven logistics management system

Across all layers, efficient data transmission and precise instruction dissemination are achieved through standardized interaction procedures and communication protocols, forming a tightly coordinated operational mechanism. Data transmission follows a hierarchical workflow of terminal acquisition-edge preprocessing-cloud aggregation, while instruction dissemination adopts a reverse flow of cloud-level macro scheduling-edge-level real-time decision-making-terminal-level precision execution. Communication protocols are tailored to heterogeneous terminal characteristics and collaborative requirements: 5G-V2X utilizes the PC5 interface to support low-latency vehicle-to-vehicle and vehicle-to-infrastructure communication, whereas RedCap relies on the New Radio (NR) interface to satisfy long-duration communication demands of low-power terminals. To address network instability in mobile environments, a breakpoint resumption mechanism is implemented, enabling seamless data recovery after transmission interruptions through data segmentation and checksum-based verification. In addition, a multi-network switching strategy spanning public networks, private networks, and satellite links is established to ensure communication continuity under complex environmental conditions. To further enhance latency performance and reliability in cross-layer communication, network slicing technology is employed to allocate dedicated communication resources for critical services. Emergency task instructions are prioritized through

transmission-level scheduling, while dynamic adjustment of transmission parameters is achieved by integrating link quality prediction mechanisms, enabling fine-grained optimization of communication performance.

### 3 CORE TECHNICAL IMPLEMENTATION

#### 3.1 Dynamic federated task offloading and scheduling technology

The primary objective of the dynamic federated task offloading and scheduling technology is to enable efficient task allocation and collaborative execution in highly dynamic logistics environments. This objective is achieved through multi-objective optimization modeling and mobility-aware algorithm design, in which trade-offs among latency, energy consumption, and computational resource utilization efficiency are jointly balanced. To this end, a multi-objective optimization problem is formulated, with the objective function defined as the minimization of a weighted sum of total latency, total energy consumption, and computational load balancing, expressed as:

$$\min \alpha \cdot T_{total} + \beta \cdot E_{total} + \gamma \cdot C_{balance} \quad (1)$$

where,  $T_{total}$  denotes the total task execution latency of the system,  $E_{total}$  represents the aggregated energy consumption of all participating nodes, and  $C_{balance}$  characterizes the computational load balancing degree across nodes. The weighting coefficients  $\alpha$ ,  $\beta$ , and  $\gamma$  are determined using the Analytic Hierarchy Process (AHP). In accordance with the logistics industry's priority on real-time performance and energy-efficient operation, the coefficients are set to 0.4, 0.3, and 0.3, respectively, satisfying  $\alpha + \beta + \gamma = 1$ . Constraints can be classified into four core categories. First, a task deadline constraint is imposed as  $T_i \leq T_{deadline}$ , where  $T_i$  denotes the execution latency of an individual task and  $T_{deadline}$  represents the predefined service deadline. Second, a computational capacity constraint is defined as  $\sum C_i \leq C_{max}$ , where  $C_i$  denotes the computational load assigned to a single node and  $C_{max}$  indicates the maximum computational capacity of the node. Third, a mobility constraint is introduced as  $v \leq v_{threshold}$  where  $v$  denotes the relative mobility speed among nodes within a cluster and  $v_{threshold}$  represents the maximum allowable relative speed required to maintain cluster stability. Fourth, a communication bandwidth constraint is specified as  $\sum B_i \leq B_{max}$ , where  $B_i$  denotes the bandwidth consumption of an individual task and  $B_{max}$  corresponds to the maximum available bandwidth of the communication link. All constraint parameters are quantitatively calibrated based on real-world logistics service requirements and device-level performance specifications.

To enable adaptive formation of dynamic edge clusters and precise task scheduling, a mobility-aware improved k-means clustering algorithm is designed. Joint clustering features are constructed based on node Global Positioning System (GPS) coordinates and signal strength, and adaptive cluster formation is achieved through dynamic adjustment of cluster centroids. When the inter-node distance exceeds 500 m or the signal strength falls below  $-85$  dBm, a cluster reconfiguration mechanism is triggered to maintain cluster stability under highly dynamic conditions. Multi-objective scheduling optimization is implemented using a Deep Q-Network (DQN) algorithm. The system state space  $S$  is defined to include node computational capacity, network bandwidth, and mobility speed. The action

space  $A$  consists of decisions such as target node selection for task offloading and resource allocation ratios. The reward function is formulated as  $R = -\lambda_1 T_i - \lambda_2 E_i + \lambda_3 C_{balance}$ . Through this formulation, the task offloading decision is transformed into a reinforcement learning problem, and algorithmic convergence is ensured through experience replay and target network update mechanisms. The pseudo-code of the core scheduling algorithm is designed in accordance with Science Citation Index (SCI)-standard algorithm description conventions. The time complexity is  $O(n \log n)$ , while the space complexity is  $O(n)$ . The execution time for a single scheduling round is maintained below 10 ms, thereby satisfying the real-time requirements of highly dynamic environments. To ensure task continuity under high-mobility scenarios, an edge node-following mechanism is introduced. When terminal mobility causes degradation in the service quality of the currently associated edge node, task migration is triggered. An incremental data synchronization strategy is adopted to reduce transmission overhead, and a pre-migration preparation mechanism is integrated to control migration latency within 5 ms. As a result, the task migration success rate is increased to over 99.8%, ensuring seamless task execution continuity.

### 3.2 LLM-driven dynamic decision generation technology

The core objective of the LLM-driven dynamic decision generation technology is to achieve precise mapping between logistics business requirements and system execution through fine-grained prompt engineering and agent-based architectural design while enhancing decision rationality via tool invocation and digital twin collaboration. Prompt engineering is centered on role anchoring and task formalization. Through system-level prompts, the LLM is explicitly positioned as a logistics scheduling expert, with its task scope constrained to core scheduling scenarios such as route optimization and resource allocation. Output formats are restricted to standardized structured workflows, ensuring direct compatibility with downstream execution modules. The tool invocation registry is constructed around the full decision lifecycle and includes core interfaces for route simulation, cost evaluation, and workflow generation. Invocation conditions are determined based on the LLM's semantic parsing of business instructions: when scenario simulation or multi-option comparison is required, corresponding tools are automatically triggered. In addition, parameter validation mechanisms are embedded to verify the legality and rationality of input parameters, thereby ensuring the accuracy and security of tool execution.

The decision generation process adopts a closed-loop mechanism that tightly integrates LLM with digital twin systems, forming a complete pipeline consisting of natural-language instruction input, semantic parsing by the LLM, tool invocation triggering, simulation-based evaluation via digital twins, optimal strategy generation, and workflow encapsulation for output. To enable quantitative selection among multiple decision alternatives, a decision utility evaluation function is introduced:

$$U(s) = \omega_1 \cdot \frac{1}{T(s)} + \omega_2 \cdot \left( 1 - \frac{C(s)}{C_{max}} \right) + \omega_3 \cdot R(s) \quad (2)$$

where,  $U(s)$  denotes the overall utility of decision scheme  $s$ ,  $T(s)$  represents execution latency,  $C(s)$  denotes execution cost,  $C_{max}$  is the maximum acceptable cost for the

given scenario, and  $R(s)$  indicates execution reliability. The weighting coefficients  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$  satisfy  $\omega_1 + \omega_2 + \omega_3 = 1$  and are calibrated using AHP in accordance with the real-time performance, economic efficiency, and reliability requirements of logistics scenarios. The digital twin framework provides dual support for the decision-making process. On one hand, real-time scenario data are generated through multi-granularity modeling, supplying accurate contextual information for LLM-based instruction interpretation. On the other hand, multi-strategy simulation is conducted based on real-time data, and quantitative metrics for each alternative are computed using the above utility function. These metrics serve as objective evidence for LLM-based optimal strategy selection, thereby significantly enhancing the scientific rigor of the generated decisions.

To mitigate the uncertainty risks associated with LLM outputs, a multi-layer robustness assurance framework is established. An LLM output uncertainty quantification function is introduced to assess the reliability of generated strategies:

$$U_{unc}(o) = \frac{1}{N} \sum_{i=1}^N |P(o_i) - P_{th}| \quad (3)$$

where,  $U_{unc}(o)$  denotes the uncertainty measure of LLM output  $o$ ,  $N$  represents the number of critical decision points within the output strategy,  $P(o_i)$  indicates the confidence score associated with the  $i$ -th decision point, and  $P_{th}$  denotes a predefined confidence threshold. When  $U_{unc}(o)$  exceeds the uncertainty threshold  $U_{th}$ , a three-tier fallback mechanism is automatically activated. First, a rule-based engine is triggered to invoke predefined safe scheduling strategies, ensuring baseline system operation, while a human review process is simultaneously initiated. For non-critical issues with optimization potential, prompt optimization strategies are applied to adjust input parameters and guide the LLM toward secondary decision generation. This “quantitative assessment + dual-layer safeguarding” mechanism effectively reduces uncertainty in LLM-driven decision-making and significantly enhances system robustness in complex logistics environments.

### 3.3 Privacy-enhanced cross-domain knowledge transfer technology

The privacy-enhanced cross-domain knowledge transfer technology is grounded in precise threat modeling and is designed to achieve a coordinated balance between cross-domain knowledge sharing and privacy preservation through the integration of cryptographic techniques and federated learning mechanisms. The threat model explicitly delineates three categories of core threats, together with their boundaries and attack assumptions. First, semi-honest edge nodes are assumed to strictly follow protocol execution procedures while attempting to infer sensitive original information from locally stored intermediate data. Second, a curious cloud server, while performing global parameter aggregation, is assumed to attempt reconstruction of privacy-sensitive details embedded in cross-domain knowledge from aggregated results. Third, eavesdropping attacks are considered to target data transmission links, with the aim of intercepting and analyzing transmitted data during cross-domain knowledge transfer. These attack assumptions are defined in accordance with realistic logistics scenarios, under which adversaries possess limited computational capabilities, are unable to breach fundamental system security boundaries, and do not conduct active data tampering attacks.

To address the aforementioned threats, the privacy protection scheme is implemented through the coordinated use of homomorphic encryption-based secure aggregation and an improved FedMobile algorithm. The core component is the global parameter aggregation formulation of the improved FedMobile algorithm, expressed as:

$$\theta_{global} = Dec \left( \sum_{k=1}^K w_k \cdot Enc(\theta'_k) \right) \quad (4)$$

where,  $\theta_{global}$  denotes the globally aggregated model parameters for knowledge transfer,  $Enc()$  and  $Dec()$  represents partial homomorphic encryption and decryption operations, respectively,  $K$  denotes the number of edge nodes participating in federated learning,  $w_k$  is the weight coefficient assigned to the  $k$ -th node, and  $\theta'_k$  represents the node's local model parameters after noise injection. To further mitigate privacy inference attacks, a differential privacy-based noise injection formulation is introduced to perturb the local parameters:

$$\theta'_k = \theta_k + Lap \left( \frac{\Delta f}{\epsilon} \right) \quad (5)$$

where,  $\theta_k$  denotes the original local model parameters of the node,  $Lap()$  represents the Laplace noise distribution,  $\Delta f$  indicates the global sensitivity of the model parameters, and  $\epsilon$  denotes the privacy budget. A smaller value of  $\epsilon$  corresponds to a higher level of privacy protection and a larger magnitude of noise perturbation. In practical implementation, the computational overhead of encryption and decryption is reduced through optimization of homomorphic encryption parameters. In addition, the improved FedMobile algorithm incorporates an adaptive participant selection mechanism, thereby reducing communication costs associated with mobile terminals and ensuring the feasibility of the proposed scheme.

As privacy-preserving mechanisms inevitably introduce additional system overhead, a quantitative privacy-efficiency trade-off equation is established to explicitly characterize the balance between privacy and efficiency:

$$\eta = \frac{P_{level}}{1 + \alpha D_{delay} + \beta C_{overhead}} \quad (6)$$

where,  $\eta$  denotes the comprehensive evaluation metric for the privacy-efficiency trade-off,  $P_{level}$  represents the privacy protection level quantified by  $\epsilon$  and the homomorphic encryption strength, with a normalized range of  $[0, 1]$ ,  $D_{delay}$  denotes the incremental communication latency introduced by privacy mechanisms, and  $C_{overhead}$  denotes the incremental computational overhead. The coefficients  $\alpha$  and  $\beta$  are performance loss weighting factors calibrated according to scenario-specific real-time requirements. Based on this formulation, a dynamic adaptive regulation strategy is designed. In high-sensitivity scenarios, privacy protection is prioritized by enabling high-strength encryption and small- $\epsilon$  configurations, while system overhead is mitigated through edge-side precomputation. In contrast, in real-time-critical scenarios, the value of  $\epsilon$  is moderately increased, and lightweight encryption schemes are

adopted, ensuring real-time performance within an acceptable privacy risk range. Through the explicit quantification provided by the above formulation, the trade-off relationship is clearly defined, offering a principled basis for engineering implementation and experimental validation.

## 4 EXPERIMENTAL VALIDATION

### 4.1 Experimental design

The primary objective of the experimental design is to comprehensively validate the proposed MIIoT-driven logistics management system architecture and its core technologies across five critical dimensions: real-time performance, reliability, privacy, robustness, and scalability. This ensures that the experimental protocol is explicitly aligned with the three core scientific questions and four major innovations identified earlier, thereby providing quantitative evidence for both technological effectiveness and engineering feasibility. A hybrid evaluation strategy combining simulation-based experiments and real-world deployment scenarios was adopted. The simulation environment was constructed on the NS-3 platform to emulate MIIoT communication scenarios, with parameter configurations covering mobility speeds from 0 to 120 km/h and signal-to-noise ratios ranging from  $-10$  dB to 20 dB. In parallel, a TensorFlow-based simulation framework was employed to support algorithmic performance evaluation. Through parameter calibration, the simulation scenarios were accurately mapped to real logistics environments in terms of terminal distribution, network characteristics, and service workloads. The real-world evaluation encompassed three representative logistics scenarios. In the urban distribution scenario, three logistics vehicles equipped with Huawei ME909s-521 5G modules were deployed. In the warehouse sorting scenario, five autonomous guided vehicles (AGVs) integrated with NVIDIA Jetson Xavier NX edge computing modules were utilized. In the cross-border logistics scenario, a network emulator was employed to simulate switching among public networks, private networks, and satellite links. For all real-world scenarios, complete hardware configuration specifications and physical deployment topologies were provided. The software environment was explicitly specified, including Python 3.9, PyTorch 2.0, and LangChain 0.1.0, together with their corresponding dependencies. Standardized Docker image links were provided to ensure environmental consistency and experimental reproducibility.

### 4.2 Experimental results and analysis

**Baseline performance evaluation and analysis.** Baseline performance evaluation focuses on the behavior of core system metrics across different scenarios, with the objective of validating the effectiveness of the dynamic scheduling algorithm and the dynamic edge cluster formation mechanism. The evaluation dataset encompasses both simulation-based experiments and three categories of real-world logistics scenarios. The detailed results are summarized in Table 1.

**Table 1.** Baseline performance evaluation results

Test Scenario	Variable Conditions	Task Offloading Latency (ms)	Communication Success Rate (%)	Positioning Accuracy (m)	System Energy Consumption (W)	Task Completion Rate (%)
Simulation	Speed 30 km/h, 10 terminals	18.2	99.8	0.8	42.5	99.6
	Speed 60 km/h, 10 terminals	21.5	99.5	0.9	43.2	99.5
	Speed 120 km/h, 10 terminals	27.8	98.7	1.2	45.8	99.1
	Speed 60 km/h, 30 terminals	25.3	99.2	0.9	51.3	99.3
	Speed 60 km/h, 50 terminals	32.6	98.5	1.0	58.6	98.9
Urban distribution	Real roads, 3 logistics vehicles	23.7	99.3	0.9	48.2	99.4
Warehouse sorting	Indoor warehouse, 5 AGVs	15.4	99.9	0.5	35.7	99.8
Cross-border logistics	Network switching, 2 logistics vehicles	31.2	98.3	1.1	52.6	98.7

As indicated by the results in Table 1, system performance exhibits systematic and predictable trends with increasing mobility. In the simulation scenarios, when terminal mobility increases from 30 km/h to 120 km/h, task offloading latency increases from 18.2 ms to 27.8 ms, while the communication success rate decreases from 99.8% to 98.7%, yet remains at a high level. This trend indicates that the dynamic edge cluster formation mechanism, through adaptive adjustment of clustering centroids, effectively accommodates variations in terminal mobility. Although link stability is moderately affected under high-speed conditions, no significant degradation in core system performance is observed. With increasing terminal density, task offloading latency increases correspondingly. When the number of terminals increases from 10 to 50, latency rises by 45.8%, and system energy consumption increases by 37.9%. Nevertheless, the task completion rate remains at 98.9%, demonstrating the computational load balancing capability of the dynamic scheduling algorithm. Through task decomposition and efficient offloading, system stability is maintained under high-concurrency multi-terminal conditions. In real-world evaluations, the warehouse sorting scenario achieves the best overall performance, owing to constrained mobility and a stable network environment. In contrast, the cross-border logistics scenario exhibits slightly higher latency and energy consumption due to frequent network switching. Despite this, both the communication success rate and the task completion rate consistently satisfy operational requirements, indicating strong scenario adaptability of the proposed system architecture.

**Comparative experimental results and analysis.** Comparative experiments were conducted to quantitatively evaluate the proposed approach against four representative baseline solutions across key performance metrics, thereby highlighting its technical advantages. The comparative results are illustrated in Figure 2. The selected baseline solutions include a traditional end-cloud architecture, a static edge collaborative architecture, an edge-cloud blockchain-based solution, and a federated learning solution without privacy protection.

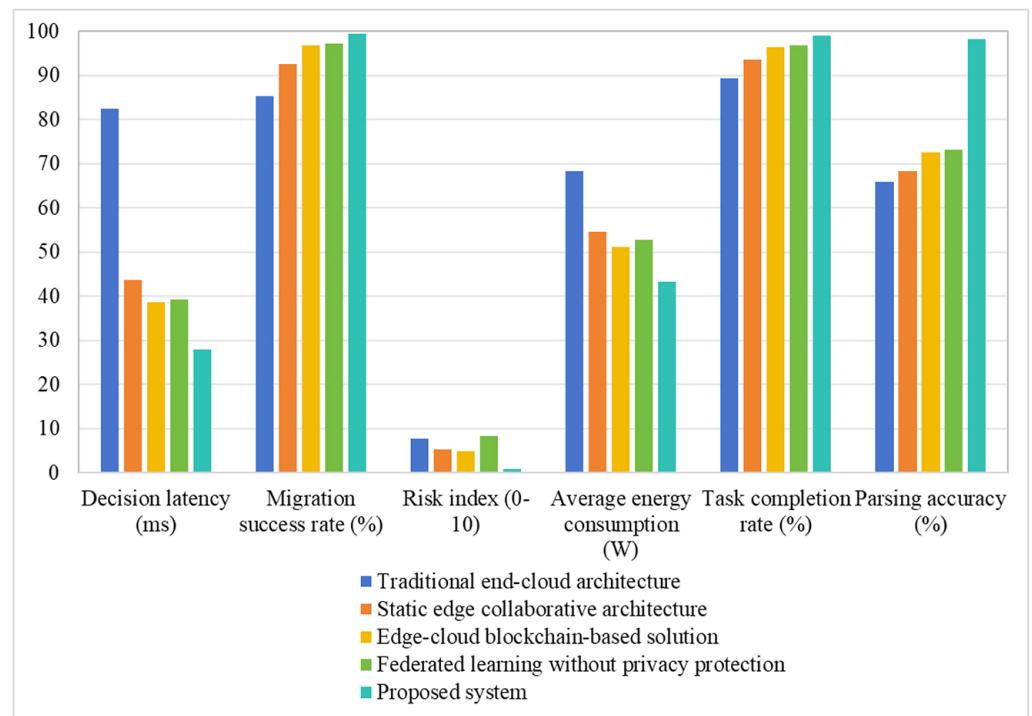


Fig. 2. Comparative experimental results

As indicated by Figure 2, the proposed system consistently outperforms all baseline solutions across all core performance metrics. In terms of edge collaborative decision latency, a reduction of 66.3% is achieved relative to the traditional end-cloud architecture, and a 28.0% reduction is observed compared with the 2024 edge-cloud blockchain-based solution. This performance gain is primarily attributed to the dynamic edge cluster mechanism, which eliminates long-distance cloud backhaul, together with the mobility-aware scheduling algorithm, which optimizes task allocation paths. The task migration success rate reaches 99.5%, representing an improvement of 2.8% over the edge-cloud blockchain-based solution. This improvement is enabled by the synergistic integration of the edge node following mechanism and the incremental data synchronization strategy, which together ensure seamless task migration in high-mobility scenarios. Privacy protection performance is particularly notable. The privacy leakage risk index is reduced to 0.9, corresponding to an 81.3% decrease relative to the best-performing baseline. This result is attributed to the combined application of homomorphic encryption-based secure aggregation and the improved FedMobile algorithm, which effectively mitigates attacks from both semi-honest nodes and curious cloud servers. The parsing accuracy for decision commands reaches 98.2%, exceeding that of the edge-cloud blockchain-based solution by 35.4%. This improvement is primarily driven by the LLM-agent-driven decision framework, in which fine-grained prompt engineering and structured tool invocation mechanisms enable precise semantic interpretation of business instructions. In addition, system energy consumption and scalability metrics exhibit superior performance, further validating the scientific soundness and practical effectiveness of the proposed “paradigm-methodology-enabling technology” innovation system.

**Extreme and failure scenario evaluation and analysis.** Extreme and failure scenario evaluations were conducted to assess system robustness under adverse conditions, encompassing network extremes, node failures, and high-load stress scenarios. The corresponding results are summarized in Table 2.

**Table 2.** Extreme and failure scenario evaluation results

Test Scenario	Test Conditions	Communication Success Rate (%)	Task Completion Rate (%)	Cluster Self-Healing Time (ms)	Task Migration Success Rate (%)	CPU Utilization (%)	Task Queuing Delay (ms)
Network extremes	Packet loss rate 30%	96.8	97.2	–	–	–	42.3
	Packet loss rate 50%	92.5	93.6	–	–	–	68.7
Node failures	Random failure of 1 node	99.2	99.4	48.6	99.6	68.3	35.2
	Random failure of 2 nodes	98.5	98.7	75.3	99.2	76.5	48.5
Load stress	5 × task volume (10 terminals)	99.6	99.5	–	–	72.4	32.8
	5 × task volume (50 terminals)	98.3	98.6	–	–	89.7	76.4
Normal operation	No extreme conditions, 20 terminals	99.7	99.6	–	99.8	56.2	24.5

Under extreme network conditions, when the packet loss rate reaches 50%, the communication success rate remains at 92.5%, with a task completion rate of 93.6%, while task queuing delay is controlled within 68.7 ms. These results demonstrate the effectiveness of the breakpoint resumption mechanism and the multi-network switching strategy, through which seamless transmission recovery is achieved via data segmentation and checksum-based verification, thereby ensuring continuity of critical services in severely degraded network environments. In node failure scenarios, even with the random failure of two edge nodes, the cluster self-healing time is limited to 75.3 ms, remaining below the predefined threshold of 100 ms, and the task migration success rate reaches 99.2%. These results indicate that the adaptive reconfiguration mechanism of dynamic edge clusters can rapidly respond to node failures by reallocating computational resources, thereby maintaining uninterrupted task execution. In load stress scenarios, when the task volume is increased by a factor of five, CPU utilization reaches 89.7% under concurrent operation of 50 terminals, with a task queuing delay of 76.4 ms, while the task completion rate remains at 98.6%. These findings confirm the system's strong scalability, demonstrating that the dynamic scheduling algorithm can effectively mitigate task surges and achieve balanced resource allocation. Compared with normal operating conditions, minor performance degradation is observed under extreme and failure scenarios; however, all key metrics remain within acceptable operational thresholds. These results validate the high robustness of the system architecture.

**Ablation study results and analysis.** Ablation experiments were conducted to quantitatively assess the necessity and contribution of the three core innovation modules by sequentially removing each component, namely the dynamic edge cluster module, the LLM-driven decision module, and the privacy-enhancement module. The corresponding results are illustrated in Figure 3.

As indicated by the results in Figure 3, each core module exerts a substantial impact on overall system performance. When the dynamic edge cluster module is removed, task offloading latency increases by 43.9%, task completion rate decreases by 3.3%, and system energy consumption rises by 12.3%. These changes demonstrate that this module constitutes a critical foundation for achieving low latency and high reliability in highly dynamic environments. Its dynamic clustering and

task migration mechanisms effectively reduce transmission overhead and balance computational loads. When the LLM-driven decision module is excluded, instruction parsing accuracy drops sharply from 98.2% to 66.5%, accompanied by a 4.9% decrease in task completion rate. This degradation confirms the essential role of the LLM-driven decision framework in accurate semantic interpretation and efficient execution of business instructions. Traditional rule-based decision mechanisms are shown to be inadequate for meeting the semantic complexity and adaptability requirements of real-world logistics scenarios. Following the removal of the privacy-enhancement module, the privacy leakage risk index increases dramatically from 0.9 to 7.6, indicating a severe deterioration in privacy protection capability. Although slight reductions in latency and energy consumption are observed, the resulting configuration fails to satisfy the privacy and security requirements associated with cross-domain data sharing. Taken together, these findings demonstrate that the three core innovation modules are mutually complementary and indispensable. Their integration collectively underpins the system's fundamental characteristics of low latency, high reliability, and privacy-enhanced operation, thereby validating the systematic coherence and rationality of the proposed innovation framework.

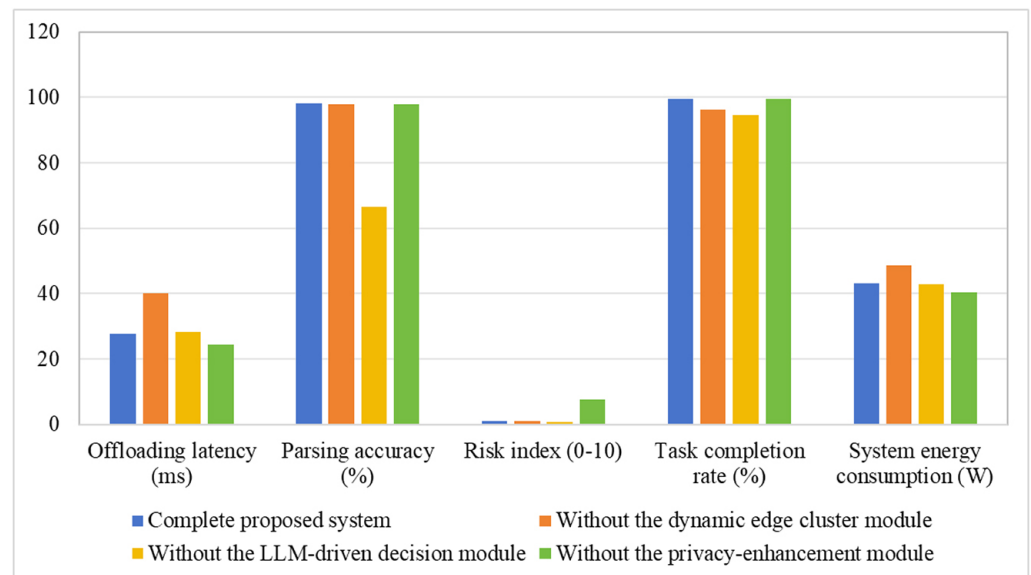


Fig. 3. Ablation study results

## 5 CONCLUSION AND OUTLOOK

To address the core challenges encountered by conventional MIoT architectures in highly dynamic logistics scenarios—namely insufficient mobility adaptation, high latency in collaborative decision-making, and elevated risks of privacy leakage—an innovative paradigm of proactive collaborative edge intelligence agents was introduced. On this basis, a three-tier hierarchical architecture comprising an intelligent terminal layer, a collaborative edge layer, and a cognitive cloud brain was constructed. Three core technologies were further developed, including dynamic federated task offloading and scheduling, LLM-driven dynamic decision generation, and privacy-enhanced cross-domain knowledge transfer, thereby forming a three-level innovation framework encompassing paradigm, methodology, and enabling technologies. Comprehensive multi-scenario experimental evaluations demonstrated

that the proposed system exhibited excellent performance adaptability under high-speed mobility, multi-terminal concurrency, and heterogeneous network switching conditions. Edge collaborative decision latency was reduced to as low as 27.8 ms, task migration success rates reached 99.5%, and the privacy leakage risk index was limited to 0.9. Comparative experiments demonstrated that key performance metrics were significantly superior to those of four representative baseline solutions, including traditional end-cloud architectures and static edge collaborative architectures. Ablation studies quantitatively verified the indispensability of the three core innovation modules, while extreme-condition and failure-scenario evaluations further confirmed the system's high robustness and scalability. The presented approach advances the theoretical foundation of collaborative intelligence in highly dynamic MIIoT scenarios and establishes a novel paradigm for the design of privacy-enhanced intelligent logistics systems. Through full-stack prototype system development and multi-scenario validation, a reproducible engineering pathway is provided to support practical deployment and industrial adoption. These contributions collectively offer substantial academic value and practical significance for promoting the deep integration of edge intelligence and the MIIoT in the logistics domain.

Building upon the limitations identified in this study, future research will be advanced along four complementary directions. First, the coverage of experimental scenarios will be expanded, with particular emphasis placed on evaluations under extreme weather conditions and large-scale terminal clusters, thereby further validating system adaptability in complex natural environments and high-concurrency settings. Second, continued efforts will be directed toward LLM lightweighting, in which model compression and quantization techniques will be employed to enable local deployment on terminal devices, thereby reducing dependence on edge and cloud computational resources. Third, industrial-scale pilot validation will be strengthened through deep collaboration with logistics enterprises, enabling the construction of real-world operational testbeds and the optimization of system alignment with practical business workflows.

## 6 REFERENCES

- [1] S. R. Talpur, A. F. Abbas, N. Khan, S. Irum, and J. Ali, "Improving opportunities in supply chain processes using the Internet of Things and blockchain technology," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 8, pp. 23–38, 2023. <https://doi.org/10.3991/ijim.v17i08.39467>
- [2] G. E. A. Eltayeb, "A proposed perspective for the successful deployment of Internet of Things in a smart home environment," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 21, pp. 117–133, 2024. <https://doi.org/10.3991/ijim.v18i21.50217>
- [3] S. Deng, H. Zhao, W. Fang, J. Yin, S. Dustdar, and A. Y. Zomaya, "Edge intelligence: The confluence of edge computing and artificial intelligence," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7457–7469, 2020. <https://doi.org/10.1109/JIOT.2020.2984887>
- [4] A. Stanković, N. Simić, and M. Stanković, "Metaheuristic optimization for stochastic job scheduling in parallel machine systems with uncertain processing and setup times," *Journal of Engineering Management and Systems Engineering*, vol. 4, no. 3, pp. 206–217, 2025. <https://doi.org/10.56578/jemse040304>
- [5] V. Pajić, M. Andrejić, and P. Chatterjee, "Enhancing cold chain logistics: A framework for advanced temperature monitoring in transportation and storage," *Mechatronics and Intelligent Transportation Systems*, vol. 3, no. 1, pp. 16–30, 2024. <https://doi.org/10.56578/mits030102>

- [6] J. Zhou *et al.*, “Stool image analysis for digital health monitoring by smart toilets,” *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 3720–3734, 2022. <https://doi.org/10.1109/JIOT.2022.3158886>
- [7] J. Liu, C. Zhou, H. Wei, J. Pi, and D. Wang, “Decoupling and collaboration: An intelligent gateway-based Internet of Things system architecture for meat processing,” *Agriculture*, vol. 15, no. 2, p. 179, 2025. <https://doi.org/10.3390/agriculture15020179>
- [8] Y. Wang, X. Wang, Y. Wang, and S. Fang, “Edge-cloud intelligence for sustainable wind turbine blade transportation: Machine-vision-driven safety monitoring in renewable energy systems,” *Energies*, vol. 18, no. 8, p. 2138, 2025. <https://doi.org/10.3390/en18082138>
- [9] M. Crespo-Aguado, R. Lozano, F. Hernandez-Goberti, N. Molner, and D. Gomez-Barquero, “Flexible hyper-distributed IoT-edge-cloud platform for real-time digital twin applications on 6G-intended testbeds for logistics and industry,” *Future Internet*, vol. 16, no. 11, p. 431, 2024. <https://doi.org/10.3390/fi16110431>
- [10] M. von Stietencron *et al.*, “Towards Logistics 4.0: An edge-cloud software framework for big data analytics in logistics processes,” *International Journal of Production Research*, vol. 60, no. 19, pp. 5994–6012, 2022. <https://doi.org/10.1080/00207543.2021.1977408>
- [11] L. Sui *et al.*, “Energy-privacy tradeoff for task matching in edge computing power networks,” *IEEE Transactions on Cloud Computing*, vol. 13, no. 4, pp. 1315–1326, 2025. <https://doi.org/10.1109/TCC.2025.3614339>
- [12] B. Li, D. D. Yang, and L. Liu, “Rotatable RIS-assisted edge computing: Orientation, task offloading, and resource optimization,” *IEEE Transactions on Vehicular Technology*, vol. 74, no. 8, pp. 13290–13295, 2025. <https://doi.org/10.1109/TVT.2025.3554227>
- [13] H. Naganawa and E. Hirata, “Enhancing policy generation with GraphRAG and YouTube data: A logistics case study,” *Electronics*, vol. 14, no. 7, p. 1241, 2025. <https://doi.org/10.3390/electronics14071241>
- [14] J. J. Romero Marras, L. De la Torre, and D. Chaos García, “WarehouseGame training: A gamified logistics training platform integrating ChatGPT, DeepSeek, and Grok for adaptive learning,” *Applied Sciences*, vol. 15, no. 12, p. 6392, 2025. <https://doi.org/10.3390/app15126392>
- [15] X. Li, W. Zhang, X. Zhao, W. Pu, p. Chen, and F. Liu, “Wartime industrial logistics information integration: Framework and application in optimizing deployment and formation of military logistics platforms,” *Journal of Industrial Information Integration*, vol. 22, p. 100201, 2021. <https://doi.org/10.1016/j.jii.2021.100201>
- [16] L. M. Belmonte, E. Segura, F. L. de la Rosa, J. L. Gómez-Sirvent, A. Fernandez-Caballero, and R. Morales, “Training industrial engineers in Logistics 4.0,” *Computers and Industrial Engineering*, vol. 184, p. 109550, 2023. <https://doi.org/10.1016/j.cie.2023.109550>

## 7 AUTHOR

**Yong Chen** earned a Ph.D. degree in Management Science and Engineering from Wuhan University of Technology, she is currently affiliated with Shangqiu Polytechnic. Her research focuses primarily on logistics management (E-mail: [13507699675@163.com](mailto:13507699675@163.com)).