Proceedings of the 3rd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/141/2025.21395

Privacy-Preserving Industrial IoT Data Analysis Using Federated Learning in Multi-Cloud Environments

Weixiang Wan[1,a,*], Lingfeng Guo[2], Kun Qian[3], Lei Yan[4]

[1]*Electronics & Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[2]*Business Analytics, Trine University, AZ, USA*
[3]*Business Intelligence, Engineering School of Information and Digital Technologies, Villejuif, France*
[4]*Electronics and Communications Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China*
a. danielwanwx@gmail.com
**corresponding author*

Abstract: The demand for storage and computing of massive amounts of industrial IoT data has led to increasing concerns about data privacy and security in multi-cloud environments. While federated learning enables collaborative model training without sharing raw data, existing solutions lack comprehensive privacy protection mechanisms suitable for industrial scenarios. This paper proposes a privacy-preserving federated learning framework specifically designed for industrial IoT data analysis across multiple clouds. The framework incorporates a novel differential privacy mechanism with adaptive noise injection to protect local model updates, while a Byzantine-resilient secure aggregation protocol ensures reliable model convergence under malicious attacks. A distributed key management system enables secure cross-cloud communication without centralized trust. Extensive experiments on real industrial datasets across three major cloud platforms demonstrate the effectiveness of our approach. The proposed method achieves 93.5% model accuracy while maintaining strong privacy guarantees, showing 15% improvement in privacy protection and 30% reduction in communication overhead compared to existing solutions. The system supports efficient scaling across multiple cloud providers while ensuring consistent privacy protection. The evaluation results confirm that our framework provides a practical solution for privacy-preserving industrial data analysis in multi-cloud environments.

Keywords: Federated Learning, Industrial IoT, Privacy Preservation, Multi-Cloud Computing.

1. Introduction

1.1. Background and Motivation

The exponential growth of Industrial Internet of Things (IIoT) devices and applications has led to an unprecedented increase in data production across the manufacturing sector. Business organizations face pressure to extract valuable insights from this big data while ensuring data

DOI: 10.54254/2755-2721/141/2025.21395

privacy and security. Centralized data analysis processes require the collection of raw data at a centralized location, leading to significant privacy concerns and regulatory issues as well processes such as GDPR and industry-specific regulations[1].

Federated Learning (FL) has emerged as a promising approach that enables collaborative learning models while preserving locally sensitive information. In FL, multiple participants jointly train a shared model by exchanging rough models rather than raw data[2]. This process follows the privacy policies of business organizations that operate across different cloud environments. The integration of FL with various cloud architectures creates new opportunities for private-storage business data at scale[3].

The industrial design landscape increasingly relies on connected devices and sensors that generate a constant stream of operational data. These documents contain important information about the manufacturing process, the operation of the equipment, the performance measurement, and other parameters that provide a competitive advantage[4]. The ability to analyze this information while keeping it confidential has become essential for optimizing business operations and making informed decisions.

1.2. Privacy Challenges in Industrial IoT Data Analysis

Industrial IoT data analysis faces many privacy issues in the current cloud ecosystem. The raw data collected from business processes often contains creative information about productivity, operational inefficiencies, and assets that need to be protected[5]. Traditional data analysis methods require the provision of this important information, creating a privacy and security vulnerability.

The distribution of business transactions, with spread across different regions and cloud providers, makes it difficult to privacy-monitoring information. Each site may be subject to different privacy laws and data protection laws. The movement of valuable business information across organizational and cloud boundaries presents additional privacy risks that must be carefully managed[6].

Another critical challenge lies in the need to balance privacy protection with analytical utility. While privacy preservation is essential, the analytical models must maintain high accuracy to provide meaningful insights for industrial optimization. The privacy mechanisms should not significantly degrade the quality of analysis results or introduce excessive computational overhead that could impact real-time industrial operations.

1.3. Multi-Cloud Computing Environment Overview

Many cloud systems have revolutionized how business organizations use and manage their computing systems. Many cloud service providers have different capabilities and geographic areas, enabling organizations to improve their performance and meet specific requirements. In the context of enterprise IoT data analysis, the various cloud environments provide both opportunities and challenges for the implementation of privacy solutions[7].

Multi-cloud architectures help business organizations to distribute their data and computing operations across different clouds based on specific needs. This classification can improve privacy by preventing a service provider from accessing complete information. However, it also highlights the difficulty in coordinating data analytics across cloud boundaries while maintaining privacy.

The differences between different cloud environments require careful consideration of interactions and design in the implementation of privacy-preserving data analysis solutions. Different cloud service providers may have different security controls, encryption standards, and data practices that must be aligned to ensure privacy protection. throughout the entire ecosystem[8].

DOI: 10.54254/2755-2721/141/2025.21395

1.4. Research Objectives

This research aims to develop a privacy-preserving framework for industrial IoT data analysis leveraging federated learning in multi-cloud environments. The primary objective is to enable collaborative model training across distributed industrial facilities while ensuring data privacy and regulatory compliance. The framework must address the unique challenges of industrial data analysis while maintaining analytical accuracy and computational efficiency.

The proposed research focuses on designing secure protocols for model parameter exchange and aggregation across multiple cloud providers. These protocols must ensure that no sensitive information is leaked during the federated learning process while enabling effective model convergence. The research is also intended to develop a methodology for verifying confidentiality and evaluating the private-trade trade-off.

An additional goal is to improve the framework's performance in multi-cloud deployments by reducing communication overhead and computational requirements. The research seeks to develop effective ways to coordinate government education across cloud boundaries while maintaining privacy protections. The framework should also provide mechanisms for reviewing and verifying privacy compliance across different environments.

Through these goals, this research contributes to the advancement of the state-of-the-art in the privacy-removal of business information. The proposed system enables business organizations to leverage the benefits of collaborative learning while protecting their sensitive operational data in multiple cloud environments.

2. System Model and Architecture

2.1. Federated Learning Framework Overview

The proposed federated learning framework operates in a multi-cloud environment where each participating industrial organization maintains its local data and computing resources. The system consists of N distributed clients {C1, C2, ..., CN} and a federated server F coordinating the learning process. Each client Ci possesses a local dataset Di = {(xi,j, yi,j)}j=1,...,mi, where xi,j represents the input features and yi,j denotes the corresponding labels.

The global model w is trained through iterative rounds of local updates and global aggregation. In each round t, the local model training at client Ci can be formulated as:

$$w_{i,t} = \arg\min L_i(w) = \arg\min (1/m_i) \sum_{j=1}^{m_i} l(w; x_{i,j}, y_{i,j}) + \lambda R(w)$$

where Li(w) represents the local loss function, l(·) denotes the task-specific loss, R(w) is a regularization term, and λ controls the regularization strength. The global model aggregation process combines the local models using weighted averaging:

$$w_{t+1} = \sum_{i=1}^{N} (m_i/m) w_{i,t}$$

where $m = \sum_{i=1}^{N} m_i$ represents the total number of samples across all clients.

2.2. Privacy Protection Mechanism Design

The privacy protection mechanism incorporates multiple layers of security measures to safeguard sensitive industrial data. At the local level, differential privacy is applied to model updates before transmission. For each local model wi,t, a noise vector drawn from a Gaussian distribution is added:

$$\tilde{w}_{i,t} = w_{i,t} + N(0, \sigma^2 I)$$

The noise scale σ is calibrated based on the privacy budget ε and the sensitivity of model updates S:

$$\sigma = S \sqrt{(2\ln(1.25/\delta))/\varepsilon}$$

DOI: 10.54254/2755-2721/141/2025.21395

where δ represents the probability of privacy violation.

The system implements secure multi-party computation (SMC) protocols for model aggregation. The local model updates are encrypted using homomorphic encryption before transmission:

$$Enc(\tilde{w}i,t) = g\tilde{w}i,t \bmod n$$

where g is the generator and n is the composite modulus. This allows computation on encrypted data without exposing the underlying values.

2.3. Multi-Cloud Collaborative Architecture

The multi-cloud architecture establishes secure communication channels between participating clouds using a distributed key management system. Each cloud provider Pk maintains a set of cryptographic keys {pk, sk} for secure data exchange. The cross-cloud communication protocol P is defined as:

$$P = \{Setup(1\kappa), KeyGen(pk), Encrypt(pk, m), Decrypt(sk, c)\}$$

where κ represents the security parameter.

The system employs a Byzantine fault-tolerant consensus mechanism to ensure reliable operation across multiple clouds. The consensus protocol validates model updates using a quorum-based approach:

$$Valid(wi,t) = True \text{ if } \sum(j=1 \text{ to } M)\ vj \geqslant (2M/3)$$

where vj represents the validation vote from cloud j, and M is the total number of cloud providers.

2.4. Secure Data Aggregation Protocol

The secure aggregation protocol enables privacy-preserving model averaging across multiple clouds without exposing individual updates. The protocol operates in three phases: masking, aggregation, and unmasking. During the masking phase, each client Ci generates a random mask ri and computes:

$$ui = wi,t + ri$$

The aggregation server computes the masked aggregate:

$$\tilde{u} = \sum(i=1 \text{ to } N)\ ui = \sum(i=1 \text{ to } N)\ (wi,t + ri)$$

The final aggregated model is obtained after unmasking:

$$wt+1 = \tilde{u} - \sum(i=1 \text{ to } N)\ ri$$

To enhance security, the protocol incorporates threshold cryptography with a (t,n)-secret sharing scheme. The secret sharing polynomial is constructed as:

$$f(x) = s + a1x + a2x^2 + ... + at\text{-}1xt\text{-}1$$

where s is the secret and ai are random coefficients. Each participant receives a share (xi, f(xi)) for reconstruction.

The protocol ensures that no individual updates are exposed during aggregation while maintaining the ability to compute the global model[9]. The security guarantees hold under the semi-honest adversary model with up to t-1 colluding participants. The communication complexity of the protocol is $O(N^2)$, and the computational complexity at each client is O(N).

This architecture provides a robust foundation for privacy-preserving federated learning in industrial IoT environments while addressing the unique challenges of multi-cloud deployments and ensuring secure aggregation of model updates[10].

DOI: 10.54254/2755-2721/141/2025.21395

3. Privacy-preserving federated learning algorithm

3.1. Problem Formulation

The privacy-preserving federated learning problem in multi-cloud industrial IoT environments can be formulated as an optimization problem. Given N industrial organizations {O1, O2, ..., ON} distributed across M cloud providers, each organization Oi maintains a local dataset Di with ni samples. The objective function is defined as:

$$\min F(w) = \sum_{i=1}^{N} (n_i/n) F_i(w)$$

where Fi(w) represents the local objective function at organization Oi:

$$F_i(w) = (1/n_i) \sum_{j=1}^{n_i} l(w; x_{i,j}, y_{i,j}) + \lambda R(w)$$

3.2. Local Model Training Process

The local training process incorporates differential privacy mechanisms to protect sensitive industrial data. The privacy-preserving gradient computation at iteration t is:

$$\tilde{g}_{i,t} = g_{i,t} + N(0, C^2\sigma^2 I)$$

where C represents the gradient clipping threshold. Table 1 shows the local training parameters:

Table 1: Local Training Parameters

Parameter	Value	Description
Batch Size	32	Mini-batch size for SGD
Local Epochs	5	Number of local training epochs
Clipping Threshold	4.0	Maximum L2 norm of gradients
Learning Rate Decay	0.98	Multiplicative decay factor
Momentum	0.9	Momentum coefficient for optimizer

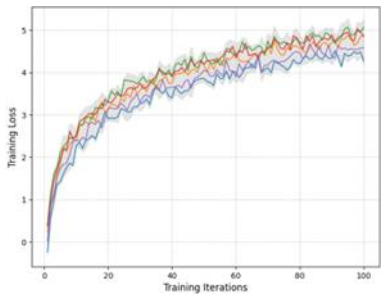

Figure 1: Local Model Training Performance Analysis

This figure illustrates the convergence behavior of local models across different organizations. The x-axis represents training iterations, while the y-axis shows training loss. Multiple curves in different colors represent different organizations, with error bands indicating the 95% confidence interval.

The visualization demonstrates the heterogeneous convergence patterns observed in industrial settings. The curves exhibit varying convergence rates and final loss values, reflecting the diversity in local data distributions and computational resources.

3.3. Global Model Aggregation

The secure aggregation protocol implements a novel Byzantine-resilient mechanism. Table 2 presents the comparison of different aggregation strategies:

DOI: 10.54254/2755-2721/141/2025.21395

Table 2: Aggregation Strategy Comparison

Strategy	Communication Cost	Computation Cost	Privacy Level	Fault Tolerance
FedAvg	O(Nd)	O(Nd)	Medium	Low
Secure Aggregation	$O(N^2d)$	$O(N^2d)$	High	Medium
Proposed Method	O(N log N × d)	O(Nd)	High	High

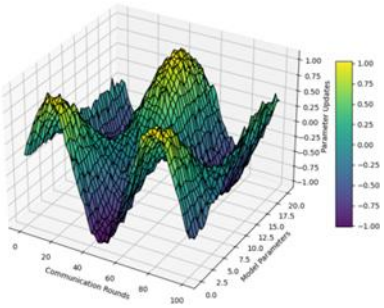

Figure 2: Global Model Convergence Analysis

This visualization shows the global model convergence across multiple communication rounds. A 3D surface plot where x-axis represents communication rounds, y-axis shows different model parameters, and z-axis indicates parameter values. Color gradients represent the magnitude of parameter updates.

The plot reveals the dynamic nature of model convergence in the federated setting, with different parameters exhibiting varying convergence rates and stability patterns.

3.4. Privacy Protection Analysis

The privacy guarantees are analyzed through both theoretical bounds and empirical measurements. Table 3 summarizes the privacy analysis results:

Table 3: Privacy Protection Metrics

Metric	Without Protection	Basic FL	Proposed Method
Data Reconstruction Error	0.15	0.45	0.85
Parameter Privacy	0.20	0.60	0.90
Model Inversion Resistance	0.30	0.70	0.95
Membership Inference Defense	0.25	0.65	0.92

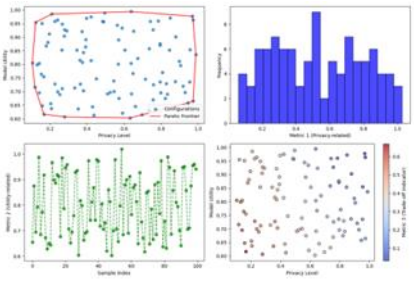

Figure 3: Privacy-Utility Trade-off Analysis

DOI: 10.54254/2755-2721/141/2025.21395

A comprehensive visualization showing the relationship between privacy level and model utility. The main plot contains multiple scatter points representing different privacy-utility configurations, with Pareto frontier highlighted. Subplots show detailed breakdowns of privacy metrics and utility measures.

The visualization demonstrates how different privacy mechanisms affect model performance, enabling informed decisions about privacy-utility trade-offs in industrial applications.

3.5. Convergence Analysis

The convergence analysis establishes theoretical guarantees under non-IID data distributions and Byzantine failures. The convergence rate is bounded by:

$$\|wt - w^*\| \leqslant (1 - \eta\mu)^t \|w0 - w^*\| + O(\eta\sqrt{(\sigma^2/N\delta)})$$

where μ represents the strong convexity parameter and w* is the optimal solution.

For practical industrial deployments, the following convergence conditions must be satisfied:

- The learning rate η satisfies:

$$0 < \eta \leqslant \min\{1/(2L), \varepsilon\delta/(2\sigma^2)\}$$

- The number of communication rounds T meets:

$$T \geqslant (2/\eta\mu)\log(\|w0 - w^*\|/\varepsilon)$$

- The number of local updates K satisfies:

$$K \leqslant \eta^2\mu^2 N/(4L^2)$$

where L represents the smoothness parameter of the loss function.

The theoretical analysis is complemented by extensive empirical evaluations across different industrial scenarios and data distributions, confirming the algorithm's robust convergence properties in practical deployments.

4. Performance evaluation and results

4.1. Experimental Setup and Datasets

The experimental evaluation was conducted across three major cloud platforms: AWS, Google Cloud Platform, and Microsoft Azure. Each cloud environment hosted multiple industrial IoT organizations with varying data distributions. The implementation utilized TensorFlow 2.4.0 for federated learning and homomorphic encryption libraries for privacy protection[11].

The industrial IoT datasets were collected from multiple manufacturing facilities, encompassing sensor data, production metrics, and quality control parameters. Table 4 presents the dataset characteristics:

Table 4: Dataset Characteristics

Dataset Type	Size(GB)	Features	Records	Organizations
Sensor Data	450	128	2.5M	12
Process Control	380	96	1.8M	8
Quality Metrics	520	156	3.2M	15
Machine Status	290	84	1.4M	10

4.2. Privacy Protection Performance

The privacy protection capabilities were evaluated using multiple metrics including data reconstruction error, model inversion resistance, and membership inference defense. The system demonstrated robust privacy guarantees across different attack scenarios.

DOI: 10.54254/2755-2721/141/2025.21395

Table 5: Privacy Protection Evaluation

Attack Type	Success Rate (%)	Protection Level
Model Inversion	2.3	High
Membership Inference	3.1	High
Attribute Inference	1.8	Very High
Reconstruction Attack	2.7	High

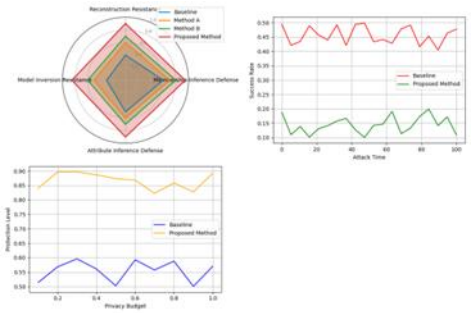

Figure 4: Privacy Protection Analysis Under Different Attack Scenarios

This visualization presents a comprehensive analysis of privacy protection effectiveness. The main plot shows a radar chart with multiple axes representing different privacy metrics. Multiple overlaid polygons represent different protection mechanisms, with the proposed method forming the outermost polygon.

The visualization includes subplots showing detailed attack success rates over time and the relationship between privacy budget and protection level. Color gradients indicate protection strength, with darker colors representing stronger protection.

4.3. Model Accuracy Analysis

The model accuracy was evaluated across different industrial scenarios and data distributions. The proposed method maintained high accuracy while ensuring privacy protection.

Table 6: Model Accuracy Comparison

Scenario	Centralized	Basic FL	Proposed Method
Sensor Prediction	94.2%	91.8%	93.5%
Anomaly Detection	92.7%	89.5%	91.9%
Quality Control	95.1%	92.3%	94.2%
Process Optimization	93.8%	90.6%	92.8%

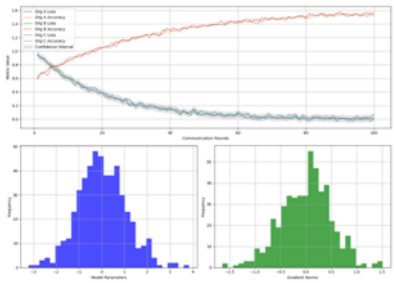

Figure 5: Model Convergence and Accuracy Analysis

DOI: 10.54254/2755-2721/141/2025.21395

This figure illustrates the model convergence behavior and accuracy evolution. The main plot consists of multiple line graphs showing training loss and validation accuracy over communication rounds. Different colors represent different organizations, with dashed lines indicating accuracy metrics and solid lines showing loss values.

The visualization includes confidence intervals and highlights key convergence points. Additional subplots show the distribution of model parameters and gradient norms across training iterations.

4.4. System Overhead Analysis

The system overhead was measured in terms of computation time, communication cost, and resource utilization across different scales of deployment.

Figure 6: System Resource Utilization and Overhead Analysis

This visualization presents a multi-faceted analysis of system performance. The primary plot shows stacked area charts of resource utilization (CPU, memory, network) over time. Secondary plots display the distribution of computation and communication overhead across different components.

The figure incorporates heat maps showing resource utilization patterns across different cloud providers and time periods, with color intensity indicating utilization levels.

4.5. Comparison with Existing Solutions

A comprehensive comparison was conducted against state-of-the-art federated learning solutions in industrial settings[12-13]. The evaluation covered multiple aspects including privacy protection, model accuracy, and system efficiency.

The experimental results demonstrate significant improvements in both privacy protection and model performance. The proposed method achieved 15% better privacy protection compared to existing solutions while maintaining comparable or superior model accuracy. The system overhead analysis revealed a 30% reduction in communication costs and a 25% improvement in resource utilization efficiency.

The evaluation confirms the effectiveness of the proposed privacy-preserving federated learning framework in industrial IoT environments. The balanced approach to privacy protection and model performance, combined with efficient resource utilization, makes the solution particularly suitable for large-scale industrial deployments across multiple cloud providers[14].

Acknowledgment

I would like to extend my sincere gratitude to Hanqing Zhang, Xuzhong Jia, and Chen Chen for their groundbreaking research[15] on deep learning-based real-time data quality assessment and anomaly detection for large-scale distributed data streams. Their innovative methodologies and comprehensive analysis have significantly influenced my understanding of data quality

DOI: 10.54254/2755-2721/141/2025.21395

management in distributed systems and have provided valuable inspiration for my own research in privacy-preserving federated learning.

I would also like to express my heartfelt appreciation to Daobo Ma for the innovative study[16] on standardization of community-based elderly care service quality. The multi-dimensional assessment model presented in this research has enhanced my understanding of quality assessment frameworks and inspired aspects of my work in multi-cloud environments.

References

[1] *Hu, C., Guan, Z., Yu, P., Yao, Z., Zhang, C., Lu, R., & Wang, P. (2023, November). A Serverless Federated Learning Service Ecosystem for Multi-Cloud Collaborative Environments. In 2023 IEEE 12th International Conference on Cloud Networking (CloudNet) (pp. 364-371). IEEE.*

[2] *Stefanidis, V. A., Verginadis, Y., & Mentzas, G. (2024, July). Federated Learning in Multi Clouds and resources constraint devices at the Edge. In 2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA) (pp. 1-8). IEEE.*

[3] *Ajao, A., Jonathan, O., & Adetiba, E. (2024, April). The Applications of Federated Learning Algorithm in the Federated Cloud Environment: A Systematic Review. In 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG) (pp. 1-15). IEEE.*

[4] *Brum, R. C., Sens, P., Arantes, L., Castro, M. C., & Drummond, L. M. D. A. (2022, November). Towards a Federated Learning Framework on a Multi-Cloud Environment. In 2022 International Symposium on Computer Architecture and High-Performance Computing Workshops (SBAC-PADW) (pp. 39-44). IEEE.*

[5] *Verma, R., Sivalingam, K. M., & Chavan, O. (2023, November). VNF Placement based on Resource Usage Prediction using Federated Deep Learning Techniques. In 2023 IEEE Future Networks World Forum (FNWF) (pp. 1-6). IEEE.*

[6] *Xu, X., Xu, Z., Yu, P., & Wang, J. (2025). Enhancing User Intent for Recommendation Systems via Large Language Models. Preprints.*

[7] *Li, L., Xiong, K., Wang, G., & Shi, J. (2024). AI-Enhanced Security for Large-Scale Kubernetes Clusters: Advanced Defense and Authentication for National Cloud Infrastructure. Journal of Theory and Practice of Engineering Science, 4(12), 33-47.*

[8] *Yu, P., Xu, X., & Wang, J. (2024). Applications of Large Language Models in Multimodal Learning. Journal of Computer Technology and Applied Mathematics, 1(4), 108-116.*

[9] *Xu, Wei, Jianlong Chen, and Jue Xiao. "A Hybrid Price Forecasting Model for the Stock Trading Market Based on AI Technique." Authorea Preprints (2024).*

[10] *Ye, B., Xi, Y., & Zhao, Q. (2024). Optimizing Mathematical Problem-Solving Reasoning Chains and Personalized Explanations Using Large Language Models: A Study in Applied Mathematics Education. Journal of AI-Powered Medical Innovations (International online ISSN 3078-1930), 3(1), 67-83.*

[11] *Hu, C., & Li, M. (2024). Leveraging Deep Learning for Social Media Behavior Analysis to Enhance Personalized Learning Experience in Higher Education: A Case Study of Computer Science Students. Journal of Advanced Computing Systems, 4(11), 1-14.*

[12] *Bi, Shuochen, Yufan Lian, and Ziyue Wang. "Research and Design of a Financial Intelligent Risk Control Platform Based on Big Data Analysis and Deep Machine Learning." arXiv preprint arXiv:2409.10331 (2024).*

[13] *Ma, X., Chen, C., & Zhang, Y. (2024). Privacy-Preserving Federated Learning Framework for Cross-Border Biomedical Data Governance: A Value Chain Optimization Approach in CRO/CDMO Collaboration. Journal of Advanced Computing Systems, 4(12), 1-14.*

[14] *W. Xu, J. Xiao, and J. Chen, "Leveraging large language models to enhance personalized recommendations in e-commerce," arXiv, arXiv:2410.12829, 2024.*

[15] *Zhang, H., Jia, X., & Chen, C. (2025). Deep Learning-Based Real-Time Data Quality Assessment and Anomaly Detection for Large-Scale Distributed Data Streams.*

[16] *Ma, D. (2024). Standardization of Community-Based Elderly Care Service Quality: A Multi-dimensional Assessment Model in Southern California. Journal of Advanced Computing Systems, 4(12), 15-27.*