China and WTO Review (CWR) Journal
2026, xx(x)
ISSN: 2384-4388
https://cwto.net

Research Article

Exposure-Aware Protection of Edge-to-Cloud Data Synchronization Workflows

Emily J. Carter [1], Benjamin R. Hughes [1], Olivia M. Bennett [1*]

[1] School of Computing and Information Systems, Faculty of Engineering and Information Technology, The University of Melbourne, Parkville, Australia
* **Corresponding Author**: o.bennett@unimelb.edu.au

ARTICLE INFO

ABSTRACT

Cloud edge architectures rely on continuous synchronization between edge gateways, regional cloud services, storage systems, and central analytics platforms. Vulnerabilities in synchronization agents, transfer protocols, gateway services, or access policies may expose sensitive edge data or allow unauthorized modification during transmission. This study proposes a policy-aware vulnerability risk quantification framework for cloud edge data synchronization pipelines. The framework evaluates edge-node sensitivity, synchronization frequency, transfer encryption, gateway vulnerability status, data classification, access-token scope, retry behavior, and policy violations. A synchronization-lineage model is built to connect vulnerable edge components with upstream data sources and downstream cloud storage assets. Experiments are conducted on a cloud edge environment containing 1,340 edge gateways, 9,780 synchronization jobs, 4,620 regional storage endpoints, 2,150 analytics ingestion services, 11,900 access tokens, and 21,300 vulnerability records from gateway images, transfer agents, and synchronization APIs. The proposed framework identifies 1,920 data-synchronization risk chains, including vulnerable transfer agents handling sensitive sensor data, weakly encrypted synchronization channels, over-scoped access tokens, and exposed regional ingestion endpoints. It groups 21,300 raw findings into 2,760 synchronization-level remediation units based on data lineage and policy impact. The median scoring time is 46 ms per synchronization job, while the full risk assessment completes in 14.9 minutes. During remediation testing, the first 420 prioritized fixes remove 980 vulnerable data-transfer chains and reduce exposed regional ingestion endpoints from 386 to 121. The framework processes 14,600 gateway-policy relationships per minute during incremental updates. The results demonstrate that policy-aware vulnerability quantification can support risk-based protection of cloud edge data movement, especially where data security depends on synchronization paths, gateway exposure, and access-token control.

Keywords: Cloud Edge Security, Data Synchronization Risk, Vulnerability Risk Quantification, Policy-Aware Risk Scoring, Edge Gateway Security, Data Transfer Protection, Cloud Infrastructure Security

INTRODUCTION

Cloud – edge computing has become an important architecture for IoT, industrial sensing, smart manufacturing, and distributed analytics. In these systems, data are collected from edge devices and gateways, temporarily processed or stored in regional cloud nodes, and then synchronized with central storage, data lakes, or analytics platforms. This architecture reduces communication delay, lowers backbone network load, and supports near-real-time decision making in distributed environments (Gui et al., 2025; Souza et al., 2024). However, it also extends the data-transfer surface across edge gateways, synchronization agents, access tokens, storage endpoints, ingestion services, and cloud-side analytics components. Recent work on open-source cloud infrastructure security has shown that vulnerability information, deployment context, policy checks, and remediation logic can be integrated into a practical security framework for cloud environments (Shao, 2026). Related studies on edge and IoT – cloud systems have also reported security risks caused by weak authentication, insecure APIs, exposed gateways, vulnerable containers, and unprotected communication channels (Chen et al., 2025; Ademilua, 2021). In industrial IoT and cyber – physical systems, these risks are more serious because attacks on edge gateways or data-transfer layers may affect not only data confidentiality but also production continuity, process integrity, and

operational safety (Latif et al., 2025; Liang et al., 2025).

Existing research has improved the protection of cloud – edge environments from several directions. Access-control and zero-trust studies have strengthened authentication and authorization through role-based, attribute-based, capability-based, and policy-driven mechanisms (Zhu & Yang, 2025; Liu et al., 2026). Other studies have focused on secure communication, device identity, edge trust management, and containerized workload protection. These approaches are useful for reducing single-point security weaknesses, but they often evaluate access policies, vulnerabilities, or data-protection controls as separate issues (Miah & Islam, 2021; Yin et al., 2026). In real cloud – edge synchronization pipelines, however, security exposure is usually produced by the combination of multiple factors. A vulnerable synchronization agent may become high risk when it runs on an internet-exposed gateway, uses an over-scoped token, transfers sensitive data without sufficient encryption, and writes results into downstream analytics storage (Anderson & Hossain, 2026; Yang, 2026). Conversely, a vulnerability with a high technical severity score may have limited practical impact if the affected component is isolated, the token scope is narrow, and the synchronized data have low sensitivity. Therefore, an effective assessment method should not only identify vulnerable components but also explain how these components are connected to downstream data assets through actual synchronization paths. Recent vulnerability management studies have moved from static severity scores toward risk-based prioritization. EPSS and related models estimate exploitation likelihood, while other approaches incorporate exploit evidence, asset context, dependency relationships, and remediation constraints into vulnerability ranking (Deng & Yang, 2026; Zhang, 2026). Graph-based cloud security models further show that cloud risks often arise from relations among users, roles, policies, data stores, services, and configurations rather than from isolated findings(Poornimadarshini, 2025; Zhuwankinyu et al., 2025). Data provenance and lineage studies provide another useful foundation by tracing how data are generated, transferred, transformed, and consumed across distributed systems (Zhang et al., 2026; Trehan & Pradhan, 2024). These studies offer important concepts for contextual risk analysis, but they remain insufficient for cloud – edge synchronization security. Most existing methods do not jointly model vulnerability severity, gateway exposure, synchronization frequency, token privilege, encryption state, data classification, retry behavior, and policy violations in a unified scoring process. As a result, security teams may receive long lists of raw vulnerability findings without knowing which synchronization chains create the greatest data-security exposure. Existing datasets and testbeds also have limitations for this problem. Public edge and industrial IoT datasets, such as Edge-IIoTset, mainly support intrusion detection, device-level attack classification, or network-traffic analysis (Qi & Qiao, 2026; Su & Wu, 2026). They are valuable for evaluating detection models, but they do not fully represent large-scale synchronization-lineage risks across gateways, transfer agents, regional cloud storage, access tokens, and ingestion services. In practical cloud – edge systems, the same vulnerability may have different risk levels depending on the data being synchronized, the access scope of the token, the frequency of the synchronization job, the encryption status of the transfer channel, and the policy requirements attached to the target data asset. This mismatch between available datasets and real synchronization scenarios limits the ability of existing methods to support remediation prioritization in operational cloud – edge environments (Paulachan et al., 2025; Manduva, 2020).

To address this gap, this study proposes a policy-aware vulnerability risk quantification framework for cloud – edge data synchronization pipelines. The framework links edge gateways, synchronization jobs, transfer agents, access tokens, data classifications, regional storage endpoints, and analytics ingestion services into a synchronization-lineage model. It then quantifies risk by considering gateway exposure, transfer protection, token privilege, retry behavior, synchronization frequency, data sensitivity, vulnerability severity, exploitation likelihood, and policy impact. The purpose is to identify high-risk data-synchronization chains rather than merely rank isolated vulnerabilities. The proposed framework also groups related findings into actionable remediation units, enabling security teams to prioritize fixes that reduce the greatest downstream data-security exposure. By connecting vulnerability management, data lineage, access-token analysis, and policy-aware scoring, this study provides a more practical basis for securing cloud – edge synchronization pipelines in IoT, industrial sensing, and distributed analytics environments.

MATERIALS AND METHODS

Sample and Study Environment Description

This study was carried out in a cloud – edge data synchronization environment that included edge data collection, regional cloud storage, and central analytics ingestion. The sample included 1,340 edge gateways, 9,780 synchronization jobs, 4,620 regional storage endpoints, 2,150 analytics ingestion services, 11,900 access tokens, and 21,300 vulnerability records. The vulnerability records were collected from gateway images, transfer agents,

synchronization APIs, and cloud service configurations. The edge gateways covered several deployment types, including industrial sensor hubs, smart-building gateways, traffic monitoring nodes, and environmental monitoring devices. Each synchronization job was used as the basic unit of analysis. For each job, the source gateway, target endpoint, synchronization interval, retry setting, transfer protocol, encryption mode, token scope, data class, and policy status were recorded. A job was included only when it had a valid source node, a downstream storage or ingestion target, and complete security and policy metadata. Jobs with missing lineage records, expired identifiers, or repeated transfer definitions were removed before analysis.

Experimental Design and Control Settings

The experiment compared the policy-aware synchronization-level risk model with three common assessment methods. The experimental group used the policy-aware model. In this model, vulnerability severity was assessed together with synchronization lineage, data sensitivity, transfer encryption, gateway exposure, access-token scope, retry behavior, and policy violations. Three control groups were used. The first control group ranked findings only by their original vulnerability severity scores. The second control group used asset-level exposure scoring, in which gateways and cloud endpoints were ranked by vulnerability count and external accessibility. The third control group used policy violation count, in which synchronization jobs were ranked by the number of non-compliant access or transfer settings. These control settings were selected because they represent common practices in vulnerability management, infrastructure exposure analysis, and compliance checking. The comparison was used to examine whether policy-aware lineage scoring could identify remediation targets with higher data-security impact than severity-only, asset-only, or rule-count-based methods.

Measurement Methods and Quality Control

Security data were collected from vulnerability scanning reports, gateway image inspection results, synchronization-agent logs, cloud access-policy records, token metadata, encryption configuration files, and storage endpoint inventories. Vulnerability records were standardized by component identifier, vulnerability ID, affected package or API, severity level, exploitability label, and discovery time. Synchronization lineage was rebuilt by linking source gateways, synchronization jobs, transfer agents, access tokens, storage endpoints, and analytics ingestion services through job identifiers and service-account mappings. Data sensitivity was divided into four levels: public, internal, confidential, and restricted. Transfer protection was measured according to protocol type, encryption status, certificate validity, and downgrade allowance. Token scope risk was measured by comparing assigned permissions with the minimum permissions required by the corresponding synchronization job. Quality control was performed in three steps. First, duplicate vulnerability records were removed according to component ID, vulnerability ID, and package version. Second, inconsistent policy records were checked against cloud audit logs and access-control snapshots. Third, lineage completeness was verified by requiring each retained record to have a valid upstream source, an active synchronization job, and a downstream target. In addition, a random 5% subset of synchronization jobs was manually checked to confirm the accuracy of lineage reconstruction and policy labeling.

Data Processing and Model Formulation

All raw records were converted into synchronization-level risk units before model calculation. Each vulnerability finding was mapped to related synchronization jobs according to the affected gateway, transfer agent, API endpoint, or token dependency. Numerical variables, including synchronization frequency, retry count, token privilege breadth, and vulnerability severity, were normalized to a range from 0 to 1. Categorical variables, including data class, encryption state, gateway exposure, and policy violation type, were encoded as binary or ordinal values according to their security meaning. The risk score of synchronization job "i" was calculated as follows (1):

$$R_i=\alpha V_i+\beta D_i+\gamma T_i+\delta E_i+\lambda P_i+\mu G_i \tag{1}$$

where Ri represents the final synchronization risk score. Vi is the normalized vulnerability severity, Di is the data sensitivity score, Ti is the access-token scope risk, Ei is the transfer encryption risk, Pi is the policy violation intensity, and Gi is the gateway exposure score. The weights α, β, γ, δ, λ, and μ were set according to security impact and adjusted through sensitivity analysis. The remediation reduction ratio was calculated as follows (2):

$$RR=Cbefore\text{-}CafterCbefore\times 100 \tag{2}$$

where C_{before} and C_{after} are the numbers of vulnerable synchronization chains before and after remediation, respectively. This metric was used to measure whether the prioritized fixes removed more high-risk data-transfer

chains than the control methods.

Model Evaluation and Validation Procedure

Model performance was evaluated in terms of risk-chain identification, remediation priority, scoring efficiency, and incremental update stability. Risk-chain identification was measured by the number and type of vulnerable synchronization paths detected by each method. Remediation priority was evaluated by applying the top-ranked fixes and comparing the number of removed vulnerable data-transfer chains, reduced exposed ingestion endpoints, and resolved over-scoped token paths. Scoring efficiency was measured by the median scoring time per synchronization job and the total assessment time for the whole environment. Incremental update stability was measured by adding new gateway-policy relationships and recording how many relationships could be processed per minute without recalculating the whole graph. To reduce the effect of a single remediation order, the top-ranked fix list was tested at four ranking windows: 100, 200, 300, and 420 fixes. The results were compared with the three control groups to determine whether the policy-aware lineage model provided clearer remediation value and better scalability for cloud – edge data synchronization security.

RESULTS AND DISCUSSION

Distribution of Synchronization Risk Chains

The policy-aware model detected 1,920 data-synchronization risk chains from 21,300 vulnerability records. These chains were mainly associated with vulnerable transfer agents, weak synchronization channels, over-scoped access tokens, and exposed regional ingestion endpoints. The results indicate that the risk of a cloud – edge synchronization job was not determined by vulnerability severity alone. Some high-severity vulnerabilities showed limited data-security impact because they were located on gateways without sensitive downstream data or write privileges. By contrast, several medium-severity vulnerabilities became high-risk cases when they were connected to confidential sensor data, frequent synchronization jobs, and broad token permissions. This pattern agrees with recent edge-computing security studies, which report that edge systems are exposed to risks from distributed nodes, heterogeneous devices, and multi-layer data transfer (Ma, 2026; Gao et al., 2026). However, most existing studies focus on general edge architecture or device-side protection. The present study further traces the full synchronization path from edge gateways to regional storage and analytics ingestion services. Therefore, the identified risk chains provide a clearer view of how vulnerable components may affect cloud-side data assets, as shown in **Figure 1.**

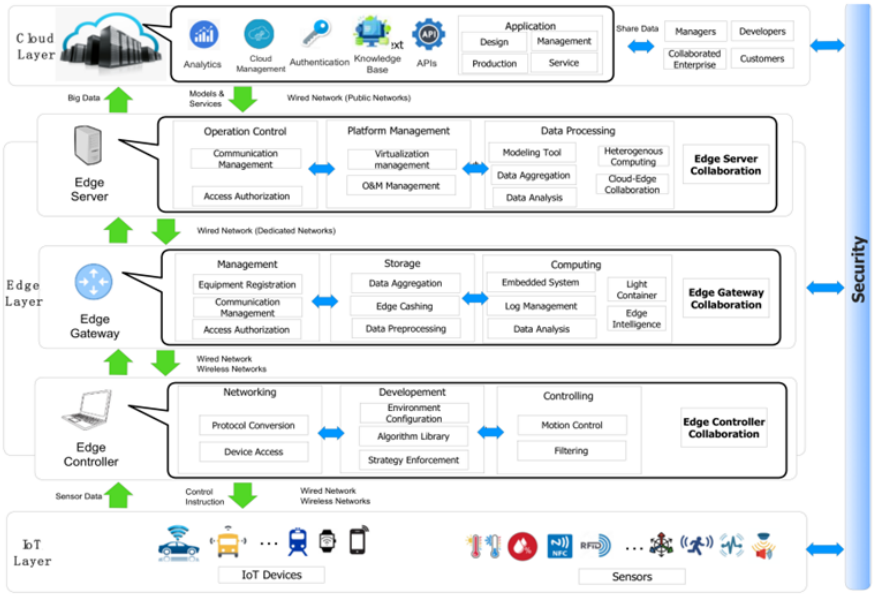

Figure 1. Risk-chain Structure of Cloud – Edge Data Synchronization, Covering Edge Gateways, Transfer Agents, Access Tokens, Regional Storage Endpoints, and Analytics Ingestion Services

Effect of Remediation Unit Grouping

The model grouped 21,300 raw findings into 2,760 synchronization-level remediation units. This step reduced repeated alerts and converted scattered vulnerability records into practical remediation targets. In cloud–edge systems, the same vulnerable transfer agent or gateway image may appear in many synchronization jobs. If each record is handled as an independent task, the remediation list becomes long and difficult to manage. The proposed grouping method addresses this problem by linking findings with the same transfer path, token dependency, policy violation, or downstream endpoint. During remediation testing, the first 420 prioritized fixes removed 980 vulnerable data-transfer chains, accounting for 51.0% of all detected chains. The number of exposed regional ingestion endpoints decreased from 386 to 121, with a reduction of 68.7%. Compared with severity-only ranking, this method avoided early remediation on isolated high-severity findings with low data impact. Compared with asset-exposure scoring, it also avoided over-prioritizing externally reachable gateways that did not process sensitive synchronized data. These findings show that remediation should be organized according to data movement and policy impact, rather than individual vulnerability records alone (Li & Liu, 2026; Riggs et al., 2023).

Scoring Efficiency and Operational Scalability

The median scoring time was 46 ms per synchronization job, and the full assessment of 9,780 synchronization jobs was completed in 14.9 min. This runtime is suitable for periodic security assessment in large cloud–edge environments. The incremental update test showed that the framework processed 14,600 gateway-policy relationships per minute. This performance was achieved because the model recalculated only the affected synchronization subgraph when gateway policies, access tokens, or endpoint settings changed (Zhao et al., 2026; Xu et al., 2026). In common vulnerability management workflows, security teams often rely on weekly or monthly scan reports. Such reports may not reflect rapid changes in cloud access policies or synchronization jobs. The proposed method is more suitable for cloud–edge systems, where synchronization tasks, retry settings, token scopes, and ingestion endpoints may change frequently. Compared with general cloud threat modeling methods, the proposed model also uses a more specific operational unit, namely the synchronization job. This unit allows security teams to update risk scores when data-transfer conditions change, without rebuilding the whole environment model.

Comparison with Existing Studies and Security Meaning

The experimental results show that policy-aware vulnerability scoring offers a different security view from traditional vulnerability ranking, general edge security analysis, and cloud threat modeling. Existing edge security studies often focus on secure architecture, device authentication, communication protection, and intrusion detection. Cloud threat modeling studies usually classify risks across data storage, processing, and transmission phases. These studies are useful, but they do not fully explain how a vulnerable synchronization component, a weak token policy, and a sensitive downstream dataset form a practical risk chain. The present study addresses this gap by using synchronization lineage as the main assessment object. The removal of 980 vulnerable chains after the first 420 fixes shows that data-path-based prioritization can produce clearer remediation benefits than isolated severity scores. The findings also indicate that access-token scope, transfer encryption, retry behavior, and policy status should be assessed together (Tiloca et al., 2025; Fang et al., 2026). A vulnerable gateway may not be urgent when it has no sensitive synchronization path. In contrast, a moderate vulnerability may become critical when it can repeatedly write restricted data to an exposed ingestion endpoint. These results support a path-based security assessment method for cloud–edge data movement, as summarized in **Figure 2.**

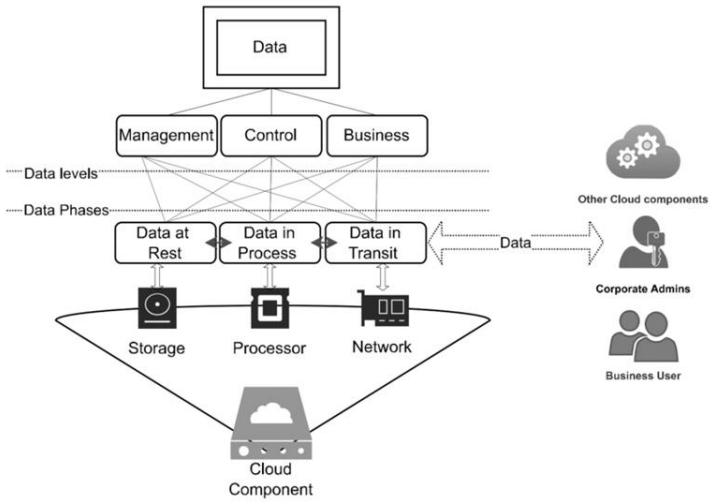

Figure 2. Effect of Policy-Aware Remediation on Vulnerable Synchronization Chains and Exposed Regional Ingestion Endpoints

CONCLUSION

This study developed a policy-aware risk quantification framework for cloud–edge data synchronization pipelines. The framework connects edge gateways, synchronization jobs, transfer agents, access tokens, data classes, regional storage endpoints, and analytics ingestion services through a synchronization-lineage model. The results show that synchronization risk cannot be explained by vulnerability severity alone. When vulnerability status, data sensitivity, transfer protection, token scope, gateway exposure, retry behavior, and policy violations were assessed together, the proposed method identified 1,920 data-synchronization risk chains from 21,300 raw findings and grouped them into 2,760 remediation units. The first 420 prioritized fixes removed 980 vulnerable data-transfer chains and reduced exposed regional ingestion endpoints from 386 to 121. These results indicate that lineage-based prioritization can provide clear remediation value. The median scoring time was 46 ms per synchronization job, and the full assessment was completed in 14.9 min, showing that the method can be applied to large cloud–edge environments. The main contribution of this study is that it uses the synchronization path as the core unit of risk assessment, rather than treating vulnerabilities, assets, and policies as separate objects. This approach can help security teams protect sensitive edge data during transfer and select fixes with higher data-security impact. However, this study also has limitations. The experiment was based on structured cloud–edge records, and the scoring weights were adjusted through sensitivity analysis rather than long-term production use. Future studies should test the framework in real operating environments, include more attack evidence, and refine the weights using actual remediation results.

REFERENCES

Ademilua, D. A. (2021). Cloud security in the era of big data and IoT: A review of emerging risks and protective technologies. *Communication in Physical Sciences, 7*(4), 590-604.

Anderson, B., & Hossain, G. (2026). A socio-technical framework for cyber-resilience in hybrid oil-renewable energy grids. *Cybersecurity, 9*(1), 185.

Chen, H., Ning, P., Li, J., & Mao, Y. (2025). Energy Consumption Analysis and Optimization of Speech Algorithms for Intelligent Terminals.

Deng, X., & Yang, J. (2026). Design of a Quantitative Futures Trading Model Incorporating Multimodal Feature Fusion and Tail Risk Control. Available at SSRN 6702398.

Fang, X., Yang, Y. H., & Wang, S. (2026). Energy-Efficient Context-Aware Multimodal AI Inference at the Edge.

Gao, G., Gao, R., Lu, C., Gao, R., & Kuang, Y. (2026, March). Security Governance Methods and Quantitative Evaluation for Enterprise SMS and Verification Code Systems. In *2026 International Conference on Generative Artificial Intelligence and Information Security (GAIIS)* (pp. 455-458). IEEE.

Gui, H., Wang, B., Lu, Y., & Fu, Y. (2025). Computational Modeling-Based Estimation of Residual Stress and Fatigue Life of Medical Welded Structures.

Latif, S., Djenouri, D., Idrees, Z., Ahmad, J., & Zou, Z. (2025). Hardware security modules for secure communications in the Industrial Internet of Things. *IEEE Communications Surveys & Tutorials*.

Li, Y., & Liu, S. (2026, May). A Study on Dynamic Optimization of Alerting Policies and Multi-Agent Decision-Making Mechanisms in Cloud Environments. In *2026 7th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)* (pp. 703-706). IEEE.

Liang, R., Fan, F., Liang, Y., & Li, S. (2025). Constructing an Adaptive Optimization Model for Ribbon Recommendation and Interface for User Habits.

Liu, H., Xu, D., Ma, Q., Xu, S., & Qiu, D. (2026). Memory Poisoning Propagation and Repair Mechanism in Multi-Agent Collaborative Environments.

Ma, Y. (2026). Industrial Financial Risk Prediction Model Based on Graph Neural Network and Knowledge Graph Inference. *Journal of Circuits, Systems and Computers, 35*(16), 2650103.

Manduva, V. C. (2020). AI-Powered Edge Computing for Environmental Monitoring: A Cloud-Integrated Approach. *International Journal of Emerging Trends in Science and Technology*, 1-21.

Miah, M. R., & Islam, M. M. (2021). Foundational Approaches to Secure Data Collection and Processing in Networked and Distributed Computing Environments. *International Journal of Business and Economics Insights, 1*(4), 32-69.

Paulachan, J., Hekal, O., Onwuchekwa, D., & Obermaisser, R. (2025). A Time-Triggered Edge–Fog–Cloud Architecture With Hierarchical Genetic Optimization for Safety-Critical Applications. IEEE Access, 13, 217406-217422.

Poornimadarshini, S. (2025). Graph-Driven Federated Identity Threat Detection for Trust and Compliance in Secure Hybrid Cloud Networks. *Transactions on Secure Communication Networks and Protocol Engineering, 2*(1), 34-41.

Qi, C., & Qiao, X. (2026). Using AI to Monitor AI: Automated Operations Through Log-Driven Intelligence. Available at SSRN 6795840.

Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., ... & Sarwat, A. I. (2023). Impact, vulnerabilities, and mitigation strategies for cyber-secure critical infrastructure. *Sensors, 23*(8), 4060.

Shao, W. (2026). Policy-Driven Vulnerability Risk Quantification framework for Large-Scale Cloud Infrastructure Data Security. *arXiv preprint arXiv:2604*.06252.

Souza, D., Iwashima, G., Farias da Costa, V. C., Barbosa, C. E., de Souza, J. M., & Zimbrão, G. (2024). Architectural trends in collaborative computing: approaches in the internet of everything era. *Future Internet, 16*(12), 445.

Su, D., & Wu, Y. (2026). Multilevel Growth and Social Network Modeling for Functional Communication Enhancement in Students with Intellectual Disabilities.

Tiloca, M., Palombini, F., Echeverria, S., & Lewis, G. (2025). Notification of Revoked Access Tokens in the Authentication and Authorization for Constrained Environments (ACE) Framework (No. rfc9770).

Trehan, A., & Pradhan, C. (2024). Automated data lineage tracking in data engineering ecosystems. *International Research Journal of Modernization in Engineering Technology and Science, 6*(12), 3305-3312.

Xu, T., Zhang, J., & Zhu, W. (2026). Fairness-Accuracy Trade-offs and Calibration Mechanisms in Machine Learning-Based Subprime Credit Scoring. Available at SSRN 6893759.

Yang, J. (2026). Stage-Coupled Computational Framework for Stratified Accessibility and Equity Analysis in Community-Based Elderly Care Services.

Yin, J., Huang, Y., & Rao, H. (2026). A Study on the Dynamic Evolution of Learning Behavior and Outcome Prediction in Digital Educational Environments. Available at SSRN 6607139.

Zhang, Z. (2026). A Study on Multimodal Product Understanding and Assembly Guidance Optimization in e-commerce for Complex Consumer Goods. Available at SSRN 6734559.

Zhang, Z., Gao, Y., & Tong, Y. (2026). Semantic-Temporal Graph Learning for Demand Forecasting and Resource Allocation in Public Information Systems. Available at SSRN 6792279.

Zhao, J., Fan, J., & Li, L. (2026). A Study on an Explainable Causal-Enhanced LLM Agent for Predicting the Forming Quality of Automotive Component Materials.

Zhu, W., & Yang, J. (2025). Causal Assessment of Cross-Border Project Risk Governance and Financial Compliance: A Hierarchical Panel and Survival Analysis Approach Based on H Company's Overseas Projects.

Zhuwankinyu, E. K., Mupa, M. N., & Tafirenyika, S. (2025). Graph-based security models for AI-driven data storage: A novel approach to protecting classified documents. *World Journal of Advanced Research and Reviews, 26*(2), 1108-1124.

ETHICAL DECLARATION

Conflict of interest: No declaration required. **Financing:** No reporting required. **Peer review:** Double anonymous peer review.