



Cloud-Scale Data Engineering for Real-Time Semiconductor Testing and AI-Powered Chip Diagnostics

Botlagunta Preethish Nandan

SAP Delivery Analytics, USA

preethishnandanbotlagunta@gmail.com

Publication History: Received: 25.04.2026; Revised: 01.05.2026; Accepted: 03.05.2026; Published: 09.05.2026

ABSTRACT: The semiconductor testing process is a key bottleneck in the chip manufacturing flow. To address this inefficiency, data engineering pipelines have been built in the cloud that unify the ingest of test vectors and rich parametric telemetry for real-time testing and AI-powered diagnosis. New chips are tested using algorithms that can decide whether to release or scrap a chip based on the test results and preceding telemetry. Because clearance decisions are made during test, failed chips can quickly be relabeled or logged without being held in test. Data science pipelines examine failed chips and use the conclusions from these analysis to detect issues in other chips in near real-time. Such AI-powered diagnostics also have the potential to help prevent future recurrences. Acting on these insights can ultimately improve yield and shorten test time for future designs.

Data pipelines are about two orders of magnitude less expensive than the test process. They enable continuous decision-making for testing, clearance, and failure analysis without manual steps and facilitate timely action based on telemetry and failed part diagnostics. Making the testing process more efficient can represent significant cost savings for high-volume products. Ultimately, delivering better and cheaper parts to customers faster improves competitiveness and profitability.

KEYWORDS: Semiconductor testing, semiconductor diagnostics, data engineering, cloudscaled development, cloud-scale data engineering.

I. INTRODUCTION

Semiconductor generation and testing remain costly and time-consuming processes that benefit from cloud-scale data engineering. High-fidelity semiconductor test vectors are computed in a cloud environment and the test operations of chips are distributed to multiple test systems. The semiconductor testing becomes a cloud-scale data generation system, with the ATE data ingested in real time into cloud-based storage and serving systems. In parallel, an AI-based visual diagnostics pipeline is enabled. High-quality visual diagnosis labels are created to train AI-based visual diagnostics models, thereby providing a scalable solution. High-speed cloud ingestion and storage systems handle the large amounts of diagnostic data generated by the physical testing operations. Real-time AI-based visual diagnosis of the chips and non-intrusive test vector noise analysis enable root-cause analysis on the tested chips. The presented cloud-scale data engineering approach and test-vector-management system enable high-fidelity and real-time semiconductor data generation for AI-powered diagnostics. The AI-powered visual-chip-diagnostic system and the cloud-scale semiconductor ATE ingestion are capable of visual analysis and scaling with only automatic test equipment data.

The successful deployment and operation of high-speed data pipelines focused on the real-time generation of high-fidelity cloud-based data for AI applications are showcased. The methodology enabled is relevant to all data-science and AI disciplines for which high-quality, wide-area data-labeling pipelines and infrastructure are critical. Deep learning is an advanced data-mining technology that continuously pushes the limits of SI systems. At the same time, rubber-stamp manufacturing processes, decreases in the size of semiconductor technologies and constant pressure to find faults at each level, increase the yield and reduce the cost of production are forcing data generation at as high speed as the manufacturing process permits. The ultimate prize is finding a solution to the problem of previous black-box development methodologies of artificial intelligence based chips.



Table 1. Overview of the Proposed Semiconductor Testing Framework

Component	Description	Purpose	Technologies/Approaches
Cloud-Scale Data Engineering	Distributed cloud infrastructure for semiconductor testing data	Real-time ingestion, storage, and analytics	Open-source cloud technologies, streaming systems
Semiconductor Test Systems	Automated Test Equipment (ATE) executing test vectors	Validate chip functionality and quality	Real-time chip testing
Data Lake	Centralized storage repository	Store telemetry, diagnostics, and test measurements	Streaming ingestion, scalable storage
AI-Powered Diagnostics	Machine learning models for failure analysis	Detect defects and improve yield	CNNs, deep learning, predictive diagnostics
Test Vector Management	Organizes and distributes test vectors	Improve test coverage and reduce redundancy	CCTV and TVD strategies
Streaming Serving Layer	Real-time data dissemination layer	Continuous monitoring and diagnostics	Cloud streaming architecture

II. BACKGROUND AND MOTIVATION

With a more connected world, there is a growing need for more interconnected devices, appliances, and systems, and semiconductors are the fundamental building blocks of these devices, which has attracted investments of hundreds of billion dollars in research facilities and product manufacturing. After an overall picture of this booming market, this section presents the two focal areas of this work with a more academic perspective: semiconductor testing for validating and certifying products, and Artificial Intelligence, drawing bridges between those two areas. Real-time data engineering processes, transported from the cloud to the edge, seize test signals and guarantee a method to generate the best test samples for advanced kaseq technologies.

An increasing number of products and systems are based on Artificial Intelligence techniques, such as information classification, image processing, anomaly detection, speech recognition, information generation, and many others. Nevertheless, many products are restricted to specific applications, and the potential empowered by these Artificial Intelligence techniques will only be accessible for all products when the development of high-quality information labeling tools and processes finally breaks the existing barriers. One of those barriers is the lack of high-quality labeled information, especially for training models that classify information in abnormal behavior and during testing for edge inference. One of the AI models in this work aims to help build high-quality labeled information repositories for semiconductors, boosting test coverage for these products that necessarily play a central role in enabling the integration of AI techniques in all products and systems.

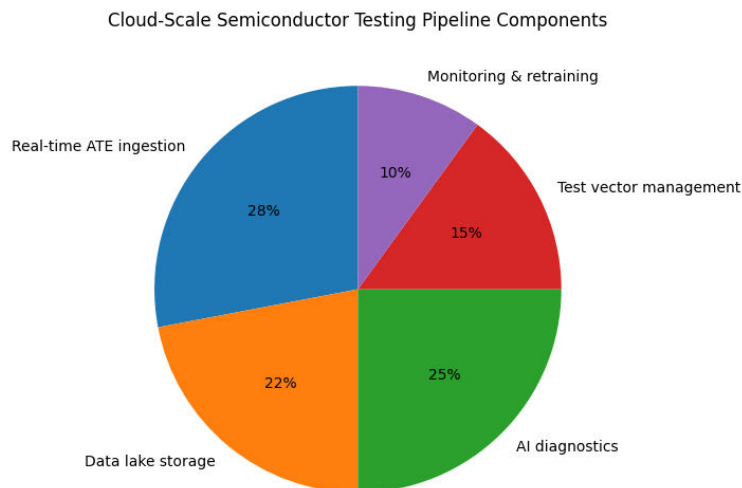




Table 2. Real-Time Data Engineering Pipeline

Pipeline Stage	Input Data	Processing Activity	Output
Data Generation	Semiconductor test vectors and telemetry	Functional testing of chips	Raw test data
Data Ingestion	ATE measurement data	Continuous upload to cloud	Streamed semiconductor data
Data Storage	Streaming data	Storage in cloud data lake	Structured datasets
AI Diagnostics	Historical and live test data	Predictive analysis and classification	Failure diagnostics
Monitoring & Analytics	Diagnostic outputs	Visualization and root-cause analysis	Yield optimization insights
Feedback Loop	AI-generated insights	Test vector improvement	Enhanced future testing

III. SYSTEM ARCHITECTURE FOR CLOUD-SCALE DATA ENGINEERING

A distributed architecture based on open-source cloud-scale data engineering principles establishes a cloud-scale solution stack that ingests, processes, and distributes semiconductor test data in real time. Cloud-scale computing enables data engineering systems that perform the requisite data processing and data movement at something approaching the required rates for industrial applications. Relatively small volumes of data are created and made available within seconds to hopefully help engineers narrow the root cause analysis as fast as possible. For semiconductor devices previously subjected to real-time diagnostic tests, the data is used to power a suite of AI-based diagnostics running in urban legend.

The first component of the architecture is a streaming serving layer designed to facilitate the dissemination of the data up and down the testing process from various levels of testing (wafer, package, board, system) to the diagnostic model developers and back. This layer is complemented by a test vector management system that organizes and processes the test vectors delivered to the tester during production testing. Together with the previous component, they provide the requisite data flow for real-time monitoring of semiconductor testing.



Mathematical Formulas:

1. Signal-to-Noise-and-Distortion Ratio (SINAD)

Used in semiconductor diagnostics and test-quality evaluation.

$$SINAD = 10\log_{10} \left(\frac{P_{signal} + P_{noise} + P_{distortion}}{P_{noise} + P_{distortion}} \right)$$

$$SINAD = 10\log_{10} \left(\frac{P_{signal} + P_{noise} + P_{distortion}}{P_{noise} + P_{distortion}} \right)$$

2. Data Throughput Equation

Represents streaming data ingestion performance in cloud pipelines.

$$Throughput = \frac{Data\ Volume}{Processing\ Time}$$

$$Throughput = \frac{Data\ Volume}{Processing\ Time}$$

3. Chip Yield Formula

Useful for semiconductor manufacturing efficiency.



$$Yield = \frac{\text{Number of Good Chips}}{\text{Total Chips Tested}}$$

$$Yield = \frac{\text{Number of Good Chips}}{\text{Total Chips Tested}}$$

4. Classification Accuracy

Applied in AI-powered chip diagnostics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

5. Precision of Failure Detection

Measures diagnostic correctness.

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

6. Recall for Fault Detection

Useful in semiconductor anomaly detection.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

7. Latency in Real-Time Streaming

$$Latency = \text{Processing Time} - \text{Arrival Time}$$

8. CNN Convolution Operation

Related to AI-based chip image diagnostics.

$$F(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n)$$

9. Mean Squared Error (MSE)

Used during AI model training.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

10. Data Compression Ratio

Useful for cloud-scale storage optimization.

$$Compression Ratio = \frac{\text{Original Size}}{\text{Compressed Size}}$$

11. Failure Probability

$$P(\text{Failure}) = \frac{\text{Failed Chips}}{\text{Total Chips}}$$



12. Test Coverage Equation

$$Coverage = \frac{Detected\ Faults}{Total\ Faults}$$

$$Coverage = \frac{Detected\ Faults}{Total\ Faults}$$

13. AI Model Loss Function

Binary classification for pass/fail chip prediction.

$$Loss = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

14. Streaming Queue Utilization

$$Utilization = \frac{Arrival\ Rate}{Service\ Rate}$$

15. Cloud Storage Scaling

$$Storage_{total} = N \times Storage_{per\ chip}$$

3.1. Data Ingestion and Streaming

The first layer of the cloud-scale data architecture addresses the use case of real-time semiconductor testing. A key requirement is the ingestion of measurement data that is generated during semiconductor tests at a 24-by-7 scale. Each test of a semiconductor device requires the execution of test vectors that are applied to the device terminals, and the test results are returned to the testing system and stored in a SQL storage system. An efficient data upload system is provided to continuously gather test measurement data from these SQL systems at the end of every test into streaming storage of a data lake.

The data-lake serving these test measurements also supports the use case for AI-generated chip test diagnostic data and leverages the streaming ingestion of chip test measurement data from the real-time chip test system.

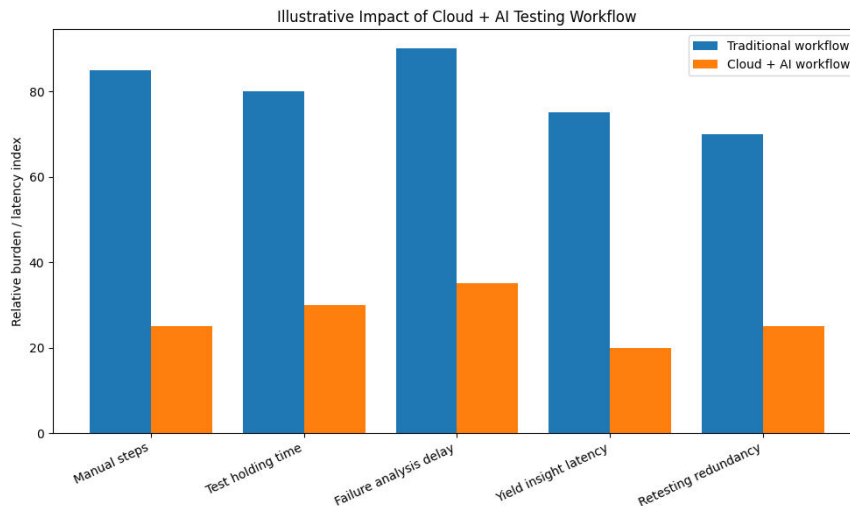


Table 3. Semiconductor Testing Workflow

Stage	Activity	Objective	Result
Test Vector Generation	Create functional test patterns	Stimulate chip functions	Test matrix
Functional Testing	Execute vectors on semiconductor devices	Validate chip behavior	Response measurements
Signal Analysis	Compare outputs with expected values	Detect deviations	SINAD assessment
Data Logging	Store simulation and test logs	Trace chip behavior	Diagnostic datasets

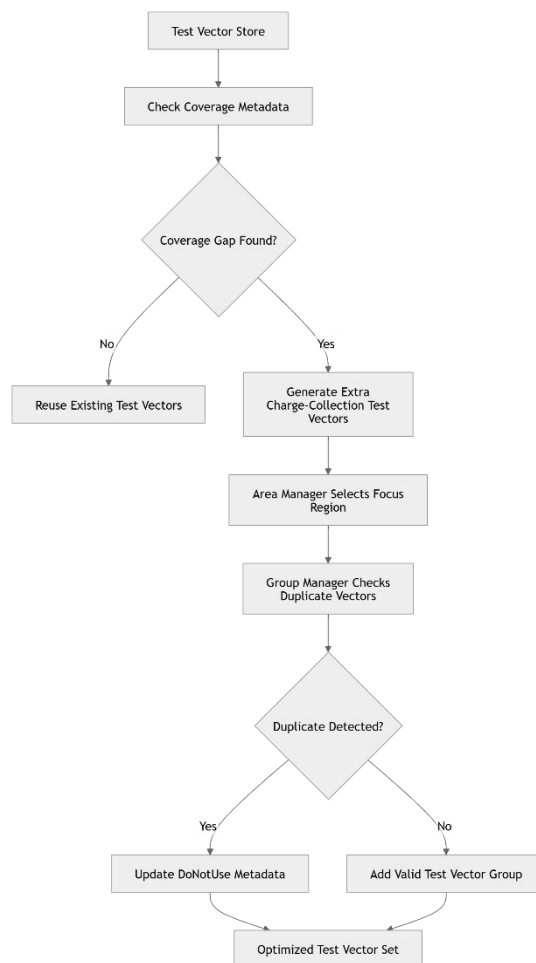


Stage	Activity	Objective	Result
Failure Analysis	Analyze abnormal responses	Identify root causes	Failure classification
AI-Based Recommendations	Predict corrective actions	Improve production quality	Yield enhancement

IV. REAL-TIME TESTING PIPELINES FOR SEMICONDUCTORS

The core of semiconductor testing consists in the assessment of chip responsiveness to thousands of functional patterns, called test vectors. Test vectors characterize the chip functions of interest and are explicitly defined by the chip designer. To disseminate these vectors, a dedicated Test Pattern Generator (TPG) is implemented. The test matrix produced by the TPG serves as input for various semiconductor testers or channels. Each test vector stimulus activates the chip, and the chip response is measured and compared to the expected output.

In parallel, an industry-standard register transfer-level simulation set-up provides logging coverage of all stimulated chip functions, offering a baseline to determine the quality of chip test results as well as the reasons for any eventual failure. The test matrix from the TPG, combined with the simulation log coverage and the results from all testing channels, enables the generation of a Signal-to-Noise-and-Distortion Ratio (SINAD) assessment for the chip functions, which is essential for diagnosis.

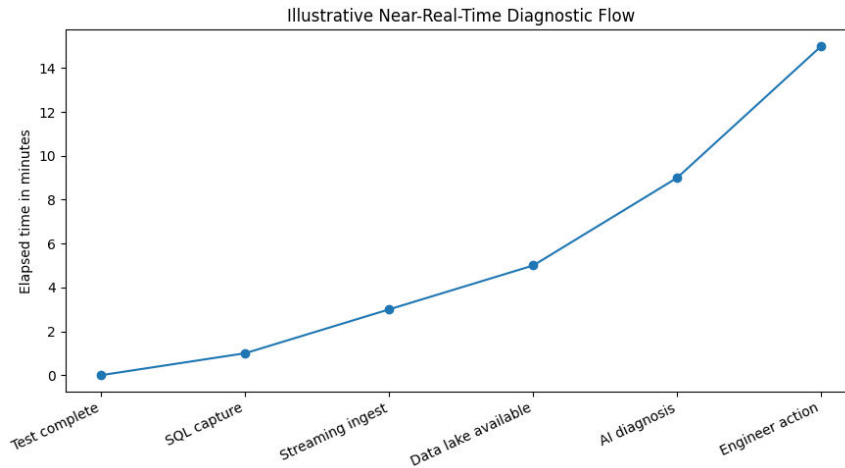


4.1. Test Vector Management

Test vectors are the input patterns used to stimulate a silicon chip when testing it after fabrication and are cached and reused in final test. Charge-Collection Test Vectors (CCTV) are specifically for detecting charge-collection defects caused by radioactive particles. They check if charge injected into the chip is trapped in the oxide near the MOSFET



channel and can leak through the oxide. These defects tend to only trigger a handful of test vectors, so a careful selection and categorization process is needed for future tests. Other test vectors are for detecting gaps in the testing space. Typically, for CTVs, if a small part of the chip has already been tested with a CCTV, the DoNotUse metadata is updated to prevent redundancy.



The extra TVD management uses groups of CTVs and is triggered when the TX store is populated enough. Four stages help decide whether to generate extra CTVs, determine which area to focus on, how many areas to cover, and the distribution of the test vectors based on missing coverage. This management is adapted for the extra TVD generator. When a duplicate is detected, the group manager update the valid DoNotUse area and moves to the group with the same area received from the area manager.

Table 4. Test Vector Management Strategies

Strategy	Description	Benefit
Charge-Collection Test Vectors (CCTV)	Detect charge-collection defects caused by radioactive particles	Improve defect identification
DoNotUse Metadata	Prevent repeated testing in already-covered areas	Reduce redundancy
Extra TVD Generation	Generate additional vectors for uncovered regions	Improve test coverage
Group-Based Vector Management	Organize vectors into coverage groups	Efficient vector allocation
Duplicate Detection Mechanism	Detect overlapping vector coverage	Optimize resource utilization

V. AI-POWERED CHIP DIAGNOSTICS

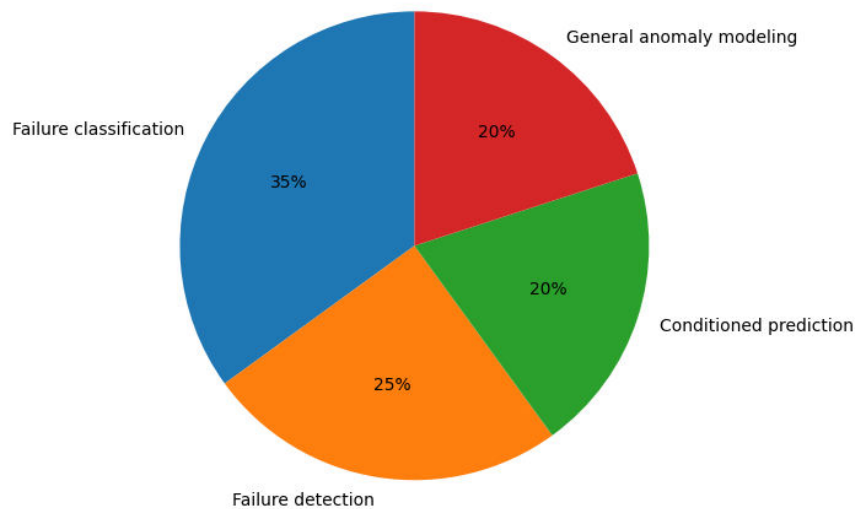
As semiconductor manufacturing is becoming increasingly complex, so are the test-based failures. Traditionally, the analysis of these failure modes was carried out manually by engineers and specialists. However, the growing number of chips to test—with the associated volume of potential test-based failures—has prompted the industry to look into solutions that improve the cadence, automation, and cost efficiency of the analysis. To address this challenge, AI has been adopted as a solution to fill the gap of human expertise. One example from the automotive industry is a Deep Learning-based failure classification model provided with redesign recommendations. Such applications necessitate the consumption of not only test-side information but also traditional design-for-test information. These relationships, however, are not always straightforward.

Four different AI problems can be identified based on the designer’s goal concerning the classification and the provided input. Failures can either be classified or detected, while the model can be conditioned either with an input feature that strongly helps to predict the failure or in the most general form. There are two main reasons that explain why traditional supervised classification is not sufficient. First, a chip can be affected by too many failure modes. Second, semiconductor manufactures follow a test-based strategy, which often leads to unbalanced data—presence of a failure in a limited



number of tested samples compared to its total amount. The proposed AI-powered chip diagnostics accounts for these issues and consists of a combination of multiple AI models.

AI Diagnostic Problem Types Discussed in the Paper



5.1. Diagnostic Models and Training Regimes

Reliable chip diagnostics not only help manufacturers identify the chip failure reasons but also aid design engineers to figure out possible design flaws. Unfortunately, there are no adequate solutions publicly available to facilitate scalable chip diagnostics modeling. Given the increasing complexity of modern chips, traditional diagnostic modeling approaches such as manually designed rules or expert systems may become impractical. Consequently, various state-of-the-art deep learning-based techniques for failing chip diagnostic modeling have been developed but generally offer poor performance. These techniques share several limitations. First, they use a multi-layer perception as a classifier, which is commonly believed to be less effective than convolutional neural networks (CNNs) when modeling classification tasks. Second, they are trained either in a single-stage or in a two-stage manner during the whole chip-level classification process. However, the two-stage training configuration introduces layer-wise category imbalance in the first-stage model, which may further introduce suboptimal first-stage models. Last but not least, these techniques lack a diagnostic failure-path analysis which helps designers to identify potential circuit design errors.

To address these limitations, researchers propose a novel diagnostic modeling framework for failing chip classification named ChipDiags, which comprises two stages: a diagnostic model and a following diagnostic sub-category analysis. A CNN-based deep learning configuration is adopted in the diagnostic modeling. The first stage investigates the chip diagnostic model training via an end-to-end level-failing-chip classification scheme, and the second stage develops a sub-category analysis system that is capable of shining a light along the failure paths of malfunctioned chips and identifying the abnormal circuit areas contributing most to the chip failure.

Table 5. AI-Powered Chip Diagnostics Models

AI Model Type	Function	Input Data	Expected Outcome
Failure Classification Model	Categorize chip failures	Test telemetry and logs	Failure categories
Failure Detection Model	Detect abnormal chip behavior	Real-time test data	Fault alerts
CNN-Based Diagnostic Model	Deep learning-based defect analysis	Chip diagnostic datasets	Improved classification accuracy
Predictive Diagnostics Model	Predict future chip failures	Historical testing patterns	Preventive maintenance



AI Model Type	Function	Input Data	Expected Outcome
Sub-Category Analysis System	Analyze failure paths	Diagnostic outputs	Root-cause identification

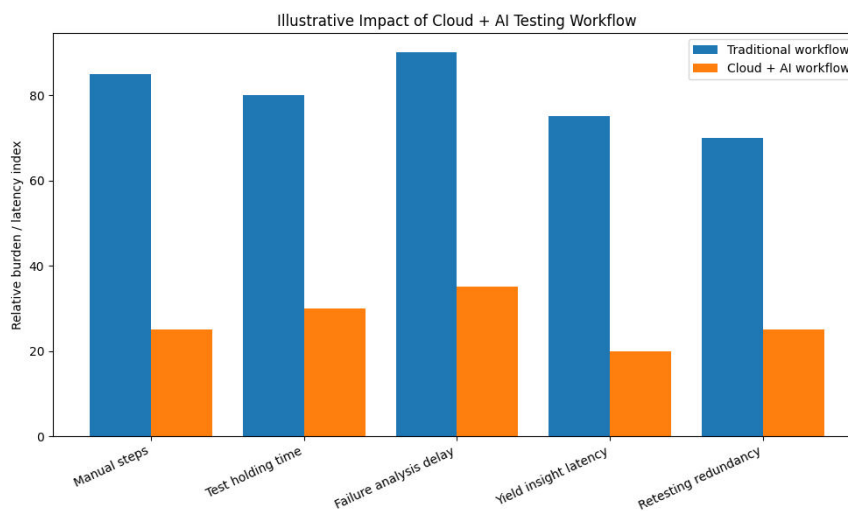
VI. CONCLUSION

The presented work illustrates a robust, scalable system architecture for the cloud-scale data engineering of semiconductor testing and diagnostics. Real-time consistency and fault tolerance requirements are met through a combination of open-source tools for data ingestion, quality control, storage, and analytics. Data is streamed in real time from semiconductor test stations during wafer manufacturing. Test vector quality is monitored and outliers highlighted for analyst review. The diagnostics subsystem deploys an advanced set of AI-powered machine learning models that reduce lab testing time in production by several orders of magnitude. These models are designed to withstand concept drift, with automatic retraining triggered by data quality monitoring.

Data engineering pipelines that serve as the foundations of semiconductor artificial intelligence are also described. New test data are ingested into the data lake on a per-test-case basis as they are generated during manufacturing. The test-condition space is monitored for gaps and the quality of surrounding test data is shown to affect AI model accuracy.

Table 6. Comparison Between Traditional and AI-Based Diagnostics

Feature	Traditional Diagnostics	AI-Powered Diagnostics
Analysis Method	Manual engineering analysis	Automated machine learning
Scalability	Limited	Highly scalable
Processing Speed	Slow	Real-time or near real-time
Failure Detection	Rule-based	Data-driven and adaptive
Handling Large Data Volumes	Difficult	Efficient
Predictive Capability	Limited	Strong predictive analytics
Cost Efficiency	High operational cost	Reduced operational cost



REFERENCES

1. Ranga Reddy, V. A. (2024). Comparing Batch vs. Streaming Approaches in Healthcare Data Warehousing Environments. *Journal of Neonatal Surgery*, 13(1), 2287–2309. Retrieved from <https://www.jneonatsurg.com/index.php/jns/article/view/10223>



2. Ranjith Kumar Peddi. (2024). AI-Based Workforce Analytics for SLA Governance and Uptime Assurance in Data Centers. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(08), 8589–8601. Retrieved from <https://eudoxuspress.com/index.php/pub/article/view/5361>
3. Loganathan, R. (2024). GENERATIVE AI-ENABLED COMPLIANCE DOCUMENTATION AND AUDIT TRAIL AUTOMATION FOR GLOBAL DATA CENTER GOVERNANCE. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 15(3), 487–504. <https://doi.org/10.61841/turcomat.v15i3.15512>
4. Mangala, N. (2026). Responsible AI Data Architecture: Embedding GDPR and PII Compliance into MLOps Pipelines at Enterprise Scale. *Canadian Journal of Marketing Research*, 16(1), 107-124.
5. Mangalampalli, B. M. (2024). AI-Enhanced Data Governance: Automating Compliance In Healthcare Analytics Platforms. *The Review of Diabetic Studies*, 191-204.
6. Kolla, T. (2025). The Future of Healthcare Analytics: Leveraging AI and Data Engineering for Personalized Medicine. *Journal of Computer Science and Technology Studies*, 7(4), 634-640.
7. Yandamuri, U. S. (2026). AI-Enabled Workflow Automation and Predictive Analytics for Enterprise Operations Management. *Management*, 3(1), 15-24.
8. Nagubandi, A. R. (2025). Cryptocurrency Market Spillovers: Risk Contagion Across Global Financial Systems.
9. Kolla, S. K. (2026). Foundation Deep Learning Models For Precision Medicine Using Multimodal Big Data. *INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES*.
10. Bandi, V. D. V. K. Autonomous Data Platforms: Converging AI, MLOps, and Cloud Engineering for Digital.
11. Davuluri, P. N. Autonomous Compliance Systems: AI, Event Streaming, and the Future of Financial Crime Prevention.
12. Pote¹, X. R., Pamisetty, A., Karthikeyan, G., & Gupta¹, D. (2025, May). Artificial Intelligence Enabled Smart Energy Conservation Systems for Intelligent Resource Management and Sustainable Future Power Grids. In *Proceedings of the International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 24)* (p. 196). Springer Nature.
13. Enterprise-Scale Gen AI Orchestration Using Small LMs and LLM Agents for Intelligent ITSM and HRSD Automation in Enterprise Ecosystems. (2025). *MSW Management Journal*, 35(2), 1889-1897.
14. Shah, M. M., & Kolla, S. H. (2026). Harvest Net: An AI-Powered Adaptive System for Yield Prediction and Resource Optimization in Agriculture. *Canadian Journal of Marketing Research*, 16(2), 181-196.
15. Amistapuram, K. (2025). GENERATIVE AI FOR CLAIMS EXCEPTIONS AND INVESTIGATIONS: ENHANCING RESOLUTION EFFICIENCY IN COMPLEX INSURANCE PROCESSES. Available at SSRN 5785482.
16. Ramana, B., Sheelam, G. K., Pandya, T., Rai, A. K., Kumar, V. A., & Kukreti, A. (2025). Exploring the Potential of NOMA in 6G Through Comparative Analysis with OMA Techniques. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1–6). IEEE. 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG). <https://doi.org/10.1109/ictbig68706.2025.11323270>
17. Chary, D. V., Meda, R., C, J. S. Mary., Narasimhachari, J. P., & A S, Y. (2025). TriFusionFormer: Tri-Modal Fusion Transformer Using Gated Modality Control and Multi-Scale Attention for Emotion Recognition. In *2025 International Conference on Communication, Computer, and Information Technology (IC3IT)* (pp. 1–8). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). <https://doi.org/10.1109/ic3it66137.2025.11341646>
18. Paleti, S., Kummari, D. N., Garapati, R. S., Sheelam, G. K., Adusupalli, B., & Singireddy, J. (2025, December). Building a Cyber-Resilient Payment Infrastructure: Transforming Payment Security with Zero Trust Architecture. In *2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT)* (pp. 1-7). IEEE.
19. Alshar, M. M., Shahdadhuri, N., Rajeshwari, M., Gupta, M., Joshi, N. R., & Singireddy, J. (2025). Enhanced Management & Performance of Remote Workforce with Cloud and AI-Driven HR Analytics. In *2025 3rd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)* (pp. 631–636). IEEE. 2025 3rd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT). <https://doi.org/10.1109/icaiccit68829.2025.11434104>
20. Annareddy, V. N., Singireddy, J., Preethish Nandan, B., Lakarasu, P., & Burugulla, J. K. R. (2025). Emotional intelligence in artificial agents: Leveraging deep multimodal big data for contextual social interaction and adaptive behavioral modelling. Available at SSRN 5241039.
21. Nuka, S. T., Chakilam, C., Chava, K., Suura, S. R., & Recharla, M. (2025). AI-driven drug discovery: transforming neurological and neurodegenerative disease treatment through bioinformatics and genomic research. *American Journal of Psychiatric Rehabilitation*, 28(1), 124-135.
22. Pandiri, L. (2025). *The Complete Compendium of Digital Insurance Solutions: Life, Health, Auto, Property, and Specialized Coverage in the Age of AI, Automation, and Intelligent Risk Management*. Deep Science Publishing.



23. Krishnaprasath, V. T., Pamisetty, V., Sharma, V., Nayak, M., Baalakumar, N. N., & Aravindh, S. (2025, May). Federated learning based artificial intelligence systems with blockchain security for global healthcare collaboration and patient centric data privacy. In *International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024)* (pp. 1277-1290). Atlantis Press.
24. Chakraborty, S., Pamisetty, A., Chandana, N., & CS, B. (2025, October). Depth-Wise Temporal Convolutional Networks with Layer Normalization for Waste Food Prediction. In *2025 2nd International Conference on Software, Systems and Information Technology (SSITCON)* (pp. 1-6). IEEE.
25. Gottimukkala, V. R. R. (2025). Agentic AI for Next-Generation Cross-Border Payments: Contextual Learning in Transaction Routing. *Journal of Informatics Education and Research*, 5(4).
26. Gupta, D. K., Purushotham, K., Dheer, G., P, S., Gottimukkala, V. R. R., & Kapoor, S. (2025). Semantic Feature Learning Using Transformer-Based Deep Neural Networks. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1–6). IEEE. 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG). <https://doi.org/10.1109/ictbig68706.2025.11323734>
27. Pallapu, S. R., Aitha, A. R., Vandhana, K., & Chelladurai, S. (2025, October). GAN-Augmented Transformer Framework for Cross-Domain Video Style Transfer. In *2025 International Conference on Communication, Computer, and Information Technology (IC3IT)* (pp. 1-6). IEEE.
28. Bhasgi, S. S., Garapati, R. S., B, Ayshwarya., Sasikala, M., & J, Srinivasan. (2025). Medical Image Fusion of Magnetic Resonance Imaging and Computed Tomography Using Learned Wavelet Complex Adapter. In *2025 International Conference on Communication, Computer, and Information Technology (IC3IT)* (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). <https://doi.org/10.1109/ic3it66137.2025.11340892>
29. Pamisetty, A., Paleti, S., Adusupalli, B., Singireddy, J., Inala, R., & Nagabhyru, K. C. (2025, September). Explainable AI Systems for Credit Scoring and Loan Risk Assessment in Digital Banking Platforms. In *2025 IEEE 13th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (pp. 1478-1483). IEEE.
30. Ranga Reddy, V. A. (2024). Comparing Batch vs. Streaming Approaches in Healthcare Data Warehousing Environments. *Journal of Neonatal Surgery*, 13(1), 2287–2309. Retrieved from <https://www.jneonatsurg.com/index.php/jns/article/view/10223>
31. AGENTIC AI FRAMEWORKS FOR AUTONOMOUS RISK DETECTION AND COMPLIANCE REMEDIATION IN ENTERPRISE DATA CENTER OPERATIONS. (2025). *Lex Localis - Journal of Local Self-Government*, 23(S6), 9672-9697. <https://doi.org/10.52152/3f90ak91>
32. Mangala, N. (2026). A Unified Architecture for Real-Time Analytics Using Microsoft Fabric OneLake. *International Journal of Human Computations and Intelligence*, 5(3), 793-807.
33. Bandi, V. D. V. K. (2026). Cognitive Data Engineering: AI-Governed Data Quality, Lineage, and Pipeline Optimization at Scale. *International Journal of Economic Practices and Theories*, 2026, 131-148.
34. Kolla, S. H. (2026). Autonomous Enterprise Agents: Orchestrating Large and Small Language Models for Scalable Decision Automation in ITSM, HRSD, and CSM Platforms. *INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES*, 24-45.
35. Yandamuri, U. S. AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology.
36. Kolla, S. H. (2026). Small Language Models as Control Planes: Designing Cost-Efficient GenAI Orchestration Layers for Enterprise-Integrated Digital Workflows. *Minnesota Journal of Business Law and Entrepreneurship*.
37. Radha, S., Gottimukkala, V. R. R., Thottara, S., Vandhana, K., & J, Gokulraj. (2025). Adaptive Video Streaming Over 5G Networks Using Deep Reinforcement Learning with Closed-Loop Feedback Mechanism for Bitrate Control. In *2025 International Conference on Communication, Computer, and Information Technology (IC3IT)* (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). <https://doi.org/10.1109/ic3it66137.2025.11341184>
38. Nagabhyru, K. C., Gadi, A. L., Seenu, A., Davuluri, P. S. L. N., Segireddy, A. R., & Pamisetty, V. (2026). Towards Automated Financial Risk Scoring in Automotive Financing with Explainable Machine Learning. In *2026 IEEE International Conference on AI Engineering and Innovations (AIEI)* (pp. 1–6). IEEE. 2026 IEEE International Conference on AI Engineering and Innovations (AIEI). <https://doi.org/10.1109/aiei69164.2026.11496822>
39. EdgeMind: A Self-Evolving AI Framework for Distributed Intelligence in IoT Ecosystems. (2026). *Journal of Informatics Education and Research*, 6(2). <https://jier.org/index.php/journal/article/view/4609>
40. Garapati, R. S., Adusupalli, B., Kaulwar, P. K., Gadi, A. L., Annapareddy, V. N., & Challa, K. (2025). The Evolution of Digital Payments: A Study on AI-Powered Transaction Monitoring Systems. In *2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT)* (pp. 1–8). IEEE. 2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT). <https://doi.org/10.1109/icicat68430.2025.11414665>



41. Devayani, G., & Nagabhyru, K. C. (2026). Wireless Sensor Networks and Digital Twins for Real-Time City Simulation. Available at SSRN 6094546.
42. Sivanand, R., Kumar, D. P., Nagabhyru, K. C., Natarajan, E. P., Pamisetty, V., & Kapila, D. (2025, September). IoT and AI for Real-Time Monitoring in Substation Automation. In 2025 International Conference on Computing and Communications (COMPUTINGCON) (pp. 1-5). IEEE.
43. P, R., Manoranjithem, V., Garapati, R. S., Singh, S., Praveen, R., & K, M. S. (2025). Random Forest–XGBoost Hybrid Model for Early Detection of Breast Cancer in Medical Imaging Datasets. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–6). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). <https://doi.org/10.1109/gwcwn66157.2025.11448354>
44. Kolla, S. H., Inala, R., & Kumar, M. V. K. (2026). Secure RAG Architectures with Small Language Models for Governance-Aligned LLM Deployment in Enterprise Service Management Platforms. *International Journal of Economic Practices and Theories*, 2026, 166-179.
45. Yerra, S. D., Kiran Kumar, D. Y., Sheelam, G. K., Praveen, R., Paul, P. M., & M, D. (2026). Enhancing Road Safety and Network Intelligence Using a Swarm Intelligence–SVM Hybrid Model in 6G-Enabled V2X Communication. In 2026 IEEE International Conference on AI Engineering and Innovations (AIEI) (pp. 1–6). IEEE. 2026 IEEE International Conference on AI Engineering and Innovations (AIEI). <https://doi.org/10.1109/aiei69164.2026.11497283>
46. FinOps Strategies for AI-Enabled Real-Time Compliance Platforms in Cloud Native Environments. (2025). *MSW Management Journal*, 35(2), 2080-2088.
47. MANGALAMPALLI, B. M., KOLLA, S. H., APPA RAO NAGUBANDI, D. R., & SEGIREDDY, A. R. (2025). AN INTELLIGENT, REAL-TIME DIGITAL FABRIC FOR HEALTHCARE AND FINANCIAL ECOSYSTEMS USING AUTONOMOUS LEARNING AND GENERATIVE SYSTEMS. *TPM–Testing, Psychometrics, Methodology in Applied Psychology*, 32(S9 (2025): Posted 15 December), 3070-3086.
48. Yandamuri, U. S. (2026). Scalable Cloud-Based Intelligent Decision Systems Leveraging AI and Big Data for Industry-Specific Optimization. *Minnesota Journal of Business Law and Entrepreneurship*, (1), 584-601.
49. Amistapuram, K. (2025). Agentic AI for Next-Generation Insurance Platforms: Autonomous Decision-Making in Claims and Policy Servicing. *Journal of Marketing & Social Research*, 2, 88-103.
50. Sheelam, G. K. (2025). Agentic AI in 6G: Revolutionizing Intelligent Wireless Systems through Advanced Semiconductor Technologies. *Advances in Consumer Research*.
51. Mangalampalli, B. M., & Kolla, T. (2026). FHIR-Based Interoperability Frameworks For Real-Time Healthcare Data Exchange: Architecture Patterns And Performance Optimization. *International Journal Of Advances in Signal and Image Sciences*, 1514-1536.
52. Bargavi, N., Athawale, S. G., Amistapuram, K., & Aitha, A. R. (2026). Safeguarding Consumer Data in Digital Insurance: Legal Frameworks and Ethical Imperatives. *International Insurance Law Review*, 34(S1), 272-284.
53. Deepika, G., Recharla, M., Deepika, S., P, Ilanchezhian., & G, Nirupashri. (2025). Adaptive Lightweight Autoencoder with Noise Estimation Module for Noise Reduction in ECG Signals. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). <https://doi.org/10.1109/ic3it66137.2025.11340876>
54. Madhavi, K. R., Gottimukkala, V. R. R., Pandiri, L., Sriram, H. K., Malempati, M., & Adusupalli, B. (2025, November). Hybrid Transformer–Federated Learning Model for Secure Release Engineering in Global Payment Networks. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1-6). IEEE.
55. Challa, K., Singireddy, J., Pamisetty, A., Garapati, R. S., Kannan, S., & Sriram, H. K. (2025, December). Harnessing Agentic AI and IT Infrastructure in Banking to Drive Consumer Insights, Operational Excellence, and Intelligent Financial Innovation. In 2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT) (pp. 1-7). IEEE.
56. Ranjith Kumar Peddi (2021). Optimizing Case Management Workflows in Global Data Center Colocation Services. *Universal Journal of Computer Sciences and Communications*, 1(1), 1-21. <https://doi.org/10.31586/ujscs.2021.1380>
57. Loganathan, R. (2022). Converging Security Architecture and Compliance Management in Enterprise Data Center Ecosystems: A Unified Control Framework. *International Journal of Scientific Research and Modern Technology*, 1(12), 295–312. <https://doi.org/10.38124/ijrsmt.v1i12.1378>
58. Mangalampalli, B. M. Generative AI Applications In Healthcare Data Mart Design And Optimization.
59. None, D. M. K., None, V. D. V. K. B., None, N. M., None, S. H. K. & None, B. M. M. (2026). Engineering Intelligent Cloud-Native Data Ecosystems for Predictive Decision-Making in Industry. *Journal of European Economic History*, 7(2), 68-88.
60. Jagtap, S., Inala, R., Venu, M., & Divya, T. V. (2025, October). Large-Scale Crowd Flow Prediction Using Temporal Convolutional Network with Spatio-Temporal Attention. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1-6). IEEE.



61. Rathor, K., Meda, R., Agnihotri, K., Sinha, P. K., Mandal, P., & Gulati, M. (2025). Detecting and Interpreting Financial Statement Fraud via Supply Chain-Based Graph Neural Network Models. In 2025 IEEE 4th International Conference for Advancement in Technology (ICONAT) (pp. 1–5). IEEE. 2025 IEEE 4th International Conference for Advancement in Technology (ICONAT). <https://doi.org/10.1109/iconat66879.2025.11362543>
62. Krishnan, M., Aitha, A. R., Amistapuram, K., Nandan, B. P., Kaulwar, P. K., & Singireddy, J. (2025). Human-in-the-Loop Hybrid Neuro-Symbolic AI Model for Reliable Data Engineering in High-Stakes Industrial Systems. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–7). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). <https://doi.org/10.1109/gcwcnc66157.2025.11448516>
63. Recharla, M., & Nuka, S. T. (2025). Translational Approaches To Commercializing Neurodegenerative Therapies: Bridging Laboratory Research With Clinical Practice. *South Eastern European Journal of Public Health*, 121–144.
64. Sudha Rani, P. R., Amistapuram, K., Pamisetty, V., Singireddy, S., Kummari, D. N., & Sheelam, G. K. (2025). Hybrid Knowledge Graph–Deep Learning Framework for Automated Exception Handling and Investigation in Complex Insurance Claims. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–6). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). <https://doi.org/10.1109/gcwcnc66157.2025.11448301>
65. Rani, P. S., Amistapuram, K., Pamisetty, V., Singireddy, S., Kummari, D. N., & Sheelam, G. K. (2025, November). Hybrid Knowledge Graph–Deep Learning Framework for Automated Exception Handling and Investigation in Complex Insurance Claims. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–6). IEEE.
66. Pandiri, L. (2025, May). Exploring Cross-Sector Innovation in Intelligent Transport Systems, Digitally Enabled Housing Finance, and Tech-Driven Risk Solutions A Multidisciplinary Approach to Sustainable Infrastructure, Urban Equity, and Financial Resilience. In 2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1–12). IEEE.
67. Naik, A. V., Sheelam, G. K., Panchakatl, N., Muthukumar, K., & Saranya, K. (2025). Comprehensive Analysis on Depression Detection From Social Media Using Deep Learning and Transformer Architectures. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–8). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). <https://doi.org/10.1109/ic3it66137.2025.11341160>
68. Madhavi, K. R., Rongali, S. K., Polineni, T. N. S., Kummari, D. N., Challa, K., & Challa, S. R. (2026). Explainable AI (XAI)-Driven Predictive Analytics Framework for Ethical and Scalable Automation in Cloud-Native Architectures with Enterprise and Healthcare Interoperability. In 2026 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 31–36). IEEE. 2026 International Conference on Electronics and Renewable Systems (ICEARS). <https://doi.org/10.1109/icears67481.2026.11416589>
69. Seenu, A., Aitha, A. R., Gottimukkala, V. R. R., Singireddy, J., Meda, R., & Garapati, R. S. (2025). Hybrid Multi-Agent Reinforcement Learning and Blockchain Framework for Real-Time Transaction Integrity in Cloud-Driven Financial Systems. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–6). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). <https://doi.org/10.1109/gcwcnc66157.2025.11448456>
70. Nandan, B. P., Kumar, M. V. K., Garapati, R. S., Bandi, V. D. V. K., Davuluri, P. S. L. N., & Mangalampalli, B. M. (2026). AI-Enhanced Semiconductor Yield Optimization Using Hybrid Deep Learning and Edge Data Analytics. In 2026 IEEE International Conference on AI Engineering and Innovations (AIEI) (pp. 1–6). IEEE. 2026 IEEE International Conference on AI Engineering and Innovations (AIEI). <https://doi.org/10.1109/aiei69164.2026.11497190>
71. Mangala, N. (2026). Beyond Medallion: Next-Generation Lakehouse Architectures for Real-Time AI-Driven Enterprise Decision Systems. *Minnesota Journal of Business Law and Entrepreneurship*, (1), 1109–1127.