Cloud-Orchestrated Edge AI for Autonomous Quality Management in Smart Manufacturing

Shashikala Valiki

Independent Researcher, India

ABSTRACT: The rapid evolution of smart manufacturing demands intelligent, real-time quality management systems capable of operating with minimal human intervention. This study presents a cloud-orchestrated edge artificial intelligence (AI) framework designed to enable autonomous quality management in smart manufacturing environments. The proposed architecture integrates edge-based AI models for low-latency defect detection, anomaly classification, and process monitoring with cloud-based orchestration for centralized coordination, model lifecycle management, and large-scale data analytics.

By deploying inference tasks at the edge, the system ensures real-time decision-making and reduced network dependency, while the cloud layer facilitates continuous model training, performance optimization, and cross-site knowledge sharing. The framework supports scalable deployment across distributed production units, enhances data security through localized processing, and enables adaptive quality control through dynamic model updates.

KEYWORDS: Smart Manufacturing Systems, Autonomous Quality Control, Intelligent Quality Inspection, Edge Artificial Intelligence, Cloud-Based AI Orchestration, Edge–Cloud Integration, Real-Time Anomaly Detection, Predictive Quality Analytics, Industrial Internet of Things (IioT)

I. INTRODUCTION

While the digitisation of the manufacturing sector holds great promise, guaranteeing quality and performance in complex-smart manufacturing scenarios remains challenging. Although intelligent inspection systems seek fault-free production, they heavily rely on external cloud for both detection and diagnosis. Quality problems thus still jeopardize productivity and profits. An innovative autonomous quality-control mechanism leveraging edge-cloud collaborative artificial intelligence to realise real-time data analysis and decision-making addresses these concerns. Experimental results on real wood materials and a numerical benchmark of logistic regression with remote support illustrate the approach. Autonomous quality control puts data lifecycle management at the heart of the solution. Proper data storage, governance and privacy also support user privacy, data protection and compliance with industry regulations.

The proposed architecture enables local detection of anomalies during production inspections, with distances-to-classification-boundaries from edge-embedded artificial-intelligence models monitored continuously to predict remaining-quality levels, identify potential faults and trigger predictive-maintenance activities in advance. These capabilities help the operator keep production on the right track from the perception perspective, thus preventing future quality degradation. The deployed-edge artificial-intelligence model also monitors the trend of the most recent analysed batches in anticipation of quality drop. One instance of deteriorating product quality prediction, detected by contrasting the trend with the normality region, is illustrated in the results.

1.1. Overview of the Study

Remote quality control leveraging edge computing and AI is an active area of research, but combining quality monitoring, predictive analytics, anomaly detection, and autonomous decision-making using a cloud–edge model is a new approach that facilitates rapid deployment. The field remains congested, however. Manufacturing companies that aim to rapidly develop reliable innovative solutions with edge–cloud features require capabilities that can process information remotely while being tightly governed in the cloud. The quest to shape these solutions led to the design of an autonomous quality-control architecture. From a particular project, an edge–cloud hierarchical framework that orchestrates real-time data acquisition and preprocessing, edge anomaly detection systems and cloud predictive analytics is proposed.

Autonomous quality-control mechanisms are built into the framework as an integral part of the architecture. A cloud system takes care of decisions about quality-control policy. Information on several quality characteristics of products undergoes analysis in real time during manufacturing. Supported by various technical advances, remote quality

monitoring and control are intensively studied topics. Quality-related data pertaining to manufacturing processes, products, and measurements from in-line sensors and signals that emerge in a process containing an edge–cloud structure can be exploited not just for quality monitoring but also for predictive analytics and real-time decision support.
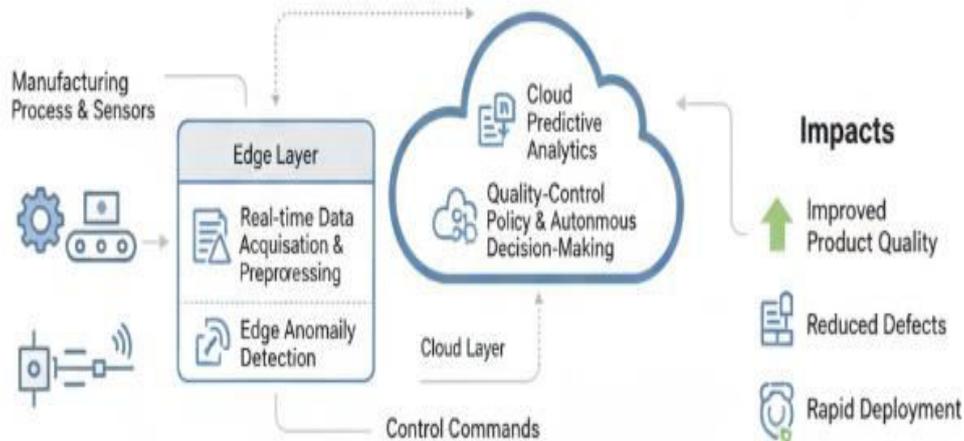

Fig 1: Hierarchical Edge–Cloud Architectures for Autonomous Quality Control: Orchestrating Real-Time Anomaly Detection and Predictive Governance in Smart Manufacturing

## II. BACKGROUND AND CONTEXT

1. Smart Manufacturing and Quality Control

Smart manufacturing systems combine the capabilities of Cyber-Physical Systems (CPS), Industrial Internet of Things (IIoT), Edge-Cloud Artificial Intelligence (AI) ecosystem and Digital Twins to enhance operational efficiency, productivity, safety and quality of products. Quality control is an essential part of smart manufacturing. It mainly consists of two processes: real-time anomaly detection and predictive quality analytics. Real-time anomaly detection verifies whether manufactured products satisfy predefined quality standards. Quality inspection is usually performed using specialized sensors and machine learning models with high predictive performance. Predictive quality analytics exploits historical quality-related data (other than real-time inspection data) to develop predictive models for identifying possible issues in product quality before the products are shipped to customers.

Traditional quality control is primarily human-driven and requires the involvement of quality control professionals for carrying out all-related activities, including setting up inspection sensors, training machine learning models, checking inspection results, performing exploratory data analysis and implementing predictive models. An autonomous and human-independent quality control framework enables manufacturing inspections without human intervention. The framework, which adopts an edge-cloud hierarchy and edge AI techniques, employs data management mechanisms that address data governance, quality, protection and privacy issues. The study focuses on real-time anomaly detection and predictive quality analytics.

2. Edge Artificial Intelligence

Edge computing is a sub-distribution routing architecture that brings computing resources closer to the network edge (where the proximity to the end-user is closest). Therefore, the distance and the cost of transmitting bulk data over bandwidth-constrained links to the central cloud are reduced. Edge computing expands and complements the cloud and can be applied in situations of high data volume and rapid decision-making demands. Edge AI combines edge computing and AI algorithms and components for executing advanced AI algorithms on edge devices. Mist computing extends cloud experts (Cloudlets) to the network edge and edge devices and enables a small data centre of resources making intelligent decision-making at the far edge with a tiny storage capability but high-speed communication.

Logistic (Sigmoid) Function $\sigma(z)=1/(1+e^{-z})$
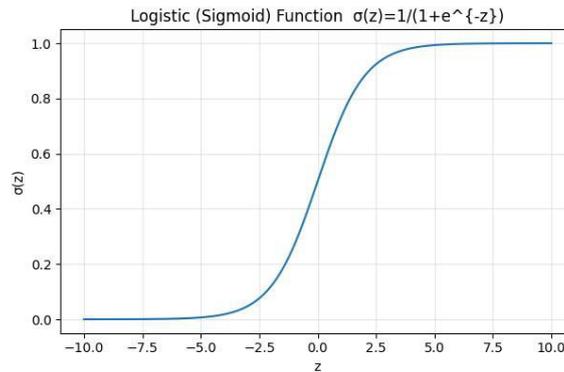
Equation 1) Real-time anomaly detection as a supervised classifier (logistic regression)
1.1 Data + model definition (binary "OK vs Defect")
Let the training dataset be:

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$$

- $\mathbf{x}_i \in \mathbb{R}^d$: feature vector (sensor values, image features, etc.)
- $y_i \in \{0,1\}$: label (0 = normal/OK, 1 = defect/anomaly)

Linear score (logit):

$$z_i = \mathbf{w}^\top \mathbf{x}_i + b$$

Sigmoid to probability:

$$p_i = P(y_i = 1 \mid \mathbf{x}_i) = \sigma(z_i) = \frac{1}{1 + e^{-z_i}}$$

1.2 Likelihood → loss (cross-entropy) step-by-step
For a Bernoulli target y_i, the probability of observing y_i is:

$$P(y_i \mid \mathbf{x}_i) = p_i^{y_i}(1 - p_i)^{(1-y_i)}$$

For the whole dataset (assuming independent samples):

$$\mathcal{L}(\mathbf{w}, b) = \prod_{i=1}^{N} p_i^{y_i} (1 - p_i)^{(1-y_i)}$$

Take log-likelihood:

$$\log \mathcal{L} = \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

We typically **minimize** negative log-likelihood → **cross-entropy loss**:

$$J(\mathbf{w}, b) = - \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

(Optionally average: divide by N.)

1.3 Gradient derivation (key training equations)
We need $\frac{\partial J}{\partial \mathbf{w}}$ and $\frac{\partial J}{\partial b}$.
First, note:

$$p_i = \sigma(z_i), \quad \frac{dp_i}{dz_i} = p_i(1 - p_i)$$

For one sample loss:

$$J_i = -[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

Differentiate w.r.t z_i:

$$\frac{\partial J_i}{\partial z_i} = \frac{\partial J_i}{\partial p_i} \cdot \frac{\partial p_i}{\partial z_i}$$

Compute $\frac{\partial J_i}{\partial p_i}$:

$$\frac{\partial J_i}{\partial p_i} = -\left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i}\right)$$

Multiply by $p_i(1 - p_i)$:

$$\frac{\partial J_i}{\partial z_i} = -\left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i}\right)p_i(1 - p_i)$$

Simplify:

- First term: $-\frac{y_i}{p_i}p_i(1 - p_i) = -y_i(1 - p_i)$
- Second term: $+\frac{1-y_i}{1-p_i}p_i(1 - p_i) = (1 - y_i)p_i$

So:

$$\frac{\partial J_i}{\partial z_i} = -y_i(1 - p_i) + (1 - y_i)p_i = p_i - y_i$$

Now use $z_i = \mathbf{w}^\top \mathbf{x}_i + b$:

$$\frac{\partial z_i}{\partial \mathbf{w}} = \mathbf{x}_i, \quad \frac{\partial z_i}{\partial b} = 1$$

Therefore:

$\frac{\partial J}{\partial \mathbf{w}} = \sum_{i=1}^{N}(p_i - y_i)\,\mathbf{x}_i \quad \frac{\partial J}{\partial b} = \sum_{i=1}^{N}(p_i - y_i)$

1.4 Gradient descent update rules (edge training / retraining)

With learning rate η:

$\mathbf{w} \leftarrow \mathbf{w} - \eta \sum_{i=1}^{N}(p_i - y_i)\,\mathbf{x}_i \quad b \leftarrow b - \eta \sum_{i=1}^{N}(p_i - y_i)$

2.1. Smart Manufacturing and Quality Control

An autonomous quality control solution in smart manufacturing aligns with the vision outlined in Industry 4.0. Quality-control applications that rely on artificial intelligence (AI) serve as examples aligned with this vision. These applications generally exploit the advances provided by edge computing technologies that enable resource-constrained components to process real-time data. For example, in Smart Manufacturing, the manufacturing enterprise effectively manages quality data acquired and analyzed in real-time by resource-constrained edge devices. Such devices apply AI models to detect quality anomalies and predict the quality of parts produced downstream in the process. However, the execution of supervision-based AI tasks typically requires advanced equipment on the cloud or in the private data center.

Research was conducted to develop a solution that satisfies the concept of autonomous edge-cloud orchestration for quality-control applications in a Smart Manufacturing environment, with a focus on the Windows OS platform. Based on this approach, resource-constrained edge devices apply AI models for autonomous quality control in the Smart Manufacturing production line. Specifically, the AI models detect deviations in the normal conditions of the produced parts and predict the quality of parts that may be produced downstream. Since the execution of the models requires the use of advanced equipment, a cloud-based data-flow management strategy is introduced to enable the use of both edge and cloud devices.

2.2. Edge Artificial Intelligence

Systems built on the edge-cloud hierarchy introduce localized intelligence into smart manufacturing systems, aiming to mitigate operations expenses through localized services and accelerate the response times and interactivity to physical-process changes at the edge. Edge intelligence lowers bandwidth consumption, reduces the traffic to cloud resources, and increases the service quality and reliability to final users. An empowered edge has an AI Neuro-Computing Unit and a data-kernel marketplace, where apps can self-assemble; the kernel forwards the values of registered variables to the apps, which exchange JSON packages with other apps. Optimal-placement algorithms enhance resource exploitation and respect real-time constraints inside an AI cloud. Business processes merge AI services and physical processes into a single smart entity, acting as a supplier and consumer of these services using interest-driven data communication.

In-depth insight into the quality of a product at the same time as its production is still a challenge in current smart-factory applications. By packaging the output of the production line and comparing it with the product at the origin, the quality-control inspection represents a major bottleneck in the flow of goods. Anomaly detection on the production line combined with predictive quality analytics and AI explainable models acts autonomously at the edge and anticipatively on cloud resources.

Scenario	End-to-end latency (ms)	Upstream bandwidth (MB/s)
Edge inference	18	0.2
Cloud inference (+network)	220	8.0
Hybrid (edge detect + cloud predict)	70	1.4

III. SYSTEM ARCHITECTURE

A hierarchical architecture that integrates edge computing and cloud orchestration enables autonomous quality control in smart manufacturing. The orchestration layer ensures proper overarching functioning of the autonomous quality control solution.

The hierarchical smart manufacturing architecture exploits edge computing and cloud orchestration. Quality control issues at smart manufacturing facilities are resolved through a set of autonomous mechanisms that utilize distributed edge devices. A dedicated edge-cloud architecture supports data acquisition and real-time detection of defects. Visual anomalies on product surfaces are detected at the edge through an edge-artificial-intelligence solution. Post anomaly detection, cloud processing orchestrates a preventive measure to ascertain quality at a predictive level. For all factory-level quality and predictive-quality concerns, a top-layer cloud implementation leverages facilities across product life-cycles, addresses anomalies timeously, and enables predictive analytics that can identify the need for product-leak testing.

3.1. Edge-Cloud Hierarchy
The proposed solution comprises an edge-cloud data acquisition and orchestration hierarchy. Quality-relevant data are continuously captured from the production process by distributed sources, packaged, and uploaded to a data repository for training, validation, and testing. The deployment of specialized models at the edge detects quality-relevant abnormalities in real time and generates edge-generated quality analytics. Time-series data generated by quality-controlling measurement instruments are also subjected to predictive analytics concerning measures of final product quality—information that is critical for production planning and needs to be relayed to the planning module in the cloud.

Data governance constitutes a major consideration at every step of the data-lifecycle management process. Data acquisition from smart-manufacturing processes involves four concerns: the exposure of sensitive data to third parties, compliance with data-associated privacy regulations such as the General Data Protection Regulation (GDPR) in Europe, the adverse impact of such exposure on the company's intellectual property (IP) and commercial interests, and protection against cybersecurity threats.
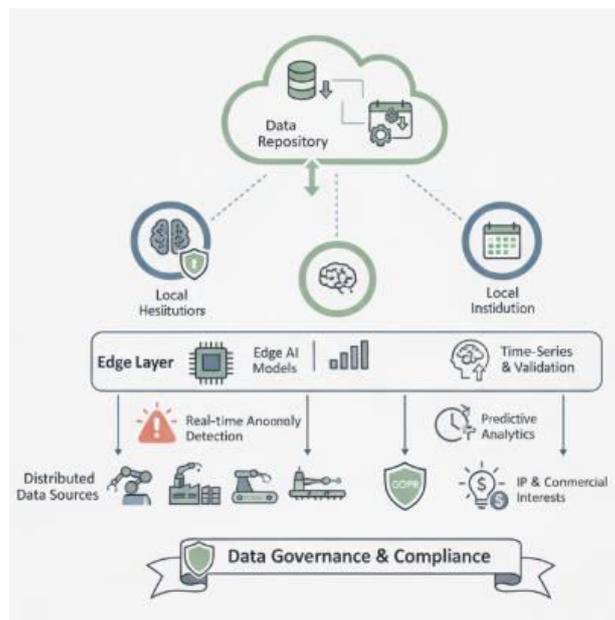

Fig 2: Secure Edge-Cloud Orchestration: A Multi-Tiered Data Governance Framework for Real-Time Quality Analytics in Smart Manufacturing

3.2. Data Acquisition and Preprocessing

An industrial Internet of Things (IoT) platform governs the data acquisition and preprocessing subsystems. The data acquisition subsystem connects to a factory's IoT platform for the collection of production and operational data, including product attributes, product impregnation statuses, process environments, equipment conditions, and inspection results. Collected data may be polluted with noise, defects, or out-of-range sensor or label values. Thus, the data preprocessing subsystem removes invalid data records and pollutants that hinder the subsequent learning and analytical processes. It detects anomalous values by leveraging process correlations of other sensor measurements or Boolean inspection modes. The conditions for anomalous-value detection and ranges for valid sensor values are factory-specific know-how. Integrating these conditions and values requires data-provenencing logics or the support of quality engineers with IoT visions by domain expertise.

Considering the heterogeneity of the acquired data and varying frequencies and proportions, the data-preprocessing subsystem aggregates all data records to a common frequency and desensitizes the signal variations by adjusting the sampling intervals. These pretreatment steps yield suitable training data for the subsequent training of the edge edge-artificial-intelligence (AI) service by leveraging a temporal sequence predictive algorithm. Agnarasen et al. relate to the prediction of an object attribute according to the values of a set of other correlated sensors. By reducing the richness of the input features through correlation analysis, attention-based temporal sequence networks can provide satisfactory object attribute detection. Grouping these networks into subnets with shared model parameters makes the training process easier without sacrificing the detection precision.

IV. AUTONOMOUS QUALITY CONTROL MECHANISMS

Autonomous quality control operations are supported through real-time detection of critical or excessive deviations from normal data patterns. Such deviations uncharacteristically affect a key quality indicator of the manufacturing processes and/or products. In addition, a predictive optimizer provides advance decision-support based on trend analysis of the data relating to that quality indicator. The predictive module evaluates the potential quality degradation and recommends technological or organizational interventions that will minimize or even avert non-conformance to quality standards.

The mechanisms are implemented at the edge level. One mechanism detects abnormal domain data streams and the other predicts the future values of the root causes, most affecting quality. Both mechanisms rely on a historical model of the factory, and warn factory personnel about critical developments in real time and in advance of the occurrence, respectively. Data with reference to a key quality characteristic is being continuously recorded and analysed for training the mechanism for predictive quality analytics. The results provide a basis for earlier intervention and control in relation to process and product quality, and associated business losses.
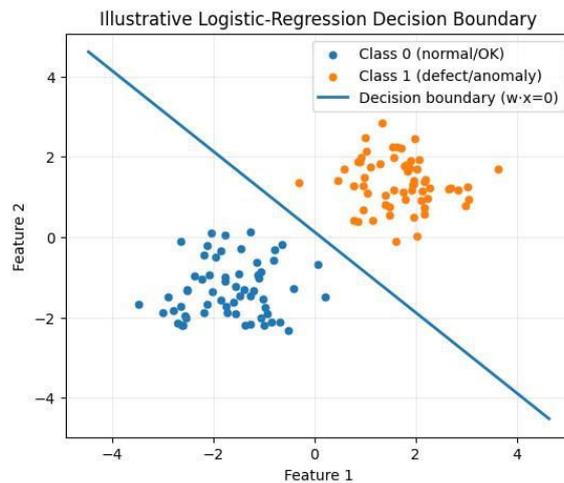
Equation 2) Distance-to-classification-boundary

For a linear classifier boundary:

$$\mathbf{w}^{\top}\mathbf{x} + b = 0$$

2.1 Signed distance derivation (step-by-step)

Distance from point $\mathbf{x}$ to the hyperplane is:

1. The hyperplane normal vector is $\mathbf{w}$.
2. The projection of $\mathbf{x}$ onto the unit normal direction $\frac{\mathbf{w}}{\|\mathbf{w}\|}$ gives distance scale.

The **signed distance**:

$$d(\mathbf{x}) = \frac{\mathbf{w}^\top \mathbf{x} + b}{\|\mathbf{w}\|}$$

- $d(\mathbf{x}) > 0$: on the "defect" side (depending on label convention)
- $d(\mathbf{x}) < 0$: on the "OK" side
- $|d(\mathbf{x})|$: margin/confidence

2.2 Mapping distance → "remaining quality level"
A common, simple mapping is to interpret the logit magnitude as confidence and use sigmoid:
Because $z = \mathbf{w}^\top \mathbf{x} + b$,

$$p(\text{defect} \mid \mathbf{x}) = \sigma(z)$$

If you want "remaining quality" as **higher = better**, define:

$$Q(\mathbf{x}) = 1 - p(\text{defect} \mid \mathbf{x}) = 1 - \sigma(z) = \sigma(-z)$$

Or in distance form (since $z = \|\mathbf{w}\| \cdot d(\mathbf{x})$):

$$Q(\mathbf{x}) = \sigma\left(-\|\mathbf{w}\| \ d(\mathbf{x})\right)$$

4.1. Real-Time Anomaly Detection
Correlating inputs and outputs in manufacturing workflows enables manufacturers to identify states of the workflow and to monitor the appearance of quality-related problems in real time. A supervised anomaly detection model classifies the quality state of the products or the overall workflow according to a set of quality-related input-output signals. During the learning phase the quality states are annotated and then used to train the classification model. The annotated states can be labelled by an expert operator, for example in a low-frequency batch production scenario, or automatically produced by a predictive maintenance model in a high-frequency scenario. The trained model can then run to classify the products' quality in real time. As for any classification model, the real-time observations are compared with the predictions. When a divergence (an anomaly) occurs, the corresponding information is propagated as an alarm (i.e., anomaly detected) to the quality-control supervisor.

For collaborative production with multiple production workflows involved, such as the example case of jewelled watch production with watch case, watch movement, and watch strap assembly workflows running on different production systems, the correlation of quality-related input-output signals of the collaborative production workflows can be built to detect the potential quality issues of product collaboration quality.

4.2. Predictive Quality Analytics
Predictive quality analytics examine quality trends over time with a view to estimating future quality across sets of similar products and providing timely alerts to possible quality issues. By preventing deviations from quality norms, especially in large-scale manufacturing, production costs can be reduced. The analytics process begins with the automatic selection of relevant continuous quality attributes or measurements from the repository of finally approved quality data. The next step is to collect the values of the selected measurements from their respective databases in a predefined window length. After prediction model development, the model is deployed for real-time inference.

Whenever the incoming values for the selected measurements exceed the trigger point for alert generation, a predicted quality alert is raised. When at least a fixed number of consecutive alerts are raised, or whenever a predefined period of continuous alerts is reached, a possible future quality issue alert is generated. The predictive quality analytics framework has been successfully implemented in a simulated hard-disk drive manufacturing unit. Data from the publicly available SECOM dataset, a process for manufacturing the production floor piece with real-time attempt and anomaly detection, and prospects for helping edge devices for past production analysis are presented.

	Pred 0	Pred 1
Actual 0 (OK)	930	30
Actual 1 (Defect)	45	95

V. DATA GOVERNANCE, SECURITY, AND PRIVACY

Data governance promotes data quality, manageability, and security. Creating a variety of data for multiple users opens the possibility of sensitive information being destroyed or released unexpectedly. This section outlines the data protection quality-control process and the applicable regulations.

The Data Protection Act 2018 regulates the processing of personal data within the United Kingdom. It is based on the EU General Data Protect Regulation, the EU Law Enforcement Directive, and the Regulation of Investigatory Powers Act. The regulation is compliant with the Council of Europe's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data and the United Nations Convention on the Rights of Persons with Disabilities. A data protection officer is designated to monitor compliance. Data are processed in accordance with the data subject's rights, including the right of access to their personal data. Appropriate technical measures are in place to protect data. Data sharing is controlled and written data-sharing agreements are put in place.

Most other jurisdictions support freedom of data publication, with appropriate privacy protections. Even when not required by law, granting unrestricted access to sensitive data can lower their value. Organisational policies are reinforced to make sure no data modification becomes irreversible. As much as possible, all modifications are reversible. Furthermore, only datasets that are legally permitted to be shared without restriction are published.

5.1. Data Management Strategies and Compliance Frameworks

As advanced manufacturing systems embrace business process optimization and data governance, so AI-augmented autonomous quality control solutions also address management and compliance related to data sensitivity, legal ownership, and security. Indeed, due to privacy concerns, errors, and the potential for AI-based solutions to present fake or misleading information, explainability, fairness, security, and quality control processes must also be considered within data management plans. Consequently, data at rest as well as in transit and use demand protection based on fundamental principles.

An in-depth explanation of a Data Management Plan Strategy considers processes, methodologies, and tools dedicated to the protection of users' sensitive data within an Intelligent Cloud-Edge System. To guarantee the proper use of data security mechanisms, a DPM framework, encompassing authentication, authorization, audit, confidentiality, integrity, and availability, provides a set of processes, methodologies, and tools designed to secure all users' data resources against threats across cloud-edge-based services. Such a DPM guarantees confidentiality by defining a role-based access mechanism, integrity through a watermarking strategy, availability by implementing redundancy mechanisms, and protection against malicious operations via an audit procedure, and it supports cloud-edge security by adopting a privacy-preserving cloud-edge service selection algorithm.
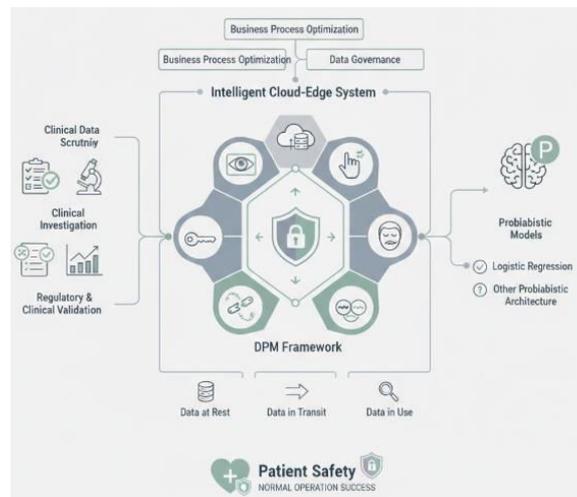

Fig 3: Securing the Intelligent Frontier: A Holistic Data Management Plan (DMP) Framework for AI-Augmented Cloud-Edge Manufacturing

5.2. Best Practices for Data Protection and Regulatory Compliance

To mitigate the risk of incidents arising from improper data management practices, the identified data governance framework must be aligned with established industry standards and best practices. These provide a roadmap for businesses to implement appropriate security control mechanisms and ensure effective data governance across the entire lifecycle of AI systems. Within the proposed architecture, data protection is crucial, as the control units contain real-time quality data and models that detect quality anomalies and predict product quality.

Moreover, European General Data Protection Regulation policies must be rigorously followed, as small manufacturing companies face fragile data protection policies. Regardless of the high value of cameras in the production environment, their data can be rarely stored or used for model training. Edge-cloud data management must follow a cloud-inclusive setup, allowing data capturing for training purposes only when strictly needed and with consent from supervisors. These best practices enable small companies to keep cloud expenses low while fulfilling GDPR rules.

Metric	Value
Accuracy	0.9318181818181818
Precision (Defect)	0.76
Recall (Defect)	0.6785714285714286
F1 (Defect)	0.7169811320754718

VI. EVALUATION AND VALIDATION

To validate autonomous quality control capabilities, an edge-cloud ecosystem orchestrates predictive quality analytics through AI-enabled cloud solutions. Edge devices enable real-time anomaly detection using AI models trained on sensor metadata and images, with inference occuring directly on-board. Sensor measurements are securely transmitted to the cloud for predictive quality modelling. Data transfer is controlled by a dedicated data governance module, which automatically selects the most relevant records for quality modelling and ensures compliance with data-security regulations. Results from a smart factory-quality dataset demonstrate the feasibility and performance of the edge AI solutions for real-time detection of anomalous samples, while future work aims to complete the end-to-end system and address data governance, security, and compliance challenges when using cloud datasets.

The growing concern of data security regulations, together with the increasing need to demonstrate compliance with regulations such as the European General Data Protection Regulation (GDPR), has promoted the development of data governance solutions and strategies for data protection when storing and sharing information. Within an edge-cloud context, these solutions and strategies have been typically considered in the cloud zone, verifying and controlling data transfer from the edge devices to the central repository. Such approaches generally allow for developing highly accurate ad hoc models while ensuring data security. However, as regulations evolve and change, the need for up-to-date data becomes crucial.

Distance-to-Classification-Boundary Trend (Illustrative)
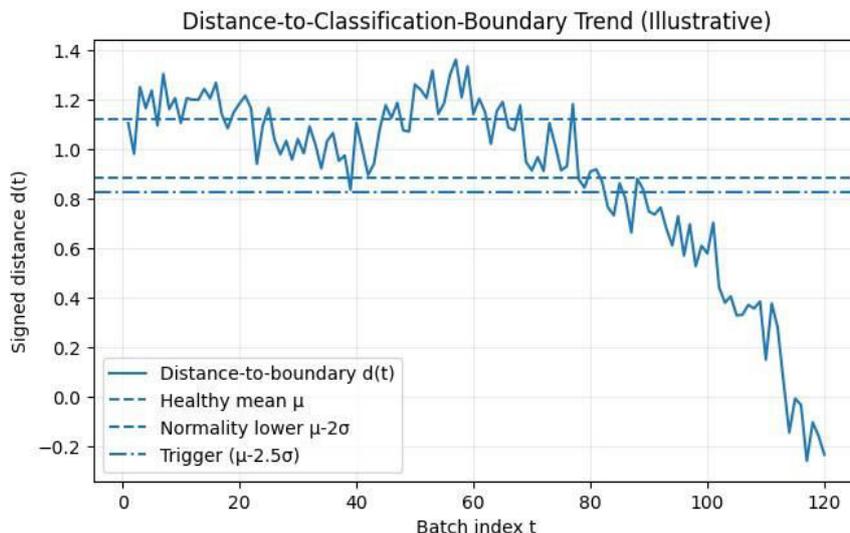
Equation 3) Trend monitoring + "normality region" detection (early warning)

Let d_t be the signed distance (or any quality indicator) for batch/time index t.

3.1 Build a normality region from healthy history

Using a baseline "healthy" window $t \in \mathcal{H}$:

Mean:

$$\mu = \frac{1}{|\mathcal{H}|} \sum_{t \in \mathcal{H}} d_t$$

Std (sample):

$$\sigma = \sqrt{\frac{1}{|\mathcal{H}| - 1} \sum_{t \in \mathcal{H}} (d_t - \mu)^2}$$

Normal band (example):

$$[\mu - 2\sigma, \ \mu + 2\sigma]$$

Trigger threshold (stricter):

$$\tau = \mu - 2.5\sigma$$

(Choose sign depending on whether decreasing distance indicates degradation.)

3.2 Smooth the trend (moving average)

Moving average over window W:

$$\bar{d}_t = \frac{1}{W} \sum_{k=0}^{W-1} d_{t-k}$$

Alert if smoothed indicator crosses trigger:

$$A_t = \begin{cases} 1 & \bar{d}_t < \tau \\ 0 & \text{otherwise} \end{cases}$$

6.1. Future Directions and Insights

Increasingly complex manufacturing environments demand effective data governance strategies that facilitate intelligent decision-making while minimizing risks to product quality, equipment reliability, and labour force safety. The emergence of edge artificial intelligence (AI) accelerates data delivery timelines and enhances information security, but the subset of fast-response quality management systems remains limited. Advanced analytics on data hovering close to the source provide accelerated support for autonomous quality control over production setups equipped with integrated vision systems. Novel mechanisms for real-time anomaly detection and predictive quality analytics have been devised.

Validated via a smart manufacturing demonstrator, the control techniques support rapid identification of non-conforming products for corrective action or removal from the normal flow, as well as prediction of product quality indicators based on sensor data and feature extraction from production images. Coordination between cloud services and an edge AI gateway harnesses the combined power of the edge and cloud compute capabilities. However, the architecture and the development of such capabilities must be enriched with data governance strategies to achieve autonomous data operation without compromising security and privacy.

VII. CONCLUSION

The proposed architecture and algorithms address the needs for autonomous and non-intrusive quality control in smart manufacturing. Orchestration of AI services and data at the edge and cloud levels aims to minimize latency and energy consumption while maximizing security and compliance. Real-time anomaly detection enables early identification of deviations from the expected production quality; predictive quality analytics avoids defective products and any associated manufacturing costs, thus ensuring financial sustainability. Making these AI services available as a catalog empowers any stakeholder to autonomously monitor the production quality through convenient interfaces and without direct involvement of skilled personnel in the development, training, maintenance, or execution of the models. Most importantly, such autonomous quality control is realized in compliance with industry regulations, proactively preserving production data confidentiality and privacy.

The research and development jointly performed at Politecnico di Milano, Palo Alto University, RMIT, and Istituto Italiano di Tecnologia open further opportunities to benefit from edge-cloud collaboration in multimodal data

management by exploiting the rich information present in the datasets involved in the aforementioned production quality control activities.
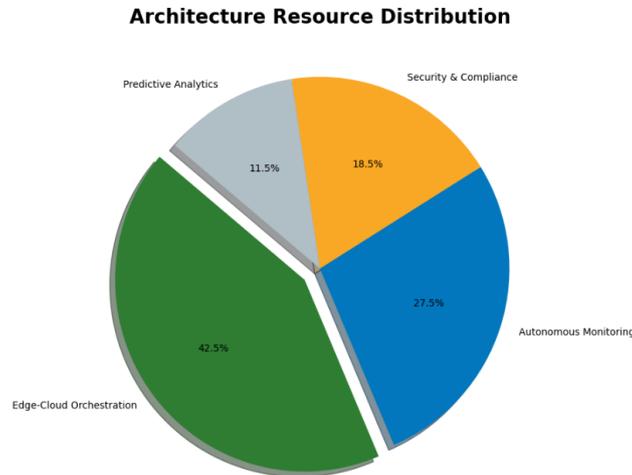

Fig 4: Architecture Resource Distribution

7.1. Final Thoughts and Implications for Industry

The proposed approach represents a step toward autonomous production quality control. Operationalizing this capability reduces the need for human oversight while gaining operational and production performance advantages. Self-managed edge AI solutions can restore their own real-time control apply of damage prediction and mitigation operational playbooks. Cloud-based dynamic workload allocation can optimize the TCO of human resources.

Evidence obtained from an actual implementation indicates that effecting predictive quality analytics improves production performance and reduces the overall cost of quality. Future work will continue extending the two autonomous quality control mechanisms—real-time anomaly detection and predictive quality analytics—through the data governance, security, and privacy lenses. Cloud-based dynamic workload orchestration support ensure enterprise-wide cost-effective delivery. The solution is expected to be especially valuable for production environments where human supervision is limited or impractical, yet a stable production quality is vital for retaining customers and ensuring safety.

REFERENCES

[1]   Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (pp. 265–283). USENIX.
[2]   Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 633-652.
[3]   Akhtar, A., Khan, M., & Nazir, S. (2021). Industrial anomaly detection: A survey of methods and applications. Computers & Industrial Engineering, 158, 107377.
[4]   IT Integration and Cloud-Based Analytics for Managing Unclaimed Property and Public Revenue. (2024). MSW Management Journal, 34(2), 1228-1248.
[5]   Alur, R., & Dill, D. L. (1994). A theory of timed automata. Theoretical Computer Science, 126(2), 183–235.
[6]   Angelopoulos, C. M., Nikoletseas, S., & Patroumpa, D. (2020). Edge computing in the Industrial Internet of Things: A survey. IEEE Internet of Things Journal, 7(10), 10665–10682.
[7]   Agentic AI in Data Pipelines: Self Optimizing Systems for Continuous Data Quality, Performance and Governance. (2024). American Data Science Journal for Advanced Computations (ADSJAC) ISSN: 3067-4166, 2(1).
[8]   Babiceanu, R. F., & Seker, R. (2016). Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. Computers in Industry, 81, 128–137.
[9]   Meda, R. (2024). Agentic AI in Multi-Tiered Paint Supply Chains: A Case Study on Efficiency and Responsiveness. Journal of Compu-tational Analysis and Applications (JoCAAA), 33(08), 3994-4015.

[10]   Bagheri, B., Yang, S., Kao, H.-A., & Lee, J. (2015). Cyber-physical systems architecture for self-aware machines in Industry 4.0 environment. IFAC-PapersOnLine, 48(3), 1622–1627.

[11]   Nagabhyru, K. C. (2024). Data Engineering in the Age of Large Language Models: Transforming Data Access, Curation, and Enterprise Interpretation. Computer Fraud and Security.

[12]   Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM.

[13]   Benosman, R., & Kang, S. B. (Eds.). (2001). Panoramic vision: Sensors, theory, and applications. Springer.

[14]   Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281–305.

[15]   Aitha, A. R. (2024). Generative AI-Powered Fraud Detection in Workers' Compensation: A DevOps-Based Multi-Cloud Architecture Leveraging, Deep Learning, and Explainable AI. Deep Learning, and Explainable AI (July 26, 2024).

[16]   Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175–1191). ACM.

[17]   Bosch, J. (2018). Speed, data, and ecosystems: The future of software engineering. IEEE Software, 35(1), 82–88.

[18]   Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932

[19]   Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT 2010 (pp. 177–186). Physica-Verlag.

[20]   Deep Learning-Driven Optimization of ISO 20022 Protocol Stacks for Secure Cross-Border Messaging. (2024). MSW Management Journal, 34(2), 1545-1554.

[21]   Burns, B., Beda, J., & Hightower, K. (2019). Kubernetes: Up & running (2nd ed.). O'Reilly Media.

[22]   Cao, Y., Jia, X., Chen, Y., Lin, S., & Zhang, X. (2020). Deep learning for industrial inspection: A survey. IEEE Transactions on Industrial Informatics, 16(8), 4876–4891.

[23]   Meda, R. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. Educational Administration: Theory and Practice.

[24]   Chen, D., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM.

[25]   Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209.

[26]   Aitha, A. R. (2023). CloudBased Micro services Architecture for Seamless Insurance Policy Administration. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 607-632.

[27]   Yandamuri, U. S. (2022). Cloud-Based Data Integration Architectures for Scalable Enterprise Analytics. International Journal of Intelligent Systems and Applications in Engineering, 10, 472-483.

[28]   Collins, E., & Nechvatal, J. (2020). NIST privacy framework: A tool for improving privacy through enterprise risk management (Version 1.0). National Institute of Standards and Technology.

[29]   Segireddy, A. R. (2024). Machine Learning-Driven Anomaly Detection in CI/CD Pipelines for Financial Applications. Journal of Computational Analysis and Applications, 33(8).

[30]   Craswell, N., Mitra, B., Yilmaz, E., Campos, D., & Voorhees, E. M. (2020). Overview of the TREC 2020 Deep Learning Track. In Proceedings of the Text REtrieval Conference (TREC 2020). NIST.

[31]   Croft, W. B., Metzler, D., & Strohman, T. (2010). Search engines: Information retrieval in practice. Addison-Wesley.

[32]   Dai, Z., Yang, Z., Yang, Y., Cohen, W. W., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-XL: Attentive language models beyond a fixed-length context. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 2978–2988). ACL.

[33]   Varri, D. B. S. (2024). Adaptive and Autonomous Security Frameworks Using Generative AI for Cloud Ecosystems. Available at SSRN 5774785.

[34]   Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019 (pp. 4171–4186). ACL.

[35]   Yandamuri, U. S. AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology.

[36]   Singireddy, J. (2024). AI-Enhanced Tax Preparation and Filing: Automating Complex Regulatory Compliance. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 2(1).

[37]   Dourish, P. (2004). What we talk about when we talk about context. Personal and Ubiquitous Computing, 8(1), 19–30.

[38]   Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407.

[39] Keerthi Amistapuram. (2024). Federated Learning for Cross-Carrier Insurance Fraud Detection: Secure Multi-Institutional Collaboration. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 6727–6738. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/3934

[40] Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. *International Journal Of Finance*, *36*(6), 682-706.

[41] Farooq, M. S., Khan, Z., Ahmad, R., Islam, S. U., & Kim, S. W. (2023). A survey on the role of industrial IoT in manufacturing for Industry 4.0. Sensors, 23(21), 8958.

[42] Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926.

[43] Fowler, M. (2018). Refactoring: Improving the design of existing code (2nd ed.). Addison-Wesley.

[44] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144.

[45] Paleti, S. (2024). Transforming Financial Risk Management with AI and Data Engineering in the Modern Banking Sector. American Journal of Analytics and Artificial Intelligence (ajaai) with ISSN 3067-283X, 2(1).

[46] Davuluri, P. N. Event-Driven Compliance Systems: Modernizing Financial Crime Detection Without Machine Intelligence.

[47] Sheelam, G. K., & Koppolu, H. K. R. (2024). From Transistors to Intelligence: Semiconductor Architectures Empowering Agentic AI in 5G and Beyond. Journal of Computational Analy- sis and Applications(JoCAAA), 33(08), 4518-4537.

[48] Gray, J., & Reuter, A. (1993). Transaction processing: Concepts and techniques. Morgan Kaufmann.

[49] Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems.

[50] Guo, J., Fan, Y., Ai, Q., & Croft, W. B. (2020). A deep look into neural ranking models for information retrieval. Information Processing & Management, 57(6), 102067.

[51] Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In Proceedings of ICLR 2016.

[52] Inala, R. Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective.

[53] He, W., Xu, L. D., & Chen, H. (2014). Internet of Things in industries: A survey. IEEE Transactions on Industrial Informatics, 10(4), 2233–2243.

[54] Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. International Journal of Scientific Research and Modern Technology, 1(12), 216-226.

[55] Hohpe, G., & Woolf, B. (2003). Enterprise integration patterns: Designing, building, and deploying messaging solutions. Addison-Wesley.

[56] Amistapuram, K. (2024). Generative AI in Insurance: Automating Claims Documentation and Customer Communication. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 15(3), 461–475. https://doi.org/10.61841/turcomat.v15i3.15474

[57] IEC. (2018). IEC 62443-3-3:2013 + AMD1:2017 + AMD2:2020 Industrial communication networks—Network and system security—Part 3-3: System security requirements and security levels. International Electrotechnical Commission.

[58] ISO. (2018). ISO/IEC 27001:2018 Information security management systems—Requirements. International Organization for Standardization.

[59] Guntupalli, R. (2024). Enhancing Cloud Security with AI: A Deep Learning Approach to Identify and Prevent Cyberattacks in Multi-Tenant Environments. Available at SSRN 5329132.

[60] Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.

[61] Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems, 20(4), 422–446.

[62] Koppolu, H. K. R., & Sheelam, G. K. (2024). Machine Learning-Driven Optimization in 6G Telecommunications: The Role of Intelligent Wireless and Semiconductor Innovation. Global Research Development (GRD) ISSN: 2455-5703, 9(12).

[63] Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.

[64] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Rouayheb, S. E., Gascón, A., Ghazi, B., Gibbons, P. B., Hastie, T., Hazy, T., Kalenichenko, D., Kamath, G., … Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.

[65] Lahari Pandiri, "AI-Powered Fraud Detection Systems in Professional and Contractors Insurance Claims," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), DOI 10.17148/IJIREEICE.2024.121206.

[66] Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. *Online Journal of Engineering Sciences*, *1*(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/ojes/article/view/1360

[67] Khattab, O., & Zaharia, M. (2020). ColBERT: Efficient and effective passage search via contextualized late interaction over BERT. In Proceedings of SIGIR 2020 (pp. 39–48). ACM.

[68] Rongali, S. K. (2023). Explainable Artificial Intelligence (XAI) Framework for Transparent Clinical Decision Support Systems. International Journal of Medical Toxicology and Legal Medicine, 26(3), 22-31.

[69] Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18–23.

[70] Lee, J., Jin, C., & Bagheri, B. (2017). Cyber physical systems for predictive production systems. Production Engineering, 11(2), 155–165.

[71] Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls.

[72] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

[73] Mashetty, S., Challa, S. R., ADUSUPALLI, B., Singireddy, J., & Paleti, S. (2024). Intelligent Technologies for Modern Financial Ecosystems: Transforming Housing Finance, Risk Management, and Advisory Services Through Advanced Analytics and Secure Cloud Solutions. Risk Management, and Advisory Services Through Advanced Analytics and Secure Cloud Solutions (December 12, 2024).

[74] Liu, T.-Y. (2009). Learning to rank for information retrieval. Foundations and Trends in Information Retrieval, 3(3), 225–331.

[75] Rongali, S. K., & Kumar Kakarala, M. R. (2024). Existing challenges in ethical AI: Addressing algorithmic bias, transparency, accountability and regulatory compliance.

[76] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

[77] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.

[78] Guntupalli, R. (2024). AI-Powered Infrastructure Management in Cloud Computing: Automating Security Compliance and Performance Monitoring. Available at SSRN 5329147.

[79] Siva Hemanth Kolla. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(3s), 495–506. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8037

[80] Nagubandi, A. R. (2023). Advanced Multi-Agent AI Systems for Autonomous Reconciliation Across Enterprise Multi-Counterparty Derivatives, Collateral, and Accounting Platforms. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 653-674.

[81] Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning (2nd ed.). MIT Press.

[82] Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965

[83] NIST. (2020). Security and privacy controls for information systems and organizations (NIST SP 800-53 Rev. 5). U.S. Department of Commerce.

[84] Chava, K. (2024). The Role of Cloud Computing in Accelerating AI-Driven Innovations in Healthcare Systems. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 2(1).

[85] Siva Hemanth Kolla. (2023). Deep Learning–Driven Retrieval-Augmented Generation for Enterprise ITSM Automation: A Governance-Aligned Large Language Model Architecture. *Journal of Computational Analysis and Applications (JoCAAA)*, *31*(4), 2489–2502. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/4774

[86] Object Management Group. (2019). Decision model and notation (DMN), version 1.3. OMG.

[87] Rongali, S. K. (2024). Federated and Generative AI Models for Secure, Cross-Institutional Healthcare Data Interoperability. Journal of Neonatal Surgery, 13(1), 1683-1694.

[88] Kolla, S. H. (2024). RETRIEVAL-AUGMENTED GENERATION WITH SMALL LLMS FOR KNOWLEDGE-DRIVEN DECISION AUTOMATION IN ENTERPRISE SERVICE PLATFORMS. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *15*(3), 476–486. https://doi.org/10.61841/turcomat.v15i3.15497

[89] AI and ML-Driven Optimization of Telecom Routers for Secure and Scalable Broadband Networks. (2024). MSW Management Journal, 34(2), 1145-1160.

[90] Singh, R., Auluck, N., & Rana, O. (2023). Edge AI: A survey. Results in Engineering, 18, 101053.