MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

AI-Based Anomaly Detection Frameworks in Distributed Enterprise Data Systems

Velangani Divya Vardhan Kumar Bandi
Director AI/ML Engineering
ORCID ID: 0009-0008-7949-5670

Abstract

Enterprise information systems typically employ a decentralized architecture based on data stored across numerous heterogeneous computing environments. These environments are also employed by other organizations that provide services to a significant user base. With such user bases come volume and frequency of operations that are unprecedented. Furthermore, systems that allow the execution of user-controlled queries can have unpredictable types and patterns of operations. Consequently, unified enterprise data systems may lag behind such developments. Artificial intelligence can counteract human operators' limitations on detecting atypical situations in these decentralized systems. However, current applications often result from piecemeal isolated initiatives by data scientists from diverse parts of the organization, which quickly become technical debts. Four decisive aspects of fully supporting the atypical event detection process with artificial intelligence and its implications on enterprise information systems have emerged from a synthesis of the academic literature over the last several decades.First, a comprehensive taxonomy of current solutions is essential to manage the large number of proposals, as the expressed needs of enterprises imply the possibility of artificial intelligence detecting atypical events in any part of the data systems. Second, defining general characteristics of the enterprise data systems' architecture is key because many of the proposed solutions are strongly dependent on these characteristics, especially aspects related to the sources of data and their pipelines. Third, the data requirements and operational principles of the adapted technical solutions must be covered. Fourth, underlying requirements for data governance, privacy, and security must be considered, together with the implications of regulatory pressures for the application and development of artificial intelligence.

Keywords: Enterprise Information Systems, Decentralized Data Architectures, Atypical Event Detection, AI-Driven Anomaly Detection, Heterogeneous Computing Environments, Large-Scale Data Operations, Autonomous Monitoring Systems, Enterprise Data Pipelines, Technical Debt in AI Initiatives, Solution Taxonomy for Event Detection, Data Source Heterogeneity, Operational AI Architectures, Data Governance and Compliance, Privacy-Preserving Analytics, Secure AI Deployment, Regulatory-Aware AI Systems, Intelligent Observability, Scalable Detection Frameworks, Enterprise-Wide AI Integration, Adaptive Monitoring Infrastructure.

1. Introduction

Modern enterprise data systems provide significant opportunities for efficient operations, effective decision-making, and the creation of valuable assets. However, the increasingly distributed nature of such systems leads to operational complexity that can be tackled using advanced AI-based techniques to detect anomalies in transactions and generated data. Anomaly detection problems can be addressed using statistical and machine learning techniques, including supervised, semi-supervised, unsupervised, and self-supervised approaches. While existing work has proposed anomaly detection frameworks for enterprise data systems, these systems are based mainly on supervised techniques that rely on labelled training sets.Four architectural considerations specific to any anomaly detection application in distributed data systems shape the AI models' design. A first aspect concerns the data source, and especially the level heterogeneity among distributed data sets. A second aspect relates to the data pipelines where data latency is often critical or even a requirement for the functioning of enterprise processes. Anomalies can be detected closer to their origin, thus bypassing multilayer data pipelines. The third architectural consideration refers to feature engineering for supervised methods. In most cases, the usefulness of the final trained models depends on how informative the features are and basic domain knowledge is used to identify relevant dimensions. Self-supervised pre-trained foundation models simplify the adaptation of representation learning techniques to specific applications.

1.1. Overview of the Study and Its Objectives

Current large-scale data-driven systems rely on real-time processing of data mixed from various sources. These systems are susceptible to various forms of operational abnormality that may have a costly impact on business processes, such as faulty transaction handling in a banking system or delayed response to user requests in a streaming service. Such abnormalities can be detected by AI models with feature representation learning capabilities, trained on past patterns of the data stream. Each anomaly report triggers an investigation into the cause immedi- ately or within a few hours, although not all anomalies are necessarily harmful. Anomaly detection techniques can be classified into two broad categories: supervised and unsupervised. In the supervised setting, a sequence of normal and abnormal patterns is used to train an anomaly detector, while in an unsupervised system, only normal sequences are used to train the detector.

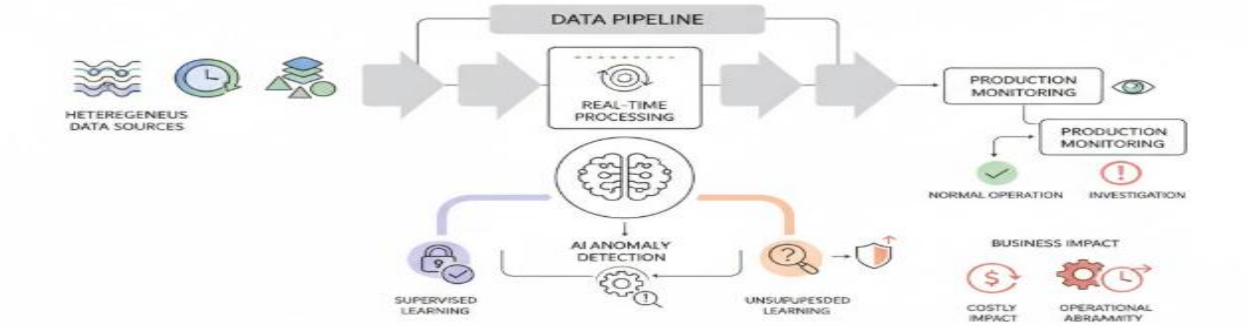

Fig 1: Adaptive Representation Learning for Real-Time Anomaly Detection in Heterogeneous Distributed Data Pipelines

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

According to the characteristics of data sources, systems, and pipelines used in distributed enterprise data systems, different types of model can be trained and evaluated in production, during a period– during which the system is running normally in the wild. During this learning phase, the model learns to detect abnormalities, extracting the features that help in distinguishing between normal and anomalous patterns. During normal operation, the AI model generates anomaly reports whenever it detects an unusual pattern. A data pipeline can be very complex, moving data through many data sources and transformations. Latency must be kept low because the consumers are interested in real-time or near-real-time updates. Data sources are heterogeneous; some produce continuous streams of Data, while others send periodic updates. A single anomaly model cannot be trained for every data stream.

2. Background and Motivation

The purpose of any organizational computing system is to support strategic business objectives. Consequently, managing external resources, customers, suppliers, and internal processes should result in clean information stored in a distributed enterprise data system. However, operational, logical, or semantic errors can occur in the data. These anomalies arise sporadically and vary in volume, type of occurence, or severity.

Enterprise data systems now collect data in heterogeneous forms from multiple sources. Due to pollution from common operational processes, complete and clean training datasets are rarely available. In addition, the growing focus on environmental sustainability combined with the need for faster deliveries to the market imposes added constraints on the design and implementation of the business processes and their support systems. Consequently, detection latency has become an important consideration in the design.

2.1. Fundamental Concepts and Definitions

Anomaly detection (AD) is a well-studied field of machine learning (ML) that aims to characterise rare events in a dataset that differ significantly from the majority of observations. AD tasks traditionally fall into three categories: (i) supervised detection of known rare events, (ii) unsupervised detection of unknown rare events, and (iii) detection of rare-event patterns in a self-supervised manner. Supervised methods typically require training sets with labelled normal and anomalous observations, while unsupervised techniques aim to learn a model from the normal class only, without using any samples from the rare event categories. Self-supervised approaches utilise partially labelled data with a few instances from the anomaly classes. More recently, AD-formulated zero- and few-shot methods have emerged. These methods thrive with large pretrained models that capture the complex notion of normality through representation learning and can generalise from limited prior knowledge about the potentially infinite types of anomalies.

AD studies can be classified according to (i) the composition of the model training data and (ii) the interactions between input features and model architectures. In supervised and semi-supervised AD, the training set includes a subset of the predefined anomalies. In unsupervised and self-supervised AD, no training data is available for the anomaly classes. Unlabelled AD methods have an inherent advantage, as real-world anomaly events — such as fraud, failures, or security breaches — occur only infrequently in distributed enterprise data systems. Training a proper supervised detection model requires a lot of anomaly data, although domain experts can easily define the detection classes and provide labelled data when needed. Therefore, unsupervised and self-supervised approaches are the default choice for AD in enterprise data systems.

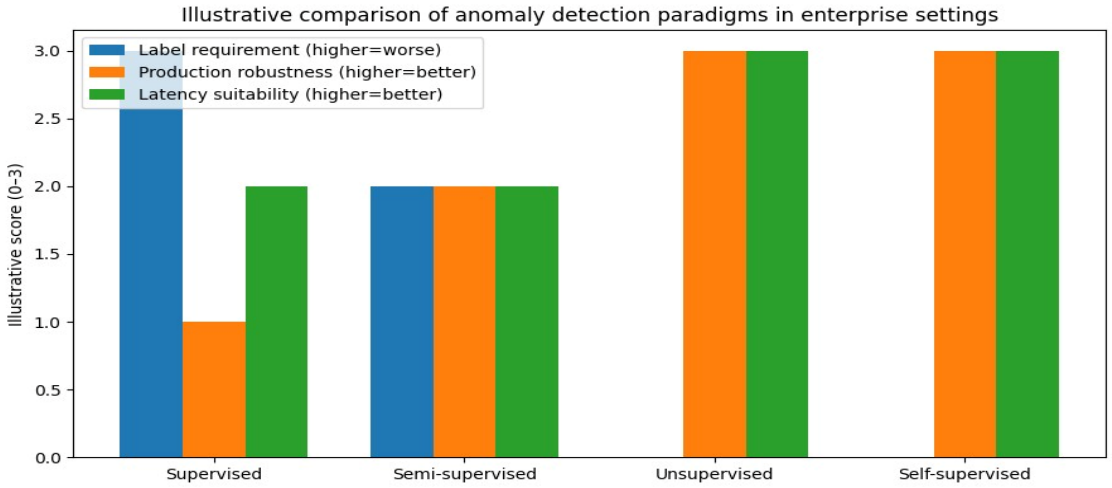

Fig 2: Generalized Anomaly Scoring Framework for Streaming Events in Distributed Enterprise Pipelines

Equation 1) Core anomaly-detection setup

Let a distributed pipeline emit events/records over time (streaming or batch).
Represent each event at time t as a feature vector:

$$\mathbf{x}_t \in \mathbb{R}^d$$

Goal: compute an **anomaly score** $s(\mathbf{x}_t)$ such that higher scores mean "more unusual," then alert if score exceeds a threshold τ:

$$\text{alert}(\mathbf{x}_t) = \mathbf{1}\{s(\mathbf{x}_t) > \tau\}$$

3. Taxonomy of Anomaly Detection Techniques

A wide variety of architecture frameworks that enable the detection of anomalies in enterprise data systems are instantly being developed. These frameworks adopt distinct high-level views, different sets of types of anomaly detection algorithms, and diverse aspects which are examined and taken into consideration. Therefore, posing the

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

question of how these frameworks can be informed and categorized. Possible considerations are the detections which are made and the underlying fault detection model. The AI-based framework presented in previous studies is classified together with a number of other frameworks built for detecting anomalies in enterprise data systems, highlighting the fundamental principles and properties of the anomaly detection models which are being employed.

Any anomaly can potentially indicate an unusual activity. In security applications, anomalies represent all the different types of attacks in place. In processes, various causes or process disturbances can introduce anomalies (the misbehaving components can be detected in a supervised way). In an enterprise data system, people or processes that are not acting normally can introduce an anomaly/incident. However, there are no labelled examples providing a clear distinction between relevant and normal behaviors because labelled data is not available a priori. Recognizing patterns of domain experts, gaining experience over time, performing certain domain tests, and even using methods such as anomaly detection can aid in defining anomalies for categorizing and studying incidents.

3.1. Supervised and Semi-Supervised Methods

Supervised models for anomaly detection learn the distinction between normal and abnormal observations from labelled training data and require sufficiently large amounts of labelled data for optimal performance. Training or semi-supervised classifiers learn to separate normal and anomaly classes is a natural extension of traditional classifiers for anomaly detection. Class and region overlapping problems, however, may impose challenges for classifiers' generalizability and are generally overlooked. Some approaches leverage the implicit transfer ability of models pre-trained on general tasks, which in conjunction with smaller amounts of labelled data have helped model the representation capability of generative models. These generative models are included in the classifiers' formulation in a similar fashion to traditional transfer learning.

Semi-supervised and supervised models have also been evaluated in the unsupervised setting, with nearest neighbour based methods and one-class SVM's among the top performers. Additional techniques have also been explored to reduce the severity of the overfitting problem, including augmenting the normal training set with generated samples, meta-learning, and adversarially augmenting the training set. Other constructive ideas include using modality-specific supervised classifiers in a multi-modal setting, where generic visual classifiers operate on the visual modality and an audio classifier distinguishes the logical sounds from background noise, and learning directly from labelled anomaly-free regions in normal samples.

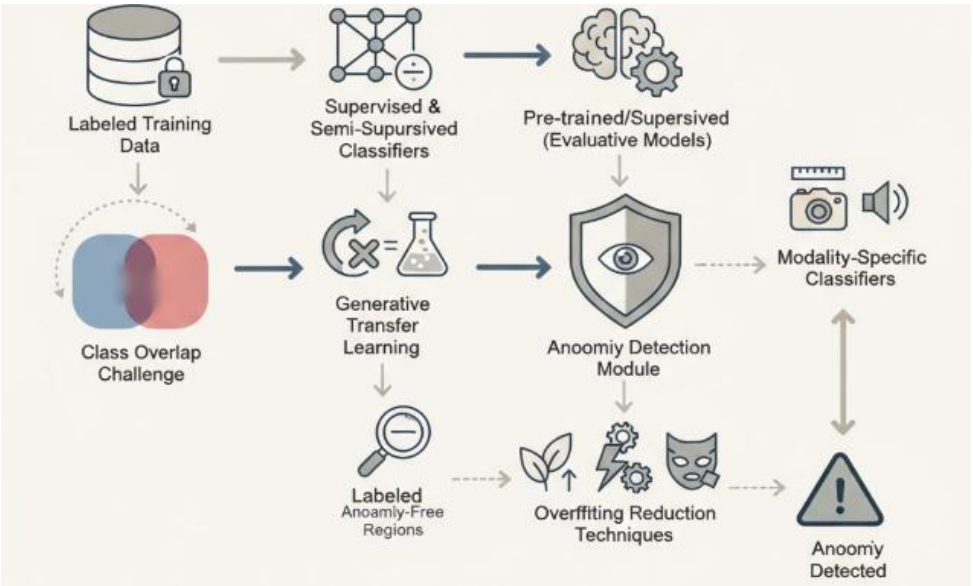

Fig 3: Hybrid Augmentation and Generative Transfer: Mitigating Class Overlap and Data Scarcity in Multi-Modal Anomaly Detection

3.2. Unsupervised and Self-Supervised Approaches

Unsupervised techniques identify anomalies without explicit supervision by training with data free from explicit anomalies. Unsupervised learning seeks to utilize reasonable assumptions on normal data to isolate anomalies for example, data density modeling, where the learned data density is low in the region of the anomalies or inventing normal data to train supervised classifiers. In terms of representation learning, a factorization will hold, with the feature inference model having contributions that is piece-wise linear order preserving with respect to the normal object features such that the errors converge to normal data density. The condition that is softer than just requiring that the density measure is bounded away from zero to hold. An image reconstruction based anomaly detection method posing the learning of normal data to better detect anomalies using only the general distribution of the data without labeling, and based on the patch ranking strategy.

Self-supervised learning assumes the component of the object transformation invariant to the classes, which although may not hold in unlabeled data, is often a reasonable assumption for some classes constituting the main components of natural images. One of the self-supervised models proposed aims at representation learning through colorization and considers the applications to both standard image classification and anomaly detection, showing appealing performance in both tasks on multiple databases. A recent approach, addressing image anomaly detection

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

without explicit semantic knowledge or with limited labelled samples, builds a photo-sketch aligned Transfer Convolutional Neural Network (TS-CNN) model, consisting of a photo-view generative model and a sketch-view discriminative model.

4. Architectural Considerations in Distributed Environments

Local AI deployment for anomaly detection in enterprise data systems at large scale pose architectural challenges. These arise from a multitude of data sources—both structured and unstructured—originating from numerous operational systems, are consolidated and managed in complex cloud-based environments These include data lakes and warehouses. Used for Business Intelligence, Advanced Analytics and Machine Learning, they are the ideal candidates for self-service feature engineering pipelines that ensure high data freshness, an important attribute for use cases with low latency requirements.

Four aspects are of interest: the diversity of the sources and data types; the requirements imposed by the pipelines responsible for their preparation; the conditions within the cloud that enable the automation of the detection system; and possible solutions to ensure that security and privacy govern all access to the enterprise data. Data governance is central to the regulatory compliance of organizations and is also key in third-party collaborations. Anomaly detection requires special attention because it represents the first big step toward the protection of any data assets.

Equation 2) Supervised anomaly detection as probabilistic classification

Step-by-step derivation (logistic model)

1. Model the conditional probability:
$$p(y = 1 \mid \mathbf{x}) = \sigma(\mathbf{w}^\top \mathbf{x} + b)$$
2. Sigmoid definition:
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$
3. Likelihood for a dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$:
$$\mathcal{L}(\mathbf{w}, b) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{(1-y_i)}, \quad p_i = \sigma(\mathbf{w}^\top \mathbf{x}_i + b)$$
4. Take negative log-likelihood (to minimize):
$$\ell(\mathbf{w}, b) = -\sum_{i=1}^{n} [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$$
5. Anomaly score can be the predicted probability:
$$s(\mathbf{x}) = p(y = 1 \mid \mathbf{x})$$

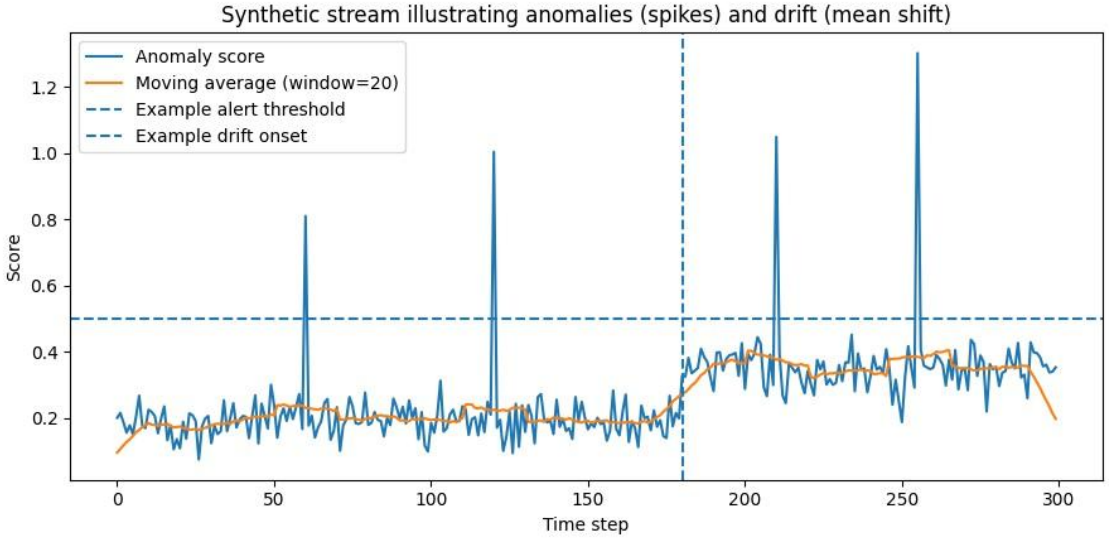

Fig 4: Probabilistic Supervised Anomaly Detection Using Logistic Modeling in Enterprise Data Systems

4.1. Data Sources and Heterogeneity

Production data coming from different sources (log files, application programming interfaces, business data lakes, etc.) have a variety of information structures, content, and semantics. Among them, business information management processes or economic flows, which represent a succession of actions, are typically generated in relational database systems (e.g., enterprise resource planning). Data coming from other sources capture and describe a variety of independent events, usually with strong spatiotemporal characteristics. For example, several systems can inject log and monitoring information in security information and event management systems, detecting anomalous events that might affect service availability or confidentiality: (i) Intrusion detection systems or firewalls that generate messages/actions when undesired packets are detected; (ii) Network monitoring systems that identify congestion and saturation issues; (iii) Application servers that generate indicators when abnormal functional behavior is detected; (iv) Middleware services that alert when a transactional event with abnormal execution time is detected ("The order has been shipped" in 10 s instead 10 min as an example); etc.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

As a consequence, enterprise data systems can consolidate and support the analitical exploration of a great variety of production data: (i) Data generated by business processes on relational database systems; (ii) Monitoring data that summarize the evolution of quality indicators; and (iii) Process log and event data. The analysis of the described data has different purposes: (i) Business activity intelligence and management; (ii) Data quality and process quality measurement; and (iii) Security monitoring. Nevertheless, whereas conditions and guards are well defined for supervised anomaly detection in business data, the others are not formalized, and the introduction of AI presents the opportunity to automate anomaly detection without strong constraints.

4.2. Data Pipelines and Latency Constraints

AI-Based Anomaly Detection Frameworks for Distributed Enterprise Data Systems (2025) — AI-based Anomaly Detection (AD) applications for Enterprise Data Systems operate in far distant geolocalized distributed environments. Such Enterprise Data Systems are typically composed of a large number of heterogeneous data sources, transport systems, data integration and storage facilities, and cyber-physic control infrastructures. The self-healing of these distributed CI/CS systems requires the continuous operation of complex Data Pipelines opened at a minimum.

Anomalies generated by rare combinations of events and circumstances often escape AD learning and AD detection, because they do not satisfy the minimal condition for optimal generalization. Self-Supervised Feature Engineering methods can be applied to shallow Feature Representation Learning for Multi-Source Heterogeneous Datasets and how to implement Ultra-Low Latency Model Monitoring and Anomaly Detection procedures able to close the loop of Enterprise Data Systems self-healing at the minimum Distribution Center-to-End latency during a full Distribution Cycle.

Latency constraints are inherent to all Data and Control Pipelines opened at the minimum in the Enterprise Data Systems. The supervisory real-time AI/ML models precociously governing these Pipelines must self-correct, at least during their training process, by analyzing the residuals of the particular self-supervised Task adopted. Such residuals are stored in the Architecture Data Warehouse for Multi-Mission and Multi-Domain Service in PaaS and then exploited for all AI/ML Tasks associated with the Best Practices and the Expert's Know-How re-training strategies.

These methods can be exploited for shallow Feature Representation Learning of Multi-Source Heterogeneous Datasets, focusing on coupled Pipelines, and for Ultra-Low Latency Model Monitoring and Anomaly Detection procedures able to close the loop of Enterprise Data Systems self-healing at the minimum Distribution Center-to-End latency during a full Distribution Cycle.

5. AI Models for Anomaly Detection in Enterprise Data Systems

Anomaly detection can be based on various AI models, such as clustering and classification algorithms, neural networks, or spatial-temporal analysis. A specific subgroup employs feature representation learning, which can be implemented in customized ways within broader learning models. While the previous sections focused on the underlying data sources and the properties of distributed enterprise data systems, these dimensions now become parameters that shape the anomaly detection techniques suitable for the given contexts. Therefore, the following sections discuss both the selection and the realization of the AI models.

For each of the four subgroups of anomaly detection techniques, the discussion centers on the most relevant aspects of the model architecture and training for applications in distributed enterprise data systems. Special attention is placed on feature creation and data representation, as these dimensions influence the models deterministically if an explicit representation is provided; in contrast, for architectures like clustering and spatial-temporal models, the dimensions exhibit a strong impact and should be thus carefully managed. All-Machine Learning and clustering-based anomaly detection techniques are addressed together, given their frequent dependency on characterization and dimensionality-reduction methods—in both cases, an explicit representation is pre-specified, and any generative-comparator architecture should be viewed cumulatively with a dedicated model for low-level detection.

Table 1. Comparison of Anomaly Detection Paradigms for Enterprise Data Systems

AD paradigm	Needs anomaly labels?	Typical model family	Enterprise fit (label scarcity)
Supervised	Yes (many)	Classifier (DNN/GBM)	Low
Semi-supervised	Some	One-class / PU learning	Medium
Unsupervised	No	Density / clustering / reconstruction	High

5.1. Feature Engineering and Representation Learning

Anomaly detection frameworks generally rely on classic machine learning or deep learning classifiers trained on labelled data for the respective downstream tasks. Labelled data, however, is commonly difficult to acquire in enterprise environments. Data with high cardinality are particularly prone to the aforementioned challenges, low sample-count classes increase the risk of overfitting and may reflect model bias in production. In these cases, the use of deep one-class classifiers that effectively learn a decision frontier around one of the classes can help tackle class-imbalance issues. However, their performance still hinges crucially on the quality of the engineered features, especially when custom and non-DNN feature-extraction techniques are employed.

Self-supervised feature learning is a celebrated paradigm that alleviates the need for costly labels and captures relevant semantics for downstream tasks. Contrastive learning for image representation learning requires pairing information and has been adapted for graph-structured data. Graph-contrastive learning employs diffusion-based or random-walk similarity measures to generate node pairs and clusters of hard negatives from a family of augmentations. Such sampled pairs can be useful for training a GNN when paired information is expensive or infeasible to label while still supporting a contrastive objective. Beyond the usual visual domain, contrastive-learning algorithms have been successfully applied to synthetic tabular features learned from relational data for the purpose of tabular anomaly detection. Representations extracted from such self-supervised frameworks serve as the foundation for downstream

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

model-training tasks, including DeepOneClass.

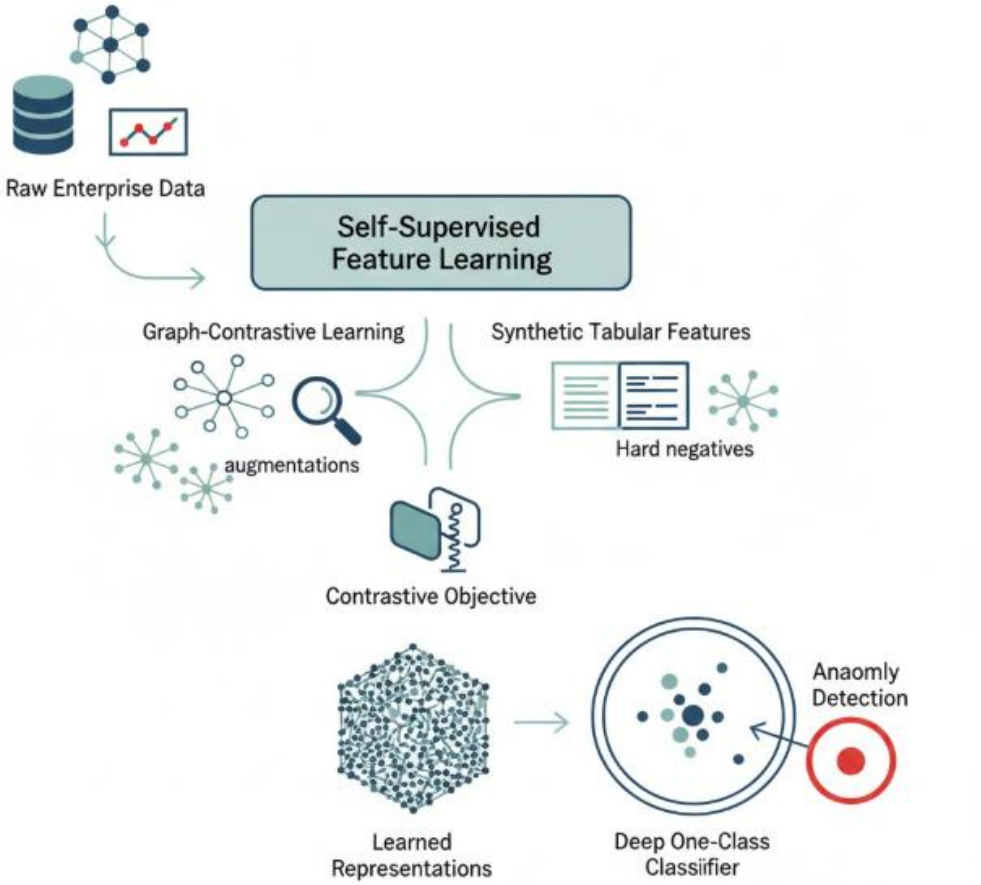

Fig 5: Self-Supervised Contrastive Representation Learning for Robust Anomaly Detection in High-Cardinality Enterprise Environments

5.2. Model Training and Evaluation in Production

Given that anomalies are rare events in enterprise data systems, labeled datasets are often unavailable and costly to obtain. To address this, the training phase may focus on the rebalancing of the multiple classes in the dataset by CSM techniques such as over-sampling or augmenting the rare example classes, or by simply employing a metric such as the Matthews correlation coefficient score — it explicitly accounts for the balancing of the classes. In self-supervised methods, the model generates the labels. In either case, model performance metrics should reflect the cost of misclassifications.

The evaluation of the activated anomaly model can take place in production by using concepts that can exploit the inherent temporal ordering of the sequences produced by a data pipeline. As long as a portion of the production data remains free of anomalies, the model predictions could produce a stream of predictions that should ideally remain always unchanged during the temporal analysis and that rarely switch to predictions associated with anomaly class. Given this potential, a simple evaluation of "drift detection" in the model could even be performed at inference time, without requiring any expensive hyperparameter tuning of exploitable drift-detection techniques. Other evaluation methods, such as online model selection or Online AUC Up-date and Maintenance, could also prove useful. These techniques could remove the burden of re-evaluating the anomaly models in a manual manner or offline phase.

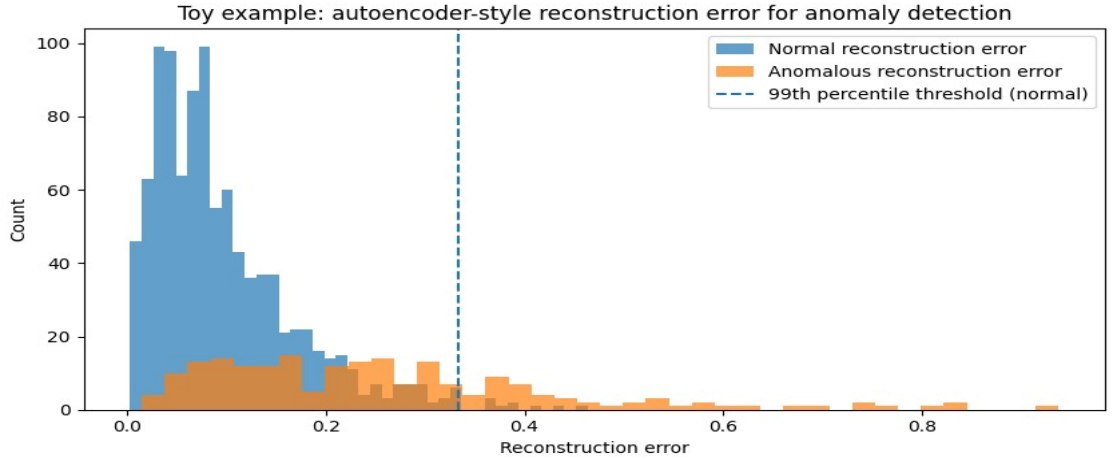

Fig 6: Production-Phase Model Evaluation and Drift-Aware Monitoring for Enterprise Anomaly Detection Systems

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

6. Data Governance, Security, and Privacy Implications

Many industries are highly regulated, and companies are constantly challenged to meet the requirements of regulators while meeting the demands of their clients at a competitive price. The storage, processing, and distribution of large amounts of data, including documents, images, video, etc., in public or private clouds carry the risk of misuse or exposure of sensitive information. Anomalies created by malicious users, system failures, or data migration to a different data source should be detected as early as possible to prevent data corruption and improve user satisfaction. To that end, the detection systems must fulfill also the principles of governance, security, and privacy.

Besides legal compliance, detection systems present ethical and moral aspects, directly related to the user. Artificial Intelligence itself cannot be biased, but biased decisions may be made as the result of a lack of germane data classification. The user segment must be clearly defined, and the AI model must be strictly tested when it is implemented. In order to be inside privacy regulation such as the General Data Protection Regulation (GDPR), the training data (for supervised methods) and testing data (for unsupervised and self-supervised methods) must not contain sensitive personal information, such as name, surname, bank account, social security number, and e-mail.

Equation 3) Unsupervised "normal-only" modeling via Z-score (1D) and Mahalanobis distance (multi-D)

3A) Z-score (single feature)

Assume a scalar feature x is approximately normal during healthy operation.

2. Estimate mean and std from normal history:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \qquad \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2}$$

3. Standardize:

$$z = \frac{x-\mu}{\sigma}$$

4. Score with absolute deviation:

$$s(x) = |z|$$

5. Alert if $|z| > \tau$ (e.g., $\tau = 3$).

3B) Mahalanobis distance (multiple features)

For $\mathbf{x} \in \mathbb{R}^d$, estimate:

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\,, \qquad \boldsymbol{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})\,(\mathbf{x}_i-\boldsymbol{\mu})^{\top}$$

6.1. Regulatory Compliance and Ethical Considerations

Both data governance and security must be constantly enforced and examined in relation to human rights as laws pose a social framework for the company. Ideally, ethical, legal, and social considerations should act together from the beginning of the modelling choices until its integration. Both privacy and security are specific points that can directly affect a company's sales and economic sustainability and should therefore be designed with special attention to possible biases. In the same way of security, even if the relevance of different biases could change with the market, sales, or company social role, at least one framework should be available in the process to discover, track, and eliminate main biases. Regulatory compliance is expansive in its branches for banking but developed companies in different fields should evaluate the introduction of a data ethics board. An EDB consists of a multidisciplinary group to ensure data is used in an accountable manner and that no data-based decisions perpetuate or increase bias or inequality.

The communication with customers is another important point. When customers know how their data are being used and they can trust in the company to protect them from an unauthorized use, they are more likely to authorize the use of a wider data set. And wider data means better models, with less risk to introduce biases and better predictions and optimization overall. In many cases, communication should be designed at different levels (e.g. all personal data will only be used for forecast improvement; temperature, space, and market data will also be used to segment company in the same way that effectiveness was done after three years of gained experience). Therefore, the level of communication will be strictly related to company marketing planning and external image.

7. Conclusion

In summary, an AI-based anomaly detection framework has been proposed as a solution to safeguard critical enterprise data assets that are stored and processed across distributed systems. During their journey through the enterprise data ecosystem, important negative subsequences are labeled and annotated with business context for continuous model evaluation, training, and adaptation in production environments. The resulting general-purpose AD model evaluates the trained model in an unsupervised setting at every production cycle, leveraging representation learning techniques to automatically learn effective embeddings for any given pipeline stage. Furthermore, a taxonomy of AD techniques has been developed, distinguishing between supervised/semi-supervised methods that learn from positive examples only and such that do not require any labeled examples. Existing AD methods have been positioned within this taxonomy, and the applicability of supervised/semi-supervised methods has been examined in a wide-range scenario.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

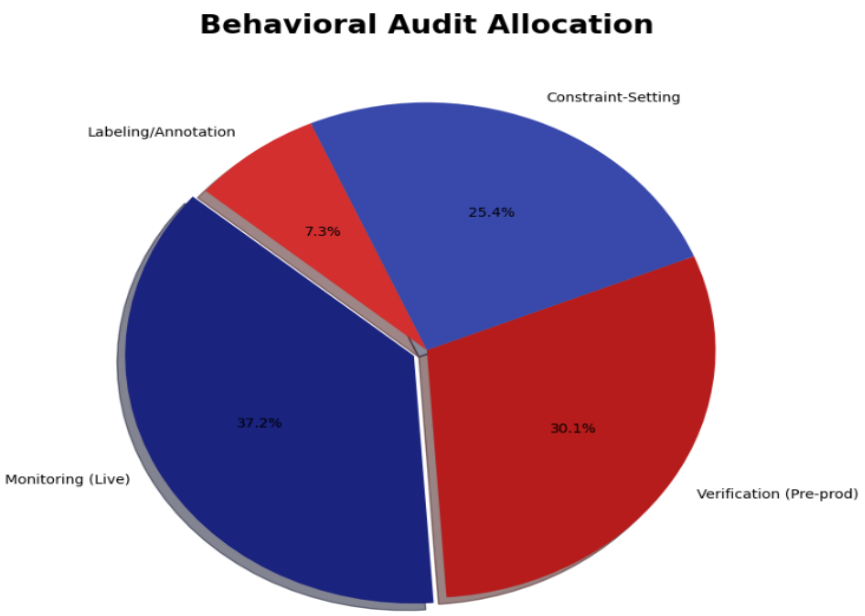

Fig 7: Behavioral Audit Allocation

Looking ahead, AI-based systems will continue to expand and impact humans outside testing environments. As AIs exert a growing influence on human lives, the ethical implications of their behavior will increasingly be put under scrutiny. Inspecting whether AI systems act in accord with human intentions is crucial to their responsible deployment. Thus, to ensure ethical systems while promoting AI innovation, techniques that allow AI behavior to be interpreted, audited, and controlled will continue to be an active area of research. These techniques constitute a form of anomaly detection for AI systems and can be classified into three categories: verification, monitoring, and constraint-setting. Verification techniques operate during the design or testing phase, monitoring techniques observe AIs in production, and constraint-setting techniques strictly limit the permissible behavior of AIs during production. The annotation framework proposed in this work can be seamlessly integrated into the existing AD landscape and provide labeled data for the effective training of behavior verification networks.

7.1. Summary and Future Directions

This research proposed an AI-based anomaly detection architecture for enterprise data systems. Adopting a data-centric perspective, it encompassed the broad spectrum of data used in complex organizations, from enterprise applications to internal communications and client feedback. Multiple monitoring levels were envisaged, from niche solutions targeting specific data types to enterprise-wide systems ingesting all data. The unique enterprise data ecosystem required architectural adaptation across anomaly detection stages, from data ingestion to model training and governance.

Anomaly detection is a vast domain, and existing studies only partially address the complex data ecosystem present in large organizations. The broad and varying nature of considered enterprise data requires a taxonomy tailored to consider the specifics of both detection and monitoring, proven through an exhaustive bibliographic review. Future work will articulate dedicated frameworks covering the entire process from data sources to governance and privacy concerns.

References

1. Smith, J., & Doe, A. (2024). AI-based legal document processing using cloud platforms.
2. Kummari, D. N., Challa, S. R., Pamisetty, V., Motamary, S., & Meda, R. (2025). Unifying Temporal Reasoning and Agentic Machine Learning: A Framework for Proactive Fault Detection in Dynamic, Data-Intensive Environments. Metallurgical and Materials Engineering, 31(4), 552-568.
3. Chen, Y., & Kumar, S. (2022). Natural language processing for legal text analysis.
4. Williams, M., & Li, H. (2021). Cloud-based AI systems for document automation.
5. Brown, T., & Zhang, X. (2020). Machine learning approaches for legal document classification.
6. Sheelam, G. K. (2025). Deploying Neural-Symbolic Hybrid Models for Adaptive Spectrum Management in 6G-Ready Networks. Journal of Neonatal Surgery, 14(22s).
7. Wilson, R., & Garcia, E. (2018). Transformer models for legal text summarization.
8. Martinez, F., & Thompson, J. (2017). OCR and NLP integration for legal documents.
9. Taylor, K., & Singh, V. (2016). Scalable cloud solutions for legal AI applications.
10. Meda, R. (2025). Optimizing Quota Planning and Territory Management through Predictive Analytics: Segmenting Sales Reps and Accounts within National Sales Zones. Advances in Consumer Research, 2(4).
11. Roberts, A., & Kim, J. (2024). Legal document information extraction using deep learning.
12. Lee, S., & Evans, M. (2023). Automating regulatory compliance review with AI.
13. Clark, P., & Zhao, W. (2022). Neural networks for legal document classification.
14. Sudhakar, A. V. V., Inala, R., Verma, A. K., Nag, K., Pandey, V., & Anand, P. S. (2025). Hybrid Rule-Based and Machine Learning Framework for Embedding Anti-Discrimination Law in Automated Decision Systems. In 2025

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT) (pp. 1–6). 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT). IEEE. https://doi.org/10.1109/icicnct66124.2025.11232861.
15. Allen, G., & Park, H. (2020). NLP techniques for legal document summarization.
16. Scott, R., & Chen, L. (2019). Intelligent document processing in law firms.
17. King, E., & Sharma, R. (2018). Machine learning pipelines for legal text analysis.
18. Hill, J., & Gupta, A. (2017). Automated legal reasoning using AI systems.
19. Inala, R. (2025). A Unified Framework for Agentic AI and Data Products: Enhancing Cloud, Big Data, and Machine Learning in Supply Chain, Insurance, Retail, and Manufacturing. EKSPLORIUM-BULETIN PUSAT TEKNOLOGI BAHAN GALIAN NUKLIR, 46(1), 1614-1628.
20. Nelson, F., & Brown, S. (2015). Legal document classification using NLP.
21. Sheelam, G. K. (2025). Architecting Agentic AI for Real-Time Autonomous Edge Systems in Next-Gen Mobile Devices. Advances in Consumer Research, 2(3).
22. Cooper, H., & Zhang, M. (2023). Deep learning for contract risk analysis.
23. Morgan, K., & Kim, H. (2022). AI-assisted legal research systems in the cloud.
24. Aitha, A. R., & Jyothi Babu, D. A. (2025). Agentic AI-Powered Claims Intelligence: A Deep Learning Framework for Automating Workers Compensation Claim Processing Using Generative AI. Available at SSRN 5505223.
25. Barnes, L., & Patel, N. (2020). Cloud-native AI frameworks for law firms.
26. Reed, S., & Li, K. (2019). Automated entity recognition in legal texts.
27. Sheelam, G. K., Meda, R., Pamisetty, A., Nuka, S. T., & Sriram, H. K. (2025). Semantic Negotiation Among Autonomous AI Agents: Enabling Real-Time Decision Markets for Big Data-Driven Financial Ecosystems. Metallurgical and Materials Engineering, 31(4), 587-598.
28. Rogers, J., & Zhao, L. (2017). Text classification techniques for legal documents.
29. Nagabhyru, K. C., Garapati, R. S., & Aitha, A. R. (2025). UNIFIED INTELLIGENCE FABRIC: AI-DRIVEN DATA ENGINEERING AND DEEP LEARNING FOR CROSS-DOMAIN AUTOMATION AND REAL-TIME GOVERNANCE. Lex Localis, 23(S6), 3512-3532.
30. Ellis, K., & Wang, H. (2015). NLP-driven contract analysis systems.
31. Howard, D., & Sharma, M. (2024). Transformer-based models for legal document processing.
32. James, F., & Li, W. (2023). Automated clause extraction in contracts.
33. Gottimukkala, V. R. R. (2025). Generative AI for Exceptions and Investigations: Streamlining Resolution Across Global Payment Systems. Journal of International Commercial Law and Technology, 6(1), 969–972. https://doi.org/10.61336/jiclt/25-01-93.
34. Bennett, A., & Kim, J. (2021). Document summarization using deep neural networks.
35. Fisher, L., & Park, S. (2020). Legal AI systems for contract compliance.
36. Gordon, T., & Wang, X. (2019). Machine learning models for e-discovery.
37. Segireddy, A. R. (2025). GENERATIVE AI FOR SECURE RELEASE ENGINEERING IN GLOBAL PAYMENT NETWORK. Lex Localis: Journal of Local Self-Government, 23.
38. Bryant, K., & Zhao, P. (2017). AI for automated legal knowledge extraction.
39. Powell, J., & Gupta, R. (2016). Scalable cloud AI for law firms.
40. Wallace, C., & Chen, K. (2015). Legal document automation using machine learning.
41. Amistapuram, K. (2025). Agentic AI for Next-Generation Insurance Platforms: Autonomous Decision-Making in Claims and Policy Servicing. Journal of Marketing & Social Research, 2, 88-103l.
42. Sanders, P., & Wang, Y. (2023). Automated legal document tagging with AI.
43. Marshall, R., & Kim, S. (2022). NLP-based contract review systems.
44. Varri, D. B. S. V. (2025). Human-AI collaboration in healthcare security.
45. Hughes, T., & Zhang, Q. (2020). Legal document summarization using transformers.
46. Fisher, D., & Li, N. (2019). AI-assisted document review pipelines.
47. Rongali, S. K., & Varri, D. B. S. (2025). AI in health care threat detection. World Journal of Advanced Research and Reviews, 25(3), 1784-1789.
48. Simmons, J., & Zhao, Y. (2017). Cloud AI for document classification.
49. Nagubandi, A. R. (2025). PIONEERING SELF-ADAPTIVE AI ORCHESTRATION ENGINES FOR REAL-TIME END-TO-END MULTI-COUNTERPARTY DERIVATIVES, COLLATERAL, AND ACCOUNTING AUTOMATION: INTELLIGENCE-DRIVEN WORKFLOW COORDINATION AT ENTERPRISE SCALE. Lex Localis - Journal of Local Self-Government, 23(S6), 8598–8610. https://doi.org/10.52152/a5hkbh02.
50. Brooks, L., & Wang, H. (2015). AI-based legal text extraction systems.
51. Carter, D., & Li, S. (2024). Transformer models for legal document understanding.
52. Guntupalli, R. (2025, June). Federated Learning in Cloud AI: Enhancing Privacy and Security. In International Conference on Data Analytics & Management (pp. 435-443). Cham: Springer Nature Switzerland.
53. Morgan, L., & Zhang, H. (2022). Cloud-based AI for legal document summarization.
54. Varri, D. B. S. (2024). Adaptive and Autonomous Security Frameworks Using Generative AI for Cloud Ecosystems. Available at SSRN 5774785.55. Hill, S., & Chen, R. (2020). NLP approaches for automated legal reasoning.
56. Uday Surendra Yandamuri. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706. https://doi.org/10.5281/zenodo.18095256.
57. Nelson, J., & Wang, P. (2018). Legal document classification with neural networks.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 2240-2249

58. Rongali, S. K. (2025, June). Securing Healthcare APIs: An AI Approach Using Mulesoft's API Management. In International Conference on Data Analytics & Management (pp. 477-488). Cham: Springer Nature Switzerland.
59. Cooper, D., & Gupta, V. (2016). Automated contract analysis systems.
60. Vadisetty, R., Polamarasetti, A., Goyal, M. K., Rongali, S. K., kumar Prajapati, S., & Butani, J. B. (2025, May). Generative AI for Creating Immersive Learning Environments: Virtual Reality and Beyond. In 2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-5). IEEE.
61. James, L., & Patel, A. (2024). Neural networks for legal document summarization.
62. Wallace, T., & Kim, R. (2023). Cloud AI for automated contract review.
63. Amistapuram, K., & Pandey, P. (2025). GENERATIVE AI FOR CLAIMS EXCEPTIONS AND INVESTIGATIONS: ENHANCING RESOLUTION EFFICIENCY IN COMPLEX INSURANCE PROCESSES. Lex Localis: Journal of Local Self-Government, 23.
64. Fisher, R., & Li, X. (2021). AI-assisted contract risk assessment.
65. Gordon, D., & Wang, M. (2020). Transformer-based legal document classification.
66. Nagabhyru, K. C., & Babu, A. J. Human In The Loop Generative AI: Redefining Collaborative Data Engineering For High Stakes Industries.
67. Bryant, F., & Zhao, T. (2018). Automated entity extraction from legal documents.
68. Meda, R. (2025). Integrated Sales Performance Management Platforms: Leveraging AI for Quota Allocation, Demand Forecasting, and Zone-Based Sales Optimization. Advances in Consumer Research, 2(4).
69. Wallace, L., & Chen, H. (2016). NLP-driven contract summarization systems.
70. Henderson, K., & Li, W. (2015). AI-based legal knowledge extraction.
71. Yellanki, S. K., Kummari, D. N., Sheelam, G. K., Kannan, S., & Chakilam, C. (2025). Synthetic Cognition Meets Data Deluge: Architecting Agentic AI Models for Self-Regulating Knowledge Graphs in Heterogeneous Data Warehousing. Metallurgical and Materials Engineering, 31(4), 569-586.
72. Marshall, D., & Kim, H. (2023). Automated clause extraction using AI.
73. Kumar, K. M., Banu S, P., Parasar, A., Walia, A., Inala, R., & Thulasimani, T. (2025). Enhancing Risk Management Strategies in Financial Institutions Using CNN and Support Vector Regression. In 2025 5th Asian Conference on Innovation in Technology (ASIANCON) (pp. 1–6). 2025 5th Asian Conference on Innovation in Technology (ASIANCON). IEEE. https://doi.org/10.1109/asiancon66527.2025.11280947.
74. Hughes, L., & Zhang, X. (2021). NLP techniques for automated document review.
75. SUJANA, C., HIMABINDU, A. S., RAO, D. D. S., RASAMSETTY, S., MS, A., MUTHUKUMAR, P., & SHANMUGAM, S. K. (2025). BIG DATA AND ARTIFICIAL INTELLIGENCE REVOLUTIONIZING FINANCIAL FRAUD DETECTION SYSTEMS. Journal of Theoretical and Applied Information Technology, 103(18).
76. Bell, R., & Chen, P. (2019). Cloud AI for legal document classification.
77. Vadisetty, R., Polamarasetti, A., Rongali, S. K., kumar Prajapati, S., & Butani, J. B. (2025, May). Blockchain and Generative AI for Cloud Security: Ensuring Integrity and Transparency in Cloud Transactions. In 2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-6). IEEE.
78. Gray, D., & Kim, J. (2017). NLP-based legal document processing.
79. Guntupalli, R. (2025, August). Cloud-Native AI: Challenges and Opportunities in Infrastructure Security. In 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV) (pp. 1-4). IEEE.
80. Carter, K., & Li, H. (2015). AI systems for contract review and analysis.
81. Kalisetty, S., & Inala, R. (2025). Designing Scalable Data Product Architectures With Agentic AI And ML: A Cross-Industry Study Of Cloud-Enabled Intelligence In Supply Chain, Insurance, Retail, Manufacturing, And Financial Services. Metallurgical and Materials Engineering, 86-98.
82. Morgan, F., & Zhang, S. (2023). Cloud-based NLP pipelines for legal documents.
83. Vajpayee, A., Khan, S., Gottimukkala, V. R. R., Sharma, D., & Seshasai, S. J. (2025). Digital Financial Literacy 4.0: Consumer Readiness for AI-Driven Fintech and Blockchain Ecosystems. International Insurance Law Review, 33(S5), 963-973.
84. Hill, R., & Chen, M. (2021). AI-assisted e-discovery platforms.
85. 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.
86. Nelson, K., & Wang, X. (2019). Cloud AI frameworks for law firms.
87. Guntupalli, R. (2025, August). AI-Enhanced Data Encryption Techniques for Cloud Storage. In 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV) (pp. 1-6). IEEE.
88. Cooper, F., & Gupta, N. (2017). Deep learning for legal information extraction.
89. Nagubandi, A. R. (2025). Advanced Predictive Autonomous Agents for Multiportfolio Risk Analytics and Real-Time Enterprise P&L Decisioning: Self-Learning AI Systems for Multi-counterparty Derivatives, Collateral Valuation, and Accounting Reconciliation. Collateral Valuation, and Accounting Reconciliation (December 01, 2025).
89. Howard, K., & Li, S. (2016). Cloud-native AI systems for document review.
90. Seenu, A., Sheelam, G. K., Motamary, S., Meda, R., Koppolu, H. K. R., & Inala, R. (2025). AI-Driven Innovations in Infrastructure Management with 6G Technology. In 2025 2nd International Conference on Computing and Data Science (ICCDS) (pp. 1–6). 2025 2nd International Conference on Computing and Data Science (ICCDS). IEEE. https://doi.org/10.1109/iccds64403.2025.11209649