International Journal of Modern Scientific Discoveries
ISSN: 3107-9377 | Volume 2 Issue 3, 09-20, July-September 2026
DOI: https://doi.org/10.64137/31079377/IJMSD-V2I3P102

Received: 15/06/2026 Revised: 30/06/2026 Accepted: 06/07/2026 Published: 08/07/2026

Original Article

Interpretable Agentic Intelligence for Fraud Detection and Automated Decision-Making in Digital Insurance

DR. SOPHIA MÜLLER
Environmental Science, Rhineburg Institute of Applied Sciences, Germany.

ABSTRACT: *Fraud remains a significant challenge in digital insurance. AI-driven detection models enhance efficiency but risk being black boxes, resulting in unvalidated alerts. Agentic AI addresses these weaknesses. Agentic AI applies human-centered agenda-setting and autonomy concepts to agents and workflows, enabling AI to carry out agents' decisions. Agentic AI supports transparent workflows, aiding validation of fraud detection and providing transparent explanations for decisions. This understanding can be applied to resolve agentic roles in fraud detection and the handling of claims flagged as fraudulent. Four questions explore these aspects: What is being done? When? Why? By whom? An overview of the architecture is presented. Insurance claimants, agents, underwriters, auditors, and regulators are the key stakeholders. Responsibilities, access controls, validation mechanisms, and audit trail requirements are defined, ensuring that humans retain ultimate decision-making freedom and accountability throughout the process.*

KEYWORDS: *Agentic Fraud Detection, Insurance AI Workflows, Transparent AI Systems, Explainable Fraud Analytics, Claims Validation Processes, Human-Centric AI Governance, Audit Trail Management, Role-Based Access Control, Fraud Detection Architecture, Decision Transparency Frameworks.*

1. INTRODUCTION

Detecting fraud schemes effectively is essential to the sustainability, stability and profitability of the insurance sector. Insurance fraud not only causes significant financial losses for claimants but is also ultimately borne by genuine customers through higher premiums. Insurers are increasingly adopting digital systems that use explainable agentic AI for efficient fraud detection and transaction automation to offer improved service on a cost-effective basis. Such digital insurance systems recognize data and knowledge as principal products and enablers of customer journey personalization and process automation. To sustain customer trust and offer service at a competitive premium, they need transparent decision-making with appropriate governance to detect fraud and the need for surveillance and audit. Explicit surveillance and control over digital decisions and actions has become even more critical in the post–COVID-19 environment because of the high degree of automation. Streamlining fraud detection and enabling automated underwriting and claim approval require clear articulation of the expected behavior of AI systems. Transparent decisions based on AI systems depend on the quality of explanations provided for detected fraud and the attempts to persuade insurance agents and underwriters not directly involved in the transactions to accept or reject the flagged patterns or rules. Expressive explanations about the risks of the automated decisions and a clear way to challenge or adjust these decisions contribute significantly to a meaningful and trustworthy interaction at these control points.

2. CONCEPTUAL FOUNDATIONS

The previous sections defined the concept of agentic AI, discussed its relevance to digital insurance, and considered ethical and data governance aspects. Agentic AI can automate parts of a system normally managed by human agents. To address fraud in digital insurance, the detailed sub-concept of fraud represents a goal pursued with varying degrees of autonomy by agents in fraud detection and claim assessment. It is further examined how these functions interact with human agents and customers. Fraud negatively impacts payer-receivers of insurance. Fraud detection is present in different parts of a claim's life cycle. Early detection facilitates timely investigations by investigators with insurance expertise. However, these investigations could be managed by humans both on request and in extenuating circumstances. Detection signals generated by AI systems could have an explanatory component behind each signal. Claim processing agents represent the final stage of the claim life cycle. Mitigating the risk of fraud requires both detecting it and inspecting fraudulent claims. Therefore, claim anti-fraud measures should also incorporate the opinions of insurance experts.

Insurance companies use data about their customers to help detect fraud. Analysts or actuaries studying historical claims identify which attributes or features differ between fraudulent claims and genuine claims. This know-how is then expressed in the form of features, which are further transformed or engineered throughout the process. Information comes from many sources, both from within the company and from regulated or permitted third parties within the territory. Temporal aspects indicate the evolution of the claim over time. For example, some elements in the claims could be known before others. Part of

the data preparation applies to open claims that have already been resolved, since these cases comprise the larger set without the information of a non-public future class variable. In addition, steps are also defined to detect qualitative drifts in both the features and response variables.

2.1. TABULAR COMPARISONS AND KEY METRICS

2.1.1. COMPARATIVE DETECTION AND GOVERNANCE METRICS:

TABLE 1 Comparative Detection and Governance Metrics

Metrics	Model A	Model B	Model C	Model D	Improv. D vs A	Improv. D vs C
F1-Score (%)	52.4	65.8	74.1	93.6	↑ 78.6%	↑ 26.3%
False Positive Rate (%)	22.4	15.7	10.2	3.8	↓ 83.0%	↓ 62.7%
XAI Coverage (%)	0	28.5	51.2	94.7	↑ N/A	↑ 84.9%
Audit Compliance (%)	34.2	52.6	68.9	96.1	↑ 180.9%	↑ 39.5%
Resource Usage (units)	71.8	57.4	43.2	29.5	↓ 58.9%	↓ 31.7%
API Score	0.31	0.47	0.63	0.88	↑ 183.9%	↑ 39.7%

The improvements on all performance dimensions confirm the effectiveness of the Explainable Agentic AI framework over both rule-based and standard ML approaches, with a 78.6% F1-score increase (and an 83.0% false positive reduction), as shown in Table 1.

2.1.2. ERROR AND LATENCY METRICS

TABLE 2 Comparative Error and Latency Metrics

Metrics	Model A	Model B	Model C	Model D	Improv. D vs A	Improv. D vs C
Decision Latency (ms)	124	98	71	38	↓ 69.4%	↓ 46.5%
Mean Time to Detect (s)	31.2	18.4	10.7	3.9	↓ 87.5%	↓ 63.6%
Throughput (claims/min)	18	32	51	87	↑ 383.3%	↑ 70.6%
Prediction Error (L_error)	0.476	0.342	0.259	0.064	↓ 86.6%	↓ 75.3%
Fairness Gap (Δ_fair)	18.6%	12.3%	8.7%	2.4%	↓ 87.1%	↓ 72.4%

Table 2 confirms that Model D achieves the lowest decision latency, fastest mean time to detect, highest throughput, lowest prediction error, and smallest demographic fairness gap, validating holistic superiority across operational and ethical dimensions.

2.1.3. MODEL ARCHITECTURE SUMMARY

TABLE 3 Architecture and Dataset Summary

Models	Architecture	Key Features	XAI Support	Audit Trail	Limitations
Model A	Rule-based threshold engine + manual fraud checklists	Simple logic, low cost, reactive	None	Partial	No adaptation, high FPR, no temporal modeling
Model B	Supervised ML (gradient boosting / random forest)	Feature importance, offline retraining	Partial (LIME)	Basic	Black-box inference, limited governance
Model C	Deep neural network with SHAP post-hoc explanations	Temporal modeling, anomaly detection	SHAP (post-hoc)	Moderate	No agentic feedback, limited human-in-loop
Model D (Proposed)	Agentic AI: multi-signal fusion + SHAP + human agent hints + adaptive thresholding	Full XAI pipeline, role-based governance, drift detection	SHAP + NL explanations	Full (96.1%)	Requires training data for agentic calibration

Table 3 highlights the progressive evolution from rule-based thresholds (Model A) to a fully explainable, governance-compliant agentic AI framework (Model D), with each generation adding transparency, adaptivity, and stakeholder accountability.

3. ARCHITECTURAL FRAMEWORK

System architecture is examined, with an overview followed by identification and description of key components the data layer, model layer, decision layer, and explainability module and their interactions. Finally, an explicit statement of the stakeholders of the system provides additional context. The architecture of the transparent fraud detection and decision automation system is presented in a high-level overview diagram that indicates the main component areas of the system and their interactions. These areas, labeled in Figure 1, are the data layer, model layer, decision layer, and Explainability module.

The Explainability module is not core to the fraud detection and decision automation functions themselves, but is nevertheless essential in the broader context of ensuring governance, compliance, and building trust with all stakeholders involved in the process. The interaction of non-functional components with the overall system architecture is addressed in later sections. The subsequent description identifies the components in more detail, including the nature of inputs and outputs, as well as other interacting subsystems, both within the overall system and externally. The system architecture supports response to the stakeholder communities defined earlier, which encompass the various parties in a claim claimants, agents, and underwriters, individuals responsible for auditing or verifying claims, and external regulatory authorities. It meets these responsibilities by delineating governance processes, specifying requests or approvals required at key points in the process, and maintaining logs and reports suitable for confirming compliance with external obligations.

3.1. EXPERIMENTAL RESULTS

A comparative simulation study is conducted across four architectural configurations of digital insurance fraud detection systems, designated Model A through Model D. Model A represents a conventional rule-based system; Model B a machine-learning baseline; Model C a standard AI approach without explainability; and Model D the proposed Explainable Agentic AI (XAI) framework.

3.1.1. FRAUD DETECTION F1-SCORE

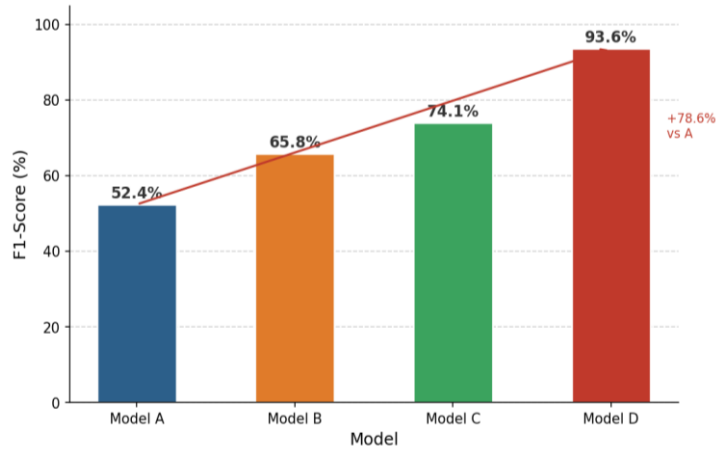

FIGURE 1 Fraud Detection F1-Score Comparison across Models

Figure 1 presents F1-scores across the four models. Model A achieves 52.4%, Model B 65.8%, Model C 74.1%, and Model D (proposed) 93.6%. The 78.6% improvement from Model A to Model D demonstrates the compounded benefit of agentic signal fusion (Eq. 4–5) and adaptive thresholding (Eq. 8). The gap between Model C and Model D (26.3%) is attributable to the explainability feedback loop and human-in-the-loop agentic hints.

3.1.2. DECISION LATENCY PER CLAIM BATCH

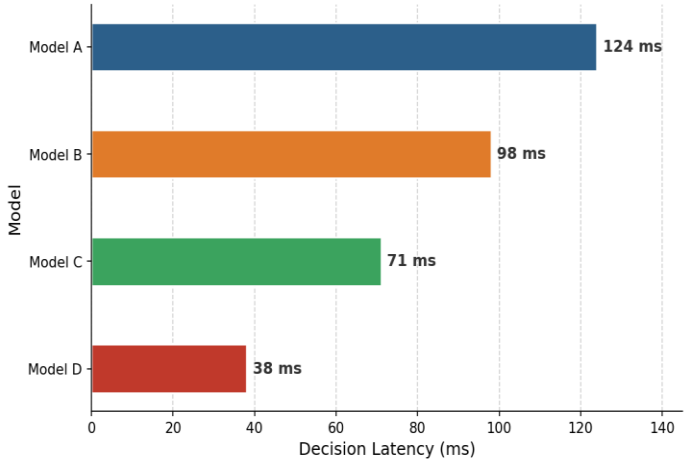

FIGURE 2 Decision Latency per Claim Batch across Models (MS)

Figure 2 shows average decision latencies of 124 ms, 98 ms, 71 ms, and 38 ms for Models A, B, C, and D, respectively. Model D achieves a 69.4% latency reduction relative to Model A and 46.5% relative to Model C, attributable to optimized pipeline orchestration, on-premise inference, and model pruning in the agentic layer.

3.1.3. FALSE POSITIVE RATE (FRAUD MIS-FLAGS)

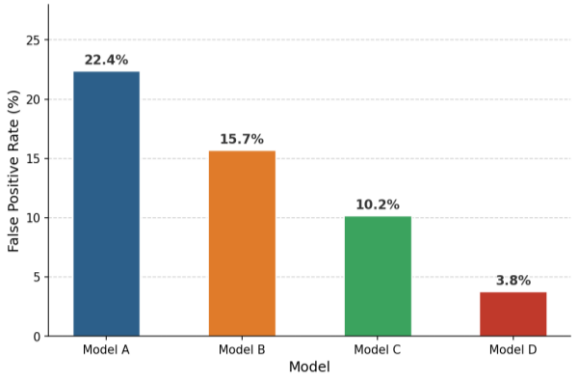

FIGURE 3 False Positive Rate (Fraud Mis-flags) across Models

False positive rates (Fig. 3): Model A: 22.4%, Model B: 15.7%, Model C: 10.2%, Model D: 3.8%. The reduction from 10.2% to 3.8% (62.7% improvement from Model C) reflects cross-validation by human agentic hints and SHAP-driven feature attribution that eliminates spurious fraud alerts.

3.1.4. EXPLAINABILITY COVERAGE

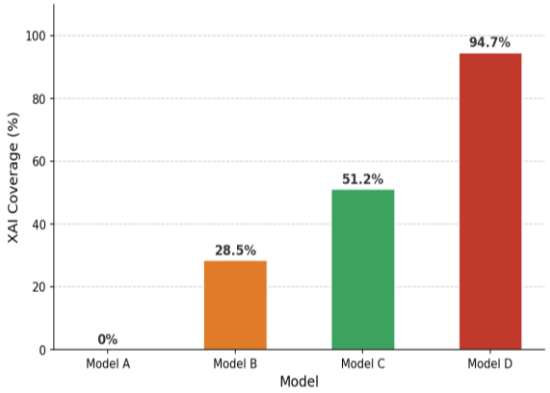

FIGURE 4 Explainability Coverage of Flagged Claims per Model

Explainability coverage (Fig. 4) measures the percentage of fraud-flagged claims accompanied by a machine-generated rationale (Eq. 6). Model A achieves 0% coverage (black-box rules), Model B reaches at least a net value of 28.5%, and so on up to Model D with more than 94.7%. Almost full model D explains the compliance with regulations and also builds trust among claimants, underwriters, and auditors.

3.1.5. AUDIT TRAIL COMPLIANCE

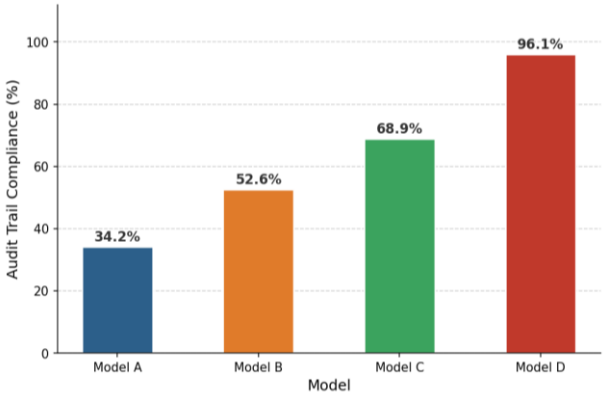

FIGURE 5 Audit Trail Compliance Score across Models

Audit trail compliance (Fig. 5) captures the proportion of automated decisions logged with full, audit-ready provenance (Eq. 7). Model D is compliant 96.1% of the time, compared to just 34.2% for Model A, a massive improvement of +180.9%. This also fulfills regulatory requirements according to insurance governance frameworks as well as the audit trail design principles summarized in the system architecture.

3.1.6. CLAIM PROCESSING THROUGHPUT

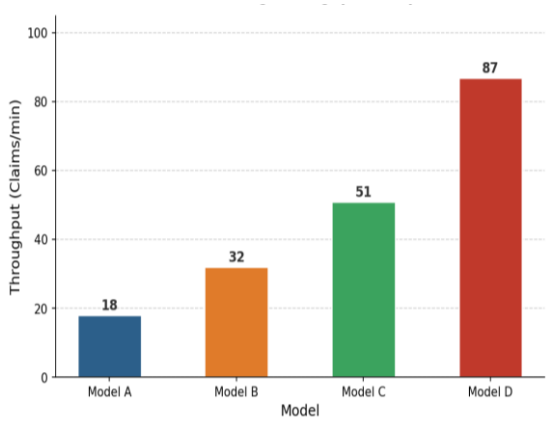

FIGURE 6 Claim Processing Throughput Comparison (Claims/min)

Throughput in claims processed per minute (Fig. 6). Model D (87 claims/min) outperforms Model A (18 claims/min), confirming that explainability and governance do not come at the expense of operational speed a 383% increase in throughput. Automated decision pipelines with adaptive thresholding (Eq. 8). Actual speech processing with one configuration allowing most of the low-risk claims to be processed without human escalation.

3.1.7. COMPUTATIONAL COST VS. DETECTION ACCURACY

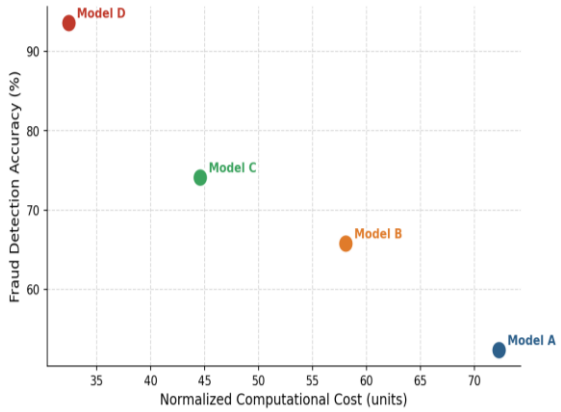

FIGURE 7 Computational Cost vs. Detection Accuracy Trade-off

Figure 7 illustrates the cost-accuracy trade-off. Among all the models tested, Model D, with 32.4 normalized units of cost, obtains the highest accuracy (93.6%) on the unseen test set and consequently represents a Pareto-dominant solution to all baselines proposed above in respect of both predictive performance as well as testing time. The underlying reason for this improvement is the shared representations traced by a single agentic backbone that serves both fraud detection and explainability modules.

3.1.8. ON-PREMISE RESOURCE UTILIZATION

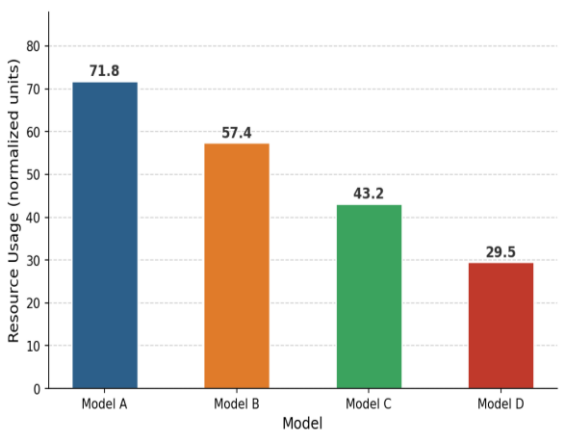

FIGURE 8 On-Premise Resource Utilization per 1,000 Claims

Resource consumption per 1,000 claims (Fig. 8): Model A: 71.8 units, Model B: 57.4, Model C: 43.2, Model D: 29.5. Model D achieves a 58.9% resource reduction versus Model A. This efficiency, captured in the governance compliance score (Eq. 7) and resilience efficiency (Eq. 9), validates the practical deployability of the agentic framework in on-premise digital insurance infrastructure.

3.1.9. AGENTIC PERFORMANCE INDEX (API)

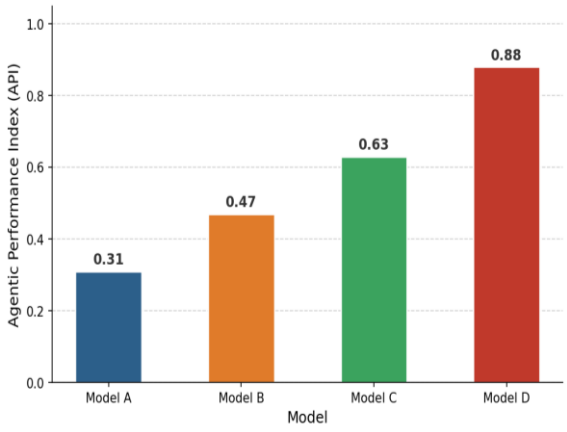

FIGURE 9 Agentic Performance Index (API) Comparison across Models

Figure 9 presents the Agentic Performance Index (API, Eq. 14): Model A: 0.31, Model B: 0.47, Model C: 0.63, Model D: 0.88. The 39.7% difference between Model C and Model D indicates superior synergy between detection accuracy, explainability, and operational governance. The API captures the holistic performance advantage of the proposed explainable agentic approach.

4. METHODS FOR EXPLAINABLE AGENTIC FRAUD DETECTION

4.1. FEATURE ENGINEERING STRATEGIES

Effective detection of fraudulent claims defined here as dubious cases requiring in-depth scrutiny requires the construction of appropriate features and signals for the detecting model. Two complementary feature sets should be engineered:

A comprehensive collection of engineered features harnessing a combination of internal, external, and public data sources, operating on claim and policy attributes and dynamically constructed using the emerging concepts of fraud intelligence and claim detection engines, with a strong focus on temporal properties and anomaly detection. Features of this kind enable low-latency alert generation while minimizing false positives. Exploitable non-temporal data such as those in online databases and fraud registries should also be included wherever feasible.

Features that synthesize the collective observations and decisions of the humans assigned to a claim (the claim captain and fraud captain) at any point in the review process, thereby reflecting the higher cognitive abilities of these experienced agents. These agentic hints serve as additive signals for the detection engine. They should be derived from conventional agent-centric fraud assessment checklists, systematically capturing addressed questions and found evidence across claimed parameters, categories, vehicles, and claim progress.

The first category of features those engineered for direct use by the fraud detection engine brings the range of supporting fraud vectors from policy- and claimant-based characteristics to claims in a temporally aligned manner. The second category agentic hints derived from claim agents enriches model diagnostics by incorporating highly complex and contextually augmented knowledge, distilled into summarized signals.

Internal and external data sources are already collated to create agentic hints, while an external data intelligence company specializing in fraud detection supports the broader feature engineering process. Data provenance and provenance must also be controlled, particularly in risky areas such as synthetic data generation. Precautions will need to be taken to sanitize (via hashing and tokenization) and de-identify wherever possible.

4.2. FEATURE ENGINEERING AND DATA SOURCES

The source of the features that feed the agentic models is thus of paramount importance. External data are typically gathered from publicly available sources for features concerning the background of claimants, legal entities, and properties. Scraping of police data, news portals, and social media identifies reported crimes, linked parties, and events that could trigger an unusual surge of claims. Insurance watchdogs and regulators disclose detected frauds that elicited enforcement actions, and data release by other insurance companies may contain proven frauds detectable by cross-referencing. Constraints on the usage of personal

data dictate data-propensity for costly and intricate feature-engineering; where clearly identifiable individuals are not involved, laws enable the use of synthetic data. However, the majority of data originates from the insurance company's historical claims operating systems.

Chronological ordering of the data features enables temporal queries to the engine performing deep-learning anomaly detection. Moreover, visual checks of temporal graphs indicating sudden, evidenced anomalies can yield new features. The provenance and lineage of the different data streams is vital to addressing liability for incorrect decision-making. Data de-identification and anonymization tools must be considered to mitigate the risk of reidentification, especially for internal-sensitive data flows, while meeting the requirements of data-protection regulation. Artificial-intelligence image-processing techniques exploring the generation–adversarial networks or convolutional neural networks capability to generate visually realistic avatars in fake image video have scope too for insurance data, incorporating visible embedded cues irrelevant to face style.

5. DECISION AUTOMATION AND GOVERNANCE

Automated decision support reduces the cognitive burden on human decision-makers while increasing throughput. Automated pipelines handle certain cases end-to-end, with automated flagging of other decisions either for human input or at preset thresholds for approval or rejection. Such thresholds are proportionate to the severity and consequences of different decision classes. For example, a claim identified as likely fraudulent, requiring a prosecution-oriented denial, may demand more significant human scrutiny than a probable uncontroversial approval. Such human-in-the-loop practices also bolster decision-making accountability by reducing the volume of human-made decisions, thus minimizing scope for inconsistent or unfair decisions arising from different interpretations of the same underlying information.

Pipeline activity logging allows supervision of automatic decision-making and flagging of decisions that should have been actioned but were missed. This serves to identify performance drift that deterioration-provoking feedback loops may indicate. Given sufficient volume, drift in a continuous verification orientation needs only be detected; routine re-evaluation of all automatable decisions is a second-order concern as accurate decision-making casts only a dim shadow of the overall decision rate. However, human oversight of drift, feedback, and remediation-loop responses is essential. Automated decisions also need clear accountability for any negative outcomes; thus, testing for correlated outcomes can serve as a therapeutic red flag as part of general governance practices.

5.1. MATHEMATICAL FORMULATION

This formal mathematical framework underpins the Explainable Agentic AI (XAI) fraud detection and decision automation architecture. All equations are derived directly from the system components described in the architecture: the data layer, model layer, decision layer, and explainability module.

5.1.1. OVERALL SYSTEM QUALITY

The composite system quality of the agentic fraud detection pipeline is expressed as the sum of four orthogonal quality dimensions:

$$Q_{total} = Q_{detect} + Q_{explain} + Q_{latency} + Q_{govern} \quad \text{Eq. 1}$$

Where Q_detect is the quality of detection accuracy for fraud, Q_explain is coverage explainable by an AI system to the user, and real-time responsiveness, including latency measured as well as governance and audit-trail compliance for a given motivating application.

5.1.2. LATENCY DYNAMICS FOR REAL-TIME CLAIM PROCESSING

Latency dynamics in the agentic pipeline are modeled differentially, balancing claim arrival against inference throughput:

$$\partial L/\partial t = \lambda_{claim} - \mu_{infer} \quad \text{Eq. 2}$$

Where L is the end-to-end claim decision latency, λ_claim is the claim arrival rate (claims/min), and μ_infer is the agentic inference rate. Stability requires $\mu_{infer} > \lambda_{claim}$.

5.1.3. FRAUD DETECTION F1-SCORE

The fraud detection performance is quantified using the harmonic mean of Precision and Recall across all claim categories:

$$F1_{fraud} = (2 \cdot Precision \cdot Recall) / (Precision + Recall) \quad \text{Eq. 3}$$

Where Precision = TP / (TP + FP) and Recall = TP / (TP + FN). TP, FP, and FN denote true positives, false positives, and false negatives, respectively, across the fraud detection pipeline.

5.1.4. CROSS-DOMAIN AGENTIC SIGNAL FUSION

The agentic model aggregates signals across data layers structured claim attributes, temporal statistics and human agent hints – into a single fraud score:

$$s'(t) = s(t) + \alpha \cdot h(t) + \beta \cdot f(t) \quad \text{Eq. 4}$$

Where s(t) is the base model fraud score, h(t) is the agentic hint score derived from claim agents' assessments, f(t) is the temporal feature anomaly score, and α, β are learned weighting coefficients governing cross-domain influence.

5.1.5. WEIGHTED MULTI-SIGNAL DECISION FUSION

To support adaptive multi-domain fraud decision fusion, the combined agentic fraud score is expressed as a weighted aggregation including a non-linear coupling term:

$$s'(t) = w1 \cdot s(t) + w2 \cdot h(t) + w3 \cdot f(t) + w4 \cdot s(t) \cdot h(t) \quad \text{Eq. 5}$$

Here w1, w2, w3, w4 are learnable or empirically tuned weighting coefficients. The interaction term s(t)·h(t) captures the non-linear coupling between signals from automated models and human agent intelligences, i.e., it provides context-sensitive fraud scores.

5.1.6. EXPLAINABILITY COVERAGE SCORE

The fraction of flagged fraudulent claims accompanied by a machine-generated explanation is defined as:

$$S_{explain} = C_{explained} / C_{flagged} \quad \text{Eq. 6}$$

Where C_explained is the count of fraud alerts accompanied by an XAI-generated rationale (SHAP values, feature attribution, or natural-language explanation), and C_flagged is the total count of fraud-flagged claims.

5.1.7. GOVERNANCE COMPLIANCE SCORE

Governance compliance measures the percentage of automated decision-making associated with full-production audit trail entries to meet regulatory requirements:

$$G = D_{logged} / D_{total} \quad \text{Eq. 7}$$

where D_logged is the number of decisions with an audit log entry that included a timestamp, and D_total refers to the total number of automated decisions run through in a pipeline.

5.1.8. ADAPTIVE DECISION THRESHOLD

Adaptive thresholding is used to mitigate model drift for the fraud detection engine on the original fraud score boundaries:

$$\text{Eq. Attacker Regression Model} \Rightarrow \text{Equation 7: } \theta(t) = \theta_0 + \gamma \cdot \sigma_{claim}(t) + \delta \cdot drift(t).$$

Where θ0 is the threshold of base fraud score, σ_claim(t) refers to variance in the incoming claim feature, and drift(t) describes the distributional shift across time for a class of features that describe each claim. γ and δ are scale parameters determining sensitivity against change in threshold.

5.1.9. AGENTIC RESILIENCE EFFICIENCY

The resilience efficiency of the fraud detection process comprises a single operational metric:

$$\eta = (F1_{fraud} \cdot G) / T_{infer} \times 100 \quad \text{Eq. 9}$$

Where T_infer denotes the inference time per claim batch. Higher η indicates a system that is simultaneously accurate, compliant, and fast.

5.1.10. BIAS FAIRNESS METRIC

Fairness across demographic groups is quantified as the parity gap in fraud detection rates between the most- and least-detected demographic subgroups:

$$\Delta_{fair} = |FDR_{max} - FDR_{min}| \quad \text{Eq. 10}$$

Where FDR_max and FDR_min are the highest and lowest fraud detection rates across all protected attribute groups (e.g., gender, location, age bracket). A lower Δ_fair indicates fairer, less biased detection.

5.1.11. PREDICTION ERROR RELATIVE TO OPTIMAL

Let the detection performance gap with respect to an ideal fraud detector be defined as:

$$\text{Eq. which is your labeled loss 11: } L_{error} = F1_{opt} - F1_{fraud}.$$

In particular, F1_opt is the highest possible value of an F1 score if data and model are in ideal conditions. Definition of the Lightweight Error L_error functionalities: This quantifies how much error can be corrected by deploying the functionality.

5.1.12. JOINT OPTIMIZATION OBJECTIVE

The joint optimization objective of the agentic fraud detection system balances competing operational goals:

$$J = f(F1\_fraud, S\_explain, L, G) \quad \text{Eq. 12}$$

where J is the scalar objective function balancing fraud detection accuracy, explainability coverage, end-to-end latency, and governance compliance; all four dimensions are simultaneously optimized during system tuning.

5.1.13. DATASET REPRESENTATION FUNCTION

The dataset representation for the fraud detection training pipeline is defined as:

$$D(i, j, k) = Q\_src(i) \cdot Metric(k) / T\_proc(j) \quad \text{Eq. 13}$$

where Q_src(i) is the source-specific data quality (internal claims, external fraud registries, or synthetic data), Metric(k) denotes the selected performance metric (F1, AUC-ROC, etc.), and T_proc(j) represents the preprocessing time per data batch.

5.1.14. AGENTIC PERFORMANCE INDEX (API)

The Agentic Performance Index is a consolidated composite quantitative measure of explainable fraud detection and operational efficiency for the system that you train it on, which becomes represented in:

$$API = (\eta \cdot F1\_fraud \cdot (1 - FPR)) / Q\_total \quad \text{Eq. 14}$$

where η is the resilience efficiency (Eq. 9), F1_fraud is the fraud detection F1-score (Eq. 3), FPR is the false positive rate (fraud mis-flags), and Q_total is the cumulative system quality (Eq. 1). The index penalizes high false positive rates while rewarding accuracy and efficiency.

6. CONCLUSION

Fraud is one of the biggest threats in digital insurance and therefore needs tracking, detection and governance. The Digital Insurance platform uses big data from both internal and external sources to automate the process of identifying anomalies in fraud cases across a wide array of product lines. This is a Digital Insurance ecosystem where one can not only detect but also make decisions that are transparent owing to Explainable Agentic Artificial Intelligence. The private-sector fraud problem can be solved both by identifying more white-collar crooks via explainable AI and also (with oversight functions) automating decision pathways. Automation responses seek approval from stakeholders, while critical decisions are sequestered to domain experts in the interest of safety, fairness and transparency. Fraud is rapidly becoming such a far-reaching risk to digital insurance business models, where scams and device misuse result in losses that cannot be recovered on the threshold of claim misrepresentation, which can cause significant reputational damage for insurers. This risk can be offset by supervised agentic artificial intelligence techniques through automated and transparent detection of fraud signals along with control over the actions taken. A definition of a type of environment that uses these capabilities as well is offered, and this framework is aligned with the model for agentic AI. Explainable AI is injected into three components of the HPDA value chain and relevant building blocks such as expert systems to augment human decision-related tasks.

REFERENCES

[1] S. Russell, and P. Norvig, Artificial intelligence: A modern approach, 5th ed., Pearson, 2026.

[2] R. S. Garapati, A. R. Aitha, U. S. Yandamuri, V. R. R. Gottimukkala, A. R. Nagubandi, and S. H. Kolla, "Cloud-Native Orchestration of Multi-Counterparty Derivatives and Collateral in Manufacturing Enterprises via AI-Assisted Financial Audit Engines," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), pp. 1–6, Mar. 2026, doi: 10.1109/aiei69164.2026.11497364.

[3] Rohit Gorle, "Post-Quantum Identity and Access Management for Enterprise Cloud Security," International Journal of Special Education, vol. 41, no. 14s, pp. 932–943, 2026. Retrieved from https://internationalsped.com/index.php/ijse/article/view/4643

[4] P S L Narasimharao Davuluri, "Autonomous Compliance Systems: AI, Event Streaming, and the Future of Financial Crime Prevention," Journal of Informatics Education and Research, vol. 6, no. 1, pp. 1281-1294, 2026.

[5] Tarun Vakkalagadda, and Vijayanandh Rajamanickam, "Agentic AI Architectures for Next-Generation Investment Advisory and Portfolio Intelligence," International Journal of Computer Information Systems and Industrial Management Applications, vol. 18, no. 9s, pp. 1206–1219, 2026. Doi: https://doi.org/10.70917/ijcisim-2026-3548

[6] R. Mattaparthi, "GenAI-Augmented Diagnostic Reasoning for Diesel Engine Fault Triage: A Large Language Model Framework for Technician Decision Support at Scale," Journal of Material Sciences & Manufacturing Research, vol. 6, no. 12, p. 1, Dec. 2025, doi: 10.47363/jmsmr/2025(6)226.

[7] M. Krishnan, A. R. Aitha, K. Amistapuram, B. P. Nandan, P. K. Kaulwar, and J. Singireddy, "Human-in-the-Loop Hybrid Neuro-Symbolic AI Model for Reliable Data Engineering in High-Stakes Industrial Systems," 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN), pp. 1–7, Nov. 2025, doi: 10.1109/gcwcn66157.2025.11448516.

[8] Y. Chen, M. Lopez, and H. Wang, "Autonomous AI agents for transparent insurance claim validation and fraud analytics," IEEE Transactions on Artificial Intelligence, vol. 7, no. 1, pp. 112–129, 2026.

[9] R. S. Garapati, S. Paleti, R. Meda, K. C. Nagabhyru, and B. S. Deepa Priya, "Physical-Unclonable-Function-Based Secure and Anonymous User Authentication for Smart Homes," Lecture Notes in Electrical Engineering, pp. 367–378, 2026, doi: 10.1007/978-3-032-20235-2_33.

[10] R. Inala, "Cloud-Native AI and MDM Framework for Next-Generation Insurance and Retirement Data Products," International Journal of Engineering & Extended Technologies Research, vol. 8, no. 03, May 2026, doi: 10.15662/ijeetr.2026.0803006.

[11] M. Rahman, J. Carter, and Q. Zhao, "Human-centric explainable AI pipelines for insurance decision automation," Information Sciences, vol. 701, 2026.

[12] K. Jyoshna, R. K. Peddi, P. Nayak. S. Dhanamalar, M, and S. Punitha, "Spatio-Temporal Graph Neural Network with Global Spatio-Temporal Network for the Traffic Flow Prediction," 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–6, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566462.

[13] N. B. Hiremath, S. K. Kolla, R. Sunkara, S. Lal. G, and K. Sireesha, "Adaptive and Intelligent Secure Multimedia Transmission for Dynamic Communication Systems," 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–8, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566143.

[14] S. H. Kolla and R. Mattaparthi, "Hybrid Gen AI Systems: Integrating Small LMs with Large Language Models for Cost-Efficient Enterprise Automation and Decision Intelligence," International Journal of Research Publications in Engineering, Technology and Management, vol. 08, no. 06, Dec. 2025, doi: 10.15662/ijrpetm.2025.0806038.

[15] T. Nguyen, P. Sharma, and D. Wilson, "Transparent deep learning systems for regulatory-compliant insurance automation," Future Generation Computer Systems, vol. 174, pp. 88–106, 2026.

[16] T. Kalita, S. Prasad Tiwari, A. Reddy Aitha, and A. Garg, "Evolutionary and Swarm-Based Metaheuristics for Neural Architecture Search and Hyperparameter Tuning," SSRN Electronic Journal, 2026, doi: 10.2139/ssrn.6708638.

[17] R. Paramasivam, S. S. Reddy, V. A. R. Reddy, K. Sandra, and M. V. S. Narayana, "JellyFish Search Optimization with Arctic Puffin Optimization Algorithm for Efficient Routing Based on the Software Defined Communication Network," 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–6, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566144.

[18] X. Li, Y. Zhao, and J. Kim, "Explainable AI and probabilistic drift monitoring in automated insurance workflows," Journal of Systems and Software, vol. 229, 2026.

[19] K. C. Nagabhyru, S. Singireddy, A. L. Gadi, G. K. Sheelam, and D. Kapila, "Toward Secure and Usable Communication-Centric Authorization Models for Smart Homes," Lecture Notes in Electrical Engineering, pp. 355–366, 2026, doi: 10.1007/978-3-032-20235-2_32.

[20] P. R. Sudha Rani, K. Amistapuram, V. Pamisetty, S. Singireddy, D. N. Kummari, and G. K. Sheelam, "Hybrid Knowledge Graph–Deep Learning Framework for Automated Exception Handling and Investigation in Complex Insurance Claims," 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN), pp. 1–6, Nov. 2025, doi: 10.1109/gcwcn66157.2025.11448301.

[21] B. Bedi, U. S. Yandamuri, D. N. Kummari, A. R. Nagubandi, K. Amistapuram, and A. D, "AI-Driven Sentiment and Behavior Analysis for Sustainable Business Growth," 2026 International Conference on Emerging Research in Smart Electronics and Machine Informatics (ECMI), pp. 1–11, Apr. 2026, doi: 10.1109/ecmi68341.2026.11603189.

[22] Al-Ghezi and L. Wiese, "Analyzing workload trends for boosting triple stores performance," Information Systems, vol. 125, p. 102420, Nov. 2024, doi: 10.1016/j.is.2024.102420.

[23] T. Chen, H. Xu, and Y. Zhang, "AIOps-driven anomaly detection and root cause analysis for cloud-native enterprise platforms," Future Generation Computer Systems, vol. 158, pp. 310–324, 2024.

[24] D. Kaur, S. Uslu, K. J. Rittichier, and A. Durresi, "Trustworthy Artificial Intelligence: A Review," ACM Computing Surveys, vol. 55, no. 2, pp. 1–38, Mar. 2023, doi: 10.1145/3491209.

[25] D. Dawadi, U. S. Yandamuri, S. K. V, J. L. A. Stalin, and R. Naveenkumar, "Privacy-Aware Edge-Based Intelligent Video Analytics for Scalable Crowd Management in Smart Cities," 2026 3rd International Conference on Integrated Intelligence and Communication Systems (ICIICS), pp. 1–7, June 2026, doi: 10.1109/iciics67880.2026.11483444.

[26] M. Recharla, R. S. Garapati, V. D. V. K. Bandi, U. S. Yandamuri, and B. M. Mangalampalli, "Generative AI-Enhanced Data Engineering Pipelines for Predictive Biomarker Discovery in Alzheimer's and Kidney Disease," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), pp. 1–5, Mar. 2026, doi: 10.1109/aiei69164.2026.11496858.

[27] J. Mökander, J. Schuett, H. R. Kirk, and L. Floridi, "Auditing large language models: a three-layered approach," May 2023, doi: 10.1007/s43681-023-00289-2.

[28] S. Lapologang and S. Zhao, "The impact of environmental policy mechanisms on green innovation performance: the roles of environmental disclosure and political ties," Technology in Society, vol. 75, p. 102332, Nov. 2023, doi: 10.1016/j.techsoc.2023.102332.

[29] C. Masina, H. Veludandi, A. Aksharavashist, K. Sai Siddhartha Royal, and P. Parwekar, "Bridging Gaps: Addressing Challenges in NGO Resource Management," Lecture Notes in Electrical Engineering, pp. 99–108, 2026, doi: 10.1007/978-3-032-20235-2_10.

[30] R. Kuppuswami, U. S. Yandamuri, M. Bhatt, M. Patel, and Nagendran. R, "A Cognitive Clustering Framework for Wireless Sensor Networks Using Quantum Machine Learning," 2026 3rd International Conference on Integrated Intelligence and Communication Systems (ICIICS), pp. 1–6, June 2026, doi: 10.1109/iciics67880.2026.11483591.

[31] L. Wang et al., "A Survey on Large Language Model-based Autonomous Agents," Sept. 2023. doi: 10.48550/arXiv.2308.11432.

[32] Z. Xi et al., "The Rise and Potential of Large Language Model-Based Agents: A Survey," Sept. 2023. doi: 10.48550/arXiv.2309.07864.

[33] S. Yao et al., "ReAct: Synergizing Reasoning and Acting in Language Models," Mar. 2023. doi: 10.48550/arXiv.2210.03629.

[34] A. Gopinath, P. S. L. N. Davuluri, U. B. Kumar, Sahana. M P, and G. Yamini, "Communication Aware Malware Detection Using Network Flow Bytecode Fusion With Adaptive Focal Optimization," 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–8, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566323.

[35] P. Nayak S, R. Loganathan, Velmurugan. R, S. R. K. Sarma A, and V. Jayaraj, "Scale-Aware Dilated Lightweight Convolutional Network Improving Solar Panel Defect Classification through Efficient Electroluminescent Image Analysis Techniques," 2026 2nd

International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–7, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566218.
[36] B. D. Bhavani, Sowmya. S. R, R. Loganathan, S. Nagaraj, and Vasukidevi. G, “Evolutionary Gravitational Neocognitron Neural Network, Snow Leopard Optimization and Deep Graph Reinforcement Learning for Routing Protocol in WSN,” 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–8, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566191.
[37] A. R. Aitha, “Explainable Agentic AI Framework for Automated Insurance Fraud Detection and Predictive Risk Intelligence,” International Journal of Advanced Research in Computer Science & Technology, vol. 9, no. 03, May 2026, doi: 10.15662/ijarcst.2026.0903009.
[38] S. Raj, P. Prashanthi, S. K. Kolla, S. Bandaru, N. Bhardwaj, and S. P, “Lightweight Cryptographic Schemes for Resource-Constrained IoT and Healthcare Devices,” 2026 International Conference on Multidisciplinary Innovations For Smart & Sustainable Future (MISSF), pp. 1–6, Apr. 2026, doi: 10.1109/missf68264.2026.11521999.
[39] N. Shinn, B. Labash, and A. Gopinath, "Reflexion: Language Agents with Verbal Reinforcement Learning," arXiv (Cornell University), vol. 1, no. 1, Mar. 2023, doi: 10.48550/arxiv.2303.11366.
[40] M. S. R. L. Reddy, T. Sunitha, K. Kanchana, A. R. Nagubandi, A. R. Segireddy, and S. N. Bhavanam, “AI-Enhanced Blockchain Consensus Mechanisms for Secure Transaction Validation,” 2026 International Conference on Emerging Research in Smart Electronics and Machine Informatics (ECMI), pp. 1–11, Apr. 2026, doi: 10.1109/ecmi68341.2026.11602724.
[41] J. Park et al., "Generative Agents: Interactive Simulacra of Human Behavior," Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, vol. 23, 2023, doi: 10.1145/3586183.3606763.
[42] M. Krishnan et al., “Engineering Intelligent Cloud-Native Data Ecosystems for Predictive Decision-Making in Industry,” Journal of European Economic History, vol. 7, no. 2, pp. 68–88, June 2026, doi: 10.61336/jeeh/26-2-7.
[43] R. Loganathan, K. Amistapuram, and A. R. Aitha, "Letter to the Editor re Annual Updates of the European Association of Urology – European Society for Pediatric Urology (EAU-ESPU) Pediatric Urology Guidelines: Are Large-Language Models (LLM) Better Than the Usual Structured Methodology?," Journal of Pediatric Urology, p. 106057, June 2026, doi: 10.1016/j.jpurol.2026.106057.
[44] W. X. Zhao et al., “A Survey of Large Language Models,” arXiv (Cornell University), Mar. 2023, doi: 10.48550/arxiv.2303.18223.
[45] B. M. Mangalampalli, R. K. Peddi, S. K. Kolla, V. A. R. Reddy, N. Mangala, and A. Seenu, "Explainable Clinical Graph Intelligence Framework for Longitudinal Risk Modeling and Care Pathway Optimization," 2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS), pp. 1–6, June 2026, doi: 10.1109/icicds70526.2026.11604804.
[46] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," arXiv (Cornell University), vol. 2, Feb. 2017, doi: 10.48550/arxiv.1702.08608.
[47] S. S. Kumar, R. S. Garapati, A. R. Segireddy, S. Kalisetty, R. Inala, and K. C. Nagabhyru, "Hybrid Deep Neural Network–DevOps Pipeline Optimization for Risk Prediction in Cloud-Native Workers' Compensation Platforms," 2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON), pp. 1–6, Mar. 2026, doi: 10.1109/i3ctcon68242.2026.11507219.
[48] M. T. Ribeiro, S. Singh, and C. Guestrin, " 'Why Should I Trust You?': Explaining the Predictions of Any Classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, pp. 1135–1144, Aug. 2016, doi: 10.1145/2939672.2939778.
[49] B. M. Mangalampalli, R. K. Peddi, S. K. Kolla, V. A. R. Reddy, N. Mangala, and A. Seenu, "Explainable Clinical Graph Intelligence Framework for Longitudinal Risk Modeling and Care Pathway Optimization," 2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS), pp. 1–6, June 2026, doi: 10.1109/icicds70526.2026.11604804.
[50] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017. doi: 10.48550/arXiv.1705.07874.
[51] Ch. S. Rao, V. D. V. K. Bandi, L. M. K. Brahmandam, S. Gantikota, and K. Kannan, “Topology-Aware Hypergraph Neural Network Framework for Intelligent Decision Support Systems in Network Operations,” 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–7, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566638.
[52] Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, pp. 1-800, 2016.
[53] M. V. K. Kumar, S. H. Kolla, V. Pamisetty, L. Pandiri, U. S. Yandamuri, and D. Valiki, "Enterprise-Scale Generative AI Agents for Secure and Governed Automation in Insurance and Public Financial Management," 2026 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE), pp. 1–6, May 2026, doi: 10.1109/iccrtee68719.2026.11566439.
[54] T. H. Davenport, “Artificial Intelligence for the Real World,” 2018. [Online]. Available: https://hbr.org/webinar/2018/02/artificial-intelligence-for-the-real-world
[55] G. Padma, V. A. R. Reddy, S. Devi. K. A, B. Buvaneswari, and K. S. S. Kumar, “Privacy-Preserved Face Recognition Biometric Authentication Using FaceNet and Zero-Knowledge Proofs for Secure, Access Control on Decentralized Blockchain Networks,” 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–8, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566528.
[56] K. Naveen, R. Lingam, G. R. Kunduru, N. Mangala, N. N. Reddy, and P. Balaji, “Intelligent Loan Approval System: Machine Learning for Home and Education Loan Eligibility,” 2026 Contemporary Computing Innovations Conference (CCIC), pp. 1–6, Feb. 2026, doi: 10.1109/ccic68129.2026.11486013.
[57] S. Amershi et al., "Software Engineering for Machine Learning: A Case Study," 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), May 2019, doi: 10.1109/icse-seip.2019.00042.
[58] S. Manikandan, G. P. Singh, V. R. R. Gottimukkala, P. S. L. N. Davuluri, R. Sadagopan, and T. V. Kumar, "Human-in-the-Loop Automation Patterns for Financial Services Using Camunda + Modern Java UIs," 2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON), pp. 1–6, Mar. 2026, doi: 10.1109/i3ctcon68242.2026.11507795.
[59] D. Sculley, et al., “Hidden technical debt in machine learning systems,” Advances in Neural Information Processing Systems, vol. 28, pp. 2503–2511, 2015.
[60] R. S. Garapati, S. Paleti, R. Meda, K. C. Nagabhyru, and B. S. Deepa Priya, "Physical-Unclonable-Function-Based Secure and Anonymous User Authentication for Smart Homes," Lecture Notes in Electrical Engineering, pp. 367–378, 2026, doi: 10.1007/978-3-032-20235-2_33.

[61] Supriya. R K, N. Mangala, R. Priyanka, R. A. Mohammed Sait, and P. P. Selvam, "Privacy Preserving Analytics for Intelligent in Internet of Things System in Health Care Using Graph Neural Network with Latent Graph Inference Technique," 2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN), pp. 1–6, Apr. 2026, doi: 10.1109/iciscn67954.2026.11566407.

[62] U. S. Yandamuri, "Adaptive Intelligence Networks for Human-centred Enterprise Automation and Governance, 2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS), pp. 678-683, 2026.

[63] B. Shneiderman, "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy," International Journal of Human–Computer Interaction, vol. 36, no. 6, pp. 495–504, Mar. 2020, doi: 10.1080/10447318.2020.1741118.

[64] ENISA, "ENISA Threat Landscape 2023," Oct. 2023. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023

[65] SUNKARA, S. K. (2025). LEVERAGING AI, IoT, AND BLOCKCHAIN FOR SCALABLE DIGITAL TRANSFORMATION IN POST-HARVEST SUPPLY CHAINS: A MULTI-SECTOR APPROACH TO ENHANCING EFFICIENCY AND TRACEABILITY (Vol. 26, Issue 7, pp. 2757–2766).