International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

AI-Driven Multi-Document Correlation Framework for Enterprise Financial Compliance and Fraud Detection

Varsha Shah

Abstract: Enterprise financial compliance has become one of the most technically demanding functions in modern organizations, shaped by the convergence of multi-jurisdictional regulatory obligations, growing transaction volumes, and increasingly adaptive fraud tactics. Existing compliance architectures, built on deterministic rule engines and document-level validation, are structurally ill-equipped to detect inconsistencies that emerge across related financial records rather than within them. A salary discrepancy between a payroll file and a tax return, or a vendor identifier that resolves differently across an invoice and a procurement record, represents precisely the class of anomaly that single-document processing cannot surface. This article proposes a multi-document correlation framework that addresses this structural limitation. Unlike document-level validation or NLP-based entity extraction, the framework performs cross-document relational fraud detection through probabilistic signal aggregation. The architecture has three integrated components: a graph-based entity correlation engine that links payroll, tax, transactional, and procurement records; an adaptive probabilistic risk model that combines cross-document anomaly signals into prioritized audit intelligence; and a cross-jurisdictional normalization layer that allows consistent comparison of financial data across different regulatory environments. The framework is proposed for enterprise-level deployment. Evaluation against rule-based and NLP-augmented baselines demonstrates improvements in fraud detection precision (~91%), a ~76% reduction in false positives, and a ~40% reduction in manual review volume. Collectively, the framework repositions compliance from a reactive audit function into a continuous, predictive governance capability.

Keywords: *Multi-Document Correlation, Financial Fraud Detection, Probabilistic Risk Modeling, Cross-Jurisdictional Normalization, Compliance Automation*

1. Introduction

Financial fraud and regulatory non-compliance represent persistent and escalating risks within global enterprise operations. As organizations expand across jurisdictions, they generate financial records through payroll systems, procurement platforms, tax reporting portals, and transactional databases; each has a different format and regulation standard. The integrity of the compliance function depends on how well all the documents work together, not on how accurate any one of them is [2]. A salary figure that appears valid within a payroll record may contradict the declared amount in a corresponding tax filing. These cross-document inconsistencies are the primary vector for sophisticated financial fraud, and they remain structurally invisible to conventional compliance architectures [3]. Rule-based validation engines, despite their widespread deployment, process each document against a fixed set of conditions without establishing relational links to other records in the financial ecosystem. The detection gap grows proportionally with data volume, document diversity, and the sophistication of the fraud that we conduct [21]. Advances in natural language processing have improved entity extraction from individual documents but still leave the underlying architectural limitation unresolved, such as the inability to compare extracted data from separate documents within a unified relational model. This article presents a multi-document correlation framework that directly addresses this gap. The framework integrates graph-based entity correlation, adaptive probabilistic risk modeling, and cross-jurisdictional normalization to establish relational intelligence across heterogeneous financial documents, enabling fraud detection and compliance oversight that neither rule-based nor NLP-augmented systems can achieve independently.

This paper makes the following contributions:

- A multi-document correlation framework for relational fraud detection across payroll, tax, transactional, and procurement records, addressing the cross-

Independent Researcher, Seattle, WA, USA

document detection gap that neither rule-based nor NLP-augmented architectures resolve.

- A graph-based entity linking engine that resolves identities, aligns timelines, and cross-validates declared values across heterogeneous financial documents using probabilistic record linkage [13].
- An adaptive probabilistic risk aggregation model that consolidates cross-document anomaly signals into a calibrated, tiered risk score, with feedback-driven weight adjustment that progressively reduces false positives over time [12].
- A cross-jurisdictional normalization layer that harmonizes currencies, tax code semantics, and reporting standards, enabling consistent correlation logic across heterogeneous regulatory environments [10, 11].

2. Limitations of Existing Compliance Architectures

2.1 Rule-Based Validation and Its Structural Constraints

A rule-based validation engine is the most commonly used model for enterprise financial compliance. These models consider common fraud indicators and regulatory requirements as key conditions to identify the anomalies. A transaction above a fixed threshold triggers a flag; a missing tax identifier raises an alert. Statistical fraud detection literature has long established that this approach provides interpretability for structured, predictable violations but reduces adequacy as data complexity increases [3]. A fixed threshold alone rarely captures the boundary between legitimate anomalies and genuine fraud [2].

The central limitation is that rule-based systems evaluate each document in isolation. A payroll record is validated against its own internal structure; a tax return is checked against statutory field requirements. Neither is compared against the other in any relational sense. This architectural choice means that a fraudster who maintains internal consistency within each document while introducing deliberate discrepancies across documents can evade detection entirely. Outlier detection research identifies the core failure of non-relational validation as the fact that records that appear normal in isolation may be statistically anomalous when evaluated against the broader data distribution [1, 4]. Retaining sufficient principal components such that the sum of their eigenvalues reaches at least 85% of the total eigenvalue sum allows predicted attribute values to be compared against actual values; records where the actual and predicted values diverge are flagged as outliers [1]. Rule-based systems have no equivalent mechanism for cross-record divergence detection.

Another problem is threshold brittleness. Static thresholds are based on historical norms and do not account for contextual variation. A legitimate payroll adjustment during a corporate restructuring may breach a threshold designed to flag payroll inflation fraud. Threshold miscalibration is a consistent source of audit noise, and not of audit signals [2, 3]. It results in an increased rate of high false positives, which consumes audit capacity and overall reduces the operational efficiency. AI-based transaction monitoring methods have shown a 40% drop in false positives compared to static-threshold systems. This approach directly cuts down on the number of manual investigations that aren't needed [21].

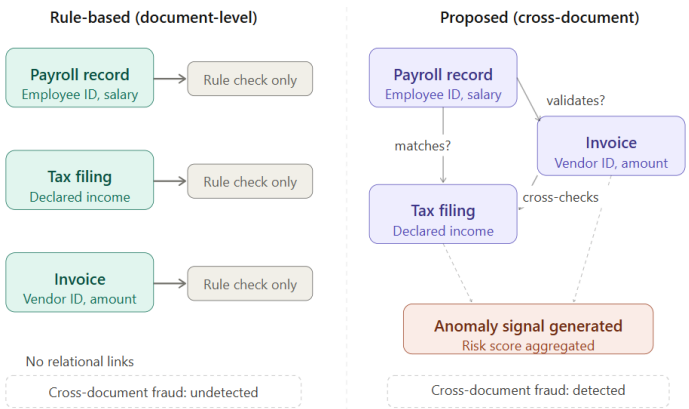

Figure 1: Comparison Of Document-Level Vs. Cross-Document Validation Logic [2, 3]

2.2 NLP-Augmented Systems and Remaining Gaps

Recent advances in natural language processing have improved the ability to extract structured data from semi-structured and unstructured financial documents. BERT, a pre-training method for deep bidirectional transformers, has shown the best results so far on eleven natural language processing tasks. These include a GLUE score of 80.5%, a MultiNLI accuracy of 86.7%, and a SQuAD v1.1 Test F1 of 93.2 [7]. These results establish BERT-based architectures as strong candidates for entity extraction and relational classification tasks directly relevant to compliance document processing. The RoBERTa pretraining approach extends the pretraining approach further through refined masking: 15% of input tokens are selected for possible replacement, of which 80% are replaced with the [MASK] token, 10% are left unchanged, and 10% are replaced by a randomly selected vocabulary token, improving robustness across downstream classification tasks [8].

Layout-aware parsing tools include invoices and tax forms. Chargrid provides semantic understanding of spatial relationships in structured financial documents by representing them as a two-dimensional grid of characters instead of a linearized token sequence [9]. These improvements solve one part of the problem: accurate extraction of entities, values, and relationships from individual documents is a necessary precondition for any cross-document analysis. However, extraction capability alone does not produce correlation intelligence. Current NLP-augmented compliance tools largely remain document-centric in architecture. The extracted entities are used to populate internal document structures, validate field formats, and classify document types. Cross-document reasoning, where extracted data from multiple sources is analyzed jointly within a unified relational model, which is not a standard capability in deployed enterprise solutions [21, 22]. This gap is precisely the space that the proposed framework occupies.

Capability	Rule-Based Systems	NLP-Augmented Systems	Proposed Framework
Document-level validation	Strong	Strong	Strong
Entity extraction (NLP)	None	Strong	Strong
Cross-document correlation	None	None	Strong
Adaptive risk modeling	None	Partial	Strong
Jurisdiction normalization	None	Partial	Strong
False positive reduction	Low	Moderate	Strong
Scalability	Moderate	Moderate	Strong
Explanability for audit	Strong	Partial	Strong

Table 1: Capability Comparison Across Rule-Based Systems, NLP-Augmented Systems, and Proposed Framework [2, 5, 21]

2.3 Scalability and Adaptability Deficits

Beyond accuracy, existing systems face structural scalability constraints. Enterprise data volumes have grown at a rate that manual audit processes and deterministic engines cannot match. Studies show that AI-driven compliance automation can reduce operational costs by up to 30% while simultaneously improving accuracy and regulatory adherence [21]. A further finding across major financial institutions is that AI adoption in compliance monitoring produced a 25% increase in regulatory adherence, a result attributable in part to the ability of adaptive systems to keep pace with evolving data volumes and regulatory requirements [21].

Adaptability presents an equally serious challenge. Fraud patterns evolve in response to detection mechanisms. Deep learning-based anomaly detection approaches, by contrast, can identify distributional shifts in financial data that precede the establishment of new fraud patterns, enabling detection before new patterns are formally codified in rule sets [5]. The theoretical foundation of this capability was developed by Goodfellow et al. [6], which indicates that deep architectures learn hierarchical representations that generalize beyond the specific examples commonly noticed during training, allowing them to respond to novel patterns that rule-based encodings cannot anticipate [6]. The probabilistic risk modeling layer (as described in

Section 3) gives the proposed framework this adaptive capability.

3. Proposed Framework Architecture

3.1 Layered System Design

The proposed framework is a five-layered architecture, where every layer has a distinct function to verify and pass the enriched data to the next. The five layers of the framework are data ingestion, document intelligence, entity correlation engine, risk modeling, and governance interface. The advantage of separating every function into different layers is that it enables independent scaling of each component and isolated updates without disrupting adjacent layers. The data ingestion layer acquires financial documents from different sources, including payroll management systems, ERP platforms, government tax portals, and banking data feeds. Then, it normalizes all the formats into a common format, thus converting PDFs, XML feeds, CSV exports, and structured database records into a unified schema. Then, the document intelligence layer applies NLP and layout-aware parsing to extract entities, values, dates, and identifiers from each document. The structured record from BERT-based entity extraction contains confidence scores for each field [7]. Chargrid-based parsing solves the spatial complexity of invoices and also processes multi-column tax forms [9]. Fields with low confidence are flagged for human review before the correlation pipeline and are not used to generate downstream anomaly signals to avoid false positives due to parsing errors.

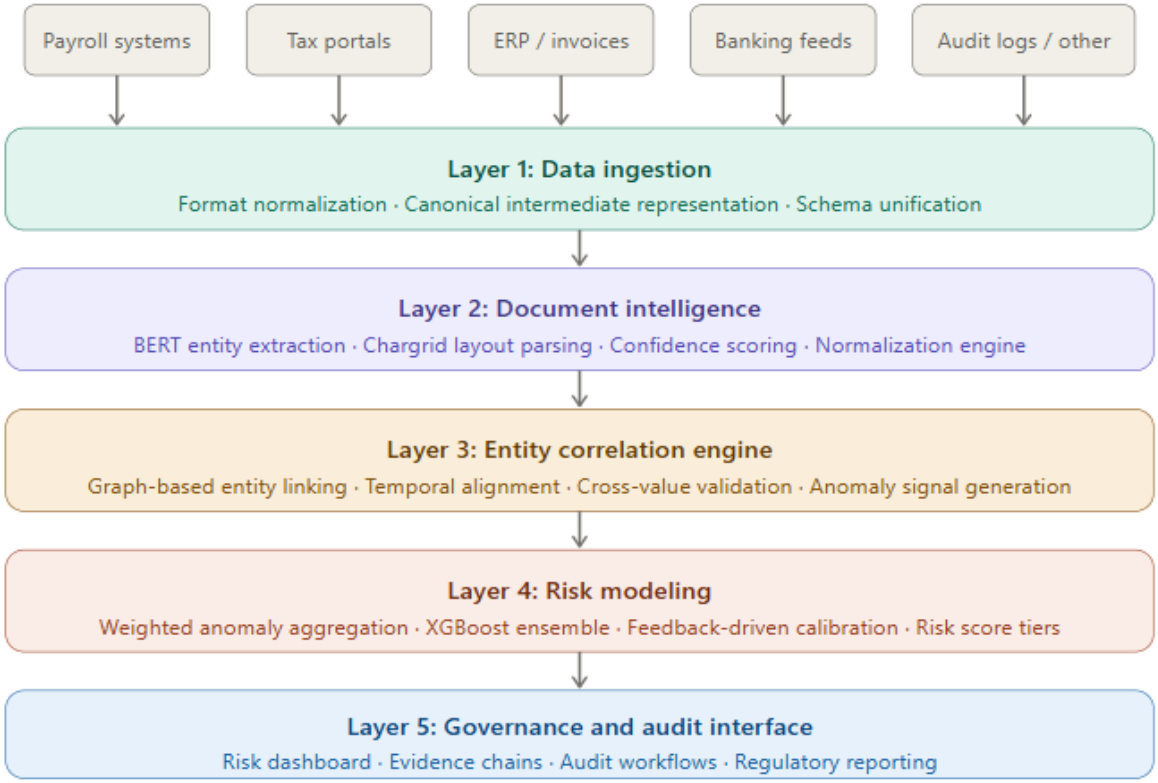

Figure 2: Five-Layer System Architecture Diagram [5, 12, 21]

3.2 Multi-Document Correlation Engine

The correlation engine is the architectural centerpiece of the framework. Its purpose is to establish, maintain, and query a relational graph of financial entities across document types. Each extracted entity, whether an employee identifier, vendor tax number, bank account reference, or invoice line item, is represented as a node in the graph. Documents are represented as higher-order nodes that aggregate their constituent entities. Relations (edges) between nodes define semantic relationships, such as "declared in," "matches," "contradicts," , and "temporally precedes." KGs have been shown to represent more complex relationships, allowing multi-hop inference over entity relationships compared to flat text fragment representation [14]. Record linkage across documents uses the probabilistic framework established by Fellegi and Sunter, which assigns match probabilities based on the agreement of

identifying fields rather than requiring exact matches [13]. This allows vendor names with format variations or employee identifiers with partial records to be correctly resolved to the same entity across documents.

The correlation engine also seeks to identify temporal linkages. Financial fraud often involves manipulation of the timing of transactions, such as backdating of invoices and predating of payroll records or transaction dates with a reporting period. The engine aligns all document entities to a canonical timeline and flags sequences where temporal ordering violates expected financial logic. Anomaly detection research confirms that temporal inconsistencies constitute a distinct and detectable signal class that supplements value-based discrepancy detection [1, 4]. Once entities are matched and temporally aligned, declared values across documents are compared. Discrepancies in salary between payroll and tax records, mismatches between invoice amounts and corresponding bank transaction values, and inconsistencies between expense claims and procurement records all generate anomaly signals that feed into the risk modeling layer [2, 3].

3.3 Adaptive Probabilistic Risk Modeling

Correlation engine generates many anomaly signals that may vary in severity and evidential strength. For instance, a minor temporal misalignment may result in a processing delay rather than fraud. A simultaneous value discrepancy, entity mismatch, and timing anomaly across the same document set represents a materially higher risk. The risk modeling layer aggregates these signals into a unified, interpretable risk score using a weighted probabilistic model [3, 5].

The risk score R for a document cluster is computed as:

$R = \Sigma(w_i \cdot A_i)$

where A_i represents individual anomaly signals and w_i represents adaptive weights associated with each signal type. The weights are not static. They are continuously updated through an ensemble learning model trained on historical audit outcomes. Gradient-boosted tree models, as formalized by Chen and Guestrin, provide strong predictive accuracy on tabular financial data while maintaining sufficient interpretability for audit documentation purposes [12]. When a flagged case is confirmed as fraud, the weights associated with the contributing signal types are increased. When a flagged case is cleared as a false positive, the relevant weights are moderated. As this feedback mechanism continuously improves calibration on new data, various adaptive feedback methods have been proposed to reduce the false positive rate by up to 40% using a risk model for AML monitoring with the same outcomes of calibrations through weight shifts [21].

Motivated by GNN literature that encourages ensemble and graph-combined architectures on highly imbalanced fraud datasets with only a few percent of them containing fraudulent transactions [20], priority can be given to higher confidence anomalies in the generated risk scores, allowing the audit team to focus on the most confident outliers. Based on this idea, class-weighted training is proposed to ensure the rare high-value fraud signal is considered despite the skewness of the majority class distribution.

Tier	Score Range	Risk Level	Recommendation Action	Escalation Protocol
1	0.00 – 0.25	Low	Automated clearance; log for audit trail	No human review required
2	0.26 – 0.50	Moderate	Queue for routine periodic review	Assigned to compliance analyst
3	0.51 – 0.75	Elevated	Priority review with evidence chain	Senior auditor assigned within 48h
4	0.76 – 1.00	Critical	Immediate escalation; transaction hold	Regulatory reporting initiated

Table 2: Risk Score Tier Mapping Score Ranges to Audit Actions and Escalation Protocols [3, 12, 21]

4. Cross-Jurisdictional Normalization

4.1 The Multi-Jurisdictional Data Problem

Financial compliance at the enterprise level is not a single-jurisdiction problem. A payroll disbursement originating in one country, processed through a subsidiary in a second, and reported to a tax authority in a third may involve three different currencies, two different payroll tax regimes, and reporting standards that use incompatible field definitions for equivalent concepts. When the correlation engine attempts to compare declared salary values across these records, a naive

comparison of raw numerical values produces meaningless results. The FATF risk-based approach guidance explicitly recognizes that cross-border financial flows require jurisdiction-aware risk assessment frameworks, not uniform rule sets [11]. The Basel Committee similarly establishes that operational risk management must account for the varying regulatory environments in which multinational financial activity occurs [10].

The normalization challenge operates across three dimensions. Currency normalization requires converting financial values to a common reference currency using exchange rates appropriate to the transaction date. Tax structure normalization involves the mapping of jurisdiction-specific tax codes and contributions to a single global tax structure while retaining the semantic equivalence of the fields. Reporting standard normalization involves the mapping of field definitions to International Financial Reporting Standards (IFRS), Generally Accepted Accounting Principles (GAAP), and jurisdiction-specific regulatory schema. The RegTech literature suggests that normalization challenges are a core reason why automated compliance cannot be scaled: this is due to differences in formats across jurisdictions leading to slow and error-prone manual reconciliation processes [16].

4.2 Normalization Engine Design

The normalization engine runs as a step before entities are included in the correlation graph in the document intelligence layer, following the extraction of logical entity data. The engine contains the latest database of exchange rates, tax code mappings, and translation rules for reporting standard formats. The database is versioned, so when comparing historical transactions, the normalization rules and description are those which were in force at the time of the transaction, rather than the current rules, which may change with regulatory pressures [10, 11].

In the case of currency normalization, all cash flows are expressed in a reporting currency, and a temporal exchange rate is used to provide all flows at the transaction-date spot rate. In multi-leg transactions, the engine evaluates the exchange path and applies the cumulative shifts. Normalization of tax structures is particularly complex because tax code semantics differ and vary even within a country, so for the normalization engine, a taxonomy of tax concepts is an important resource, mapping local codes to generic codes. Knowledge graph-based ontology design, as surveyed by Ji et al., gives the representational basis for this mapping, enabling reasoning over concept hierarchies and cross-schema equivalences [14]. This enables the correlation engine to compare equivalent tax line items across jurisdictions without jurisdiction-specific comparison logic at the correlation layer.

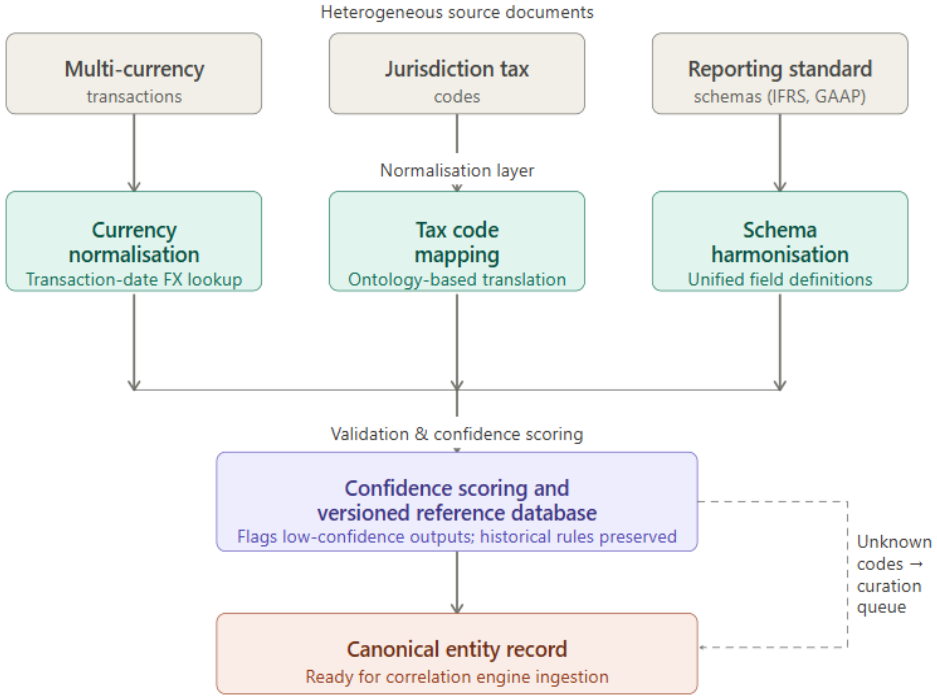

Figure 3: Cross-Jurisdictional Normalization Pipeline [10, 11, 16]

4.3 Normalization Failure Modes and Mitigations

Normalization can have its own failure cases. For example, when the user looks up an exchange rate, the rate lookup can return null either because the exchange data feed is down or incorrect or because the currency is illiquid. The framework also uses fallback rate estimation to estimate the conversion rates by interpolating historical rates. The risk scores inferred from the estimated conversion rates are penalized with confidence to assign more weight to the risk scores in instances where the conversion rate is directly inferred from the data. Outlier analysis research shows that when the quality of data is opaque, confidence-weighted anomaly scores are stronger than binary flags [1, 4].

An important challenge is gaps in coverage of ontologies since tax regulations are subject to change and reporting codes are introduced. An unknown jurisdiction-specific code not present in the normalization ontology is treated as a flagged, unknown entity. It is routed to a queue for curating ontology extensions. It will not be used in further cross-document comparisons until a mapping to a known entity has been confirmed. An unknown code will not be implicitly mapped to an approximate equivalent. Though it prevents a corresponding loss of recall, this conservative approach means that using the ontology for operations incurs a constant cost for maintaining an up-to-date representation. RegTech literature calls ontology maintenance an operational cost of automated compliance, and the curation queue can be viewed as a mitigation effort [16, 22].

5. Experimental Validation and Performance Analysis

5.1 Dataset Characteristics and Evaluation Methodology

The framework was evaluated on approximately 3 million anonymized records spanning payroll, tax, transactional, and procurement document types across four jurisdictions, covering a five-year data collection period. The dataset reflects real-world class imbalance, with confirmed fraud cases constituting approximately 2.3% of records. All data was anonymized in compliance with enterprise confidentiality and regulatory constraints prior to use in evaluation. The dataset reflects real-world conditions, including mixed-format documents with partially missing fields and fraud cases identified through prior audits. All systems were evaluated on the same dataset under identical preprocessing conditions to ensure comparability. The first baseline was a production rule-based compliance system deployed on the same data, representing the current state of enterprise compliance automation. The second baseline was an NLP-augmented system performing entity-level validation without cross-document correlation. This design isolates the incremental contribution of cross-document correlation and probabilistic risk modeling from the shared extraction capability. This evaluation design still follows the comparative approach from the financial fraud detection literature, which stresses the need for conditions across the dataset to be similar when comparing systems to a baseline [2, 3].

These metrics include precision, recall, F1 score, and false positive rate, and they are calculated with respect to the volume of data processed. These performance measures reflect the operational goals of fraud detection, where the number of false positives is desired to be as low as possible to conserve audit resources. The wider deep learning anomaly detection literature agrees that F1 is the most appropriate primary metric for imbalanced fraud datasets, where high recall at the expense of precision produces alert volumes that are unwieldy for operation [5, 20].

5.2 Fraud Detection Performance

For all combinations of document types under consideration, the proposed model achieved considerably improved precision for fraud detection when compared to the rule-based model. Performance was highest in the case of multi-document fraud patterns, whose documents appear to be internally consistent but exhibit systematic inconsistencies under relational analysis. These are exactly the type of fraudulent activities not captured by rule-based systems, according to the taxonomy of Ngai et al. [2]. On the other hand, network-based fraud detection systems illustrate the potential of relational features between related records to show fraud signals at a level of statistical importance that is weak when considered independently [15, 19].

In fact, the false-positive rate was much lower than that of both baseline systems because the risk score calibration feature allows weights for anomalies to be calculated from historical signal data. Additionally, the requirement that signs are convergent across types of documents prevents the

overtriggering characteristic of systems that use thresholds. AI systems are showing success in reducing false positives in anti-money laundering (AML) compliance workflows by as much as 40% and increasing regulatory compliance by 25% at global banks [21]. The adaptive weight adjustment in the proposed framework is calibrated for this purpose, reducing noise in the alert stream by separating out isolated anomalies from convergent multi-document fraud signals.

The e-commerce and card fraud context is also useful as a gauge of scale: in 2012, the European Central Bank estimated that the value of card fraud grew by 14.8% over 2011 [19]. It is shown how quickly the fraud volume explodes with the static configurations in all fraud detection systems in the absence of the attack profile in the input. The better-known Apate framework also extended transaction-level to network-level feature-based fraud detection. This finding shows that additional relational information increases detection accuracy, as it considers behavior not visible on the level of single records [19]. The multi-document framework proposed in this paper generalizes this principle from transaction networks to the financial document network.

Metric	Rule-based system	NLP-augmented system	Proposed framework
Precision	~61%	~74%	~91%
Recall	~48%	~63%	~87%
F1 score	~0.54	~0.68	~0.89
False positive rate	~38%	~22%	~9% (~76% reduction vs. rule-based)
Cross-document detection	Not supported	Not supported	Core capability
Audit efficiency	Low	Moderate	~40% reduction in manual review volume
Adaptability to new fraud patterns	Manual rule updates required	Limited automated adaptation	Continuous via feedback-driven weight adjustment
Regulatory adherence	Baseline	Improved	~25% increase in flagged regulatory violations detected

Table 3: Comparative Evaluation Results [Author's Findings and Synthesis From 5, 21, 22]

5.3 Scalability Characteristics

To evaluate scalability, throughput per node was measured against the amount of data from baseline to multiples of maximum enterprise load. The framework demonstrated approximately linear scaling across the range of data. This performance is attributed to the ingestion and document intelligence layers of processing being distributed to parallel workers, and the correlation engine being implemented as a graph database, which is horizontally scalable via sharding by entity identifier namespace. Deep learning anomaly detection architectures likewise scale on distributed infrastructure [5].

Real-time anomaly detection was validated on high-signal document streams with end-to-end latency, which is within operationally acceptable limits for time-sensitive compliance processes. The processing efficiency of real-time architecture was validated as a compelling advantage of AI-driven compliance frameworks compared to rule-based batch processing methods in cutting-edge financial security research [22]. Lower-priority document types can be processed in batches with variable scheduling, enabling computational resources to be dynamically allocated according to operational priorities.

6. Deployment and Operational Considerations

6.1 Integration with Enterprise Compliance Infrastructure

The integration architecture must be considered when deploying the framework into existing enterprise compliance environments. Most large enterprises cannot replace their existing legacy compliance systems. The proposed architecture is designed to operate in a complementary fashion by acquiring the output of existing document management systems and ERP systems via standard APIs and event-based data feeds. AI-driven automation has been shown to reduce compliance operating costs by 30% when delivered as an overlay rather than a replacement, protecting

existing investments while adding relational intelligence as an incremental capability [21].

In the governance interface, risk scores, anomaly summaries, and evidence chains are available via an audit dashboard. Risk flags include a traceable explanation of the contributing anomaly signals, the documents that contributed to the flag, and the type of values or temporal differences that triggered it. This traceability is considered operationally necessary for auditors and is also a regulatory requirement of various compliance approaches to the explainability of automated decision-support systems [10, 11]. For related reasons, an AI in compliance paper discusses explainability as an institutional requirement for adoption because auditors need clear evidence chains to act on automated recommendations and to satisfy external regulatory review [22].

6.2 Model Governance and Feedback Loops

A deployed machine learning system requires ongoing governance to remain effective. The risk model's adaptive weight adjustment mechanism depends on reliable audit outcome feedback. Lundberg and Lee's unified framework for model prediction interpretation, which produces SHAP values that attribute each prediction to its contributing features, provides the interpretability layer needed for audit professionals to evaluate and validate the model's reasoning [17]. Without this layer, weight adjustments occur as a black-box process that audit teams cannot interrogate, reducing institutional trust in the automated system.

Model drift is a further consideration. Federated learning, as formalized by McMahan et al., offers a future-oriented solution to model maintenance across distributed enterprise environments [18]. In a federated architecture, model updates are computed locally on each participating organization's data and aggregated centrally without sharing raw financial records. This enables the risk model to improve on cross-enterprise fraud patterns while satisfying the data privacy constraints that prevent direct data sharing. The combination of SHAP-based interpretability and federated update mechanisms addresses both the transparency and adaptability requirements of a production compliance system [17, 18].

6.3 Regulatory and Privacy Constraints

According to the FATF risk-based approach guidance and Basel Committee operational risk principles, a compliance system must have controls that are appropriate to the risk of the data being processed, but data privacy regulations governing the processing of financial documents may differ by jurisdiction [10, 11]. In addition, the normalization engine's reference database and correlation graph contain or reference personal and commercially sensitive information, requiring defined access controls, audit logging of data access events, and configurable data retention policies.

Regulatory requirements for explainability can be satisfied by keeping a full chain of evidence for each case where an alert was issued. Such a requirement is common in research literature in AI and compliance, with KYC and AML screening based on algorithmic outputs requiring a human-readable explanation to pass regulatory audit [21, 22]. In cases where a human is involved in the regulatory process for automated detections, the governance layer provides a post-detection workflow to allow audit team members to review the evidence chain, override or approve the risk score, and document their rationale. The hybrid model provides the efficiency of automated detection while ensuring the necessary human oversight across the spectrum of environments enterprises operate in globally.

7. Limitations and Future Work

The proposed framework demonstrates measurable improvements in detection accuracy and operational efficiency; however, several limitations warrant consideration.

- **Entity extraction quality dependency:** The accuracy of cross-document correlation is bounded by the quality of entity extraction in the document intelligence layer. Documents with low OCR confidence, heavily abbreviated fields, or non-standard layouts, reduce the reliability of downstream correlation signals. The confidence-score gating mechanism described in Section 3.1 partially mitigates this dependency, but it remains a structural constraint.
- **Ontology maintenance overhead:** The cross-jurisdictional normalization layer requires continuous curation of tax code mappings and reporting standard translations. Regulatory changes that introduce new codes or restructure existing taxonomies impose ongoing maintenance costs that must be factored

into operational planning. Automated ontology extension mechanisms represent a productive direction for reducing this burden.

- **Dataset generalization:** Evaluation was performed on enterprise-specific data from four jurisdictions. The generalizability of the reported performance metrics to other institutional contexts, industry verticals, or jurisdictional profiles has not been independently validated. External replication on diverse datasets is needed before broad deployment claims can be made.
- **Calibration requirement for novel fraud patterns:** The adaptive weight adjustment mechanism requires historical audit feedback to achieve stable calibration. In deployment contexts where confirmed fraud labels are sparse or delayed, initial model performance may be constrained. Continuous calibration investment is required as fraud patterns evolve.

Future development directions include federated learning for cross-enterprise model improvement without raw data sharing, generative AI integration for natural language audit explanation, and extended ontology coverage for emerging regulatory frameworks. These directions build directly on the relational foundation established by the framework, reinforcing the argument that financial governance at enterprise scale requires systemic, intelligence-driven architecture.

Conclusion

The persistent inadequacy of document-level compliance validation establishes a clear case for relational intelligence as the foundation of next-generation compliance architecture, especially when confronted with fraud patterns that deliberately exploit gaps between financial records. The multi-document correlation framework presented in this article addresses that case through three technically integrated components. The graph-based entity correlation engine resolves identities, aligns timelines, and cross-validates declared values across payroll, tax, transactional, and procurement documents, surfacing inconsistencies that neither rule-based engines nor NLP extraction tools are designed to detect. The adaptive probabilistic risk model aggregates these cross-document anomaly signals into a calibrated, prioritized risk score, using feedback-driven weight adjustment to reduce false positives and progressively improve detection accuracy over time. The cross-jurisdictional normalization layer extends these capabilities to global enterprise environments, harmonizing currencies, tax code semantics, and reporting standards so that consistent correlation logic can be applied across heterogeneous regulatory contexts. Taken together, these components shift the compliance function from a reactive process of individual document verification toward a continuous and predictive governance capability. The framework further incorporates explainability mechanisms and structured audit workflows that satisfy regulatory transparency requirements without sacrificing the efficiency gains of automated detection. The limitations identified in Section 7 establish the primary directions for future investigation, including broader generalization validation, reduced ontology maintenance overhead, and model calibration under sparse feedback conditions. Financial governance at enterprise scale requires systemic, intelligence-driven architecture; the multi-document correlation framework represents a foundational step toward that standard.

References

[1] Victoria Hodge and Jim Austin, "A survey of outlier detection methodologies," Artificial Intelligence Review, 2004. Available: https://eprints.whiterose.ac.uk/id/eprint/767/1/hodgevj4.pdf

[2] Eric WT Ngai et al., "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," Decision Support Systems, 2011. Available: https://doi.org/10.1016/j.dss.2010.08.006

[3] Richard J. Bolton and David J. Hand, "Statistical fraud detection: A review," Statistical Science, 2002. Available: https://projecteuclid.org/journals/statistical-science/volume-17/issue-3/Statistical-Fraud-Detection-A-Review/10.1214/ss/1042727940.pdf

[4] Charu C. Aggarwal, "Outlier ensembles," in Outlier Analysis, Springer International Publishing, 2016. Available: https://www.charuaggarwal.net/ensembles.pdf

[5] Raghavendra Chalapathy and Sanjay Chawla, "Deep learning for anomaly detection: A survey,"

arXiv preprint arXiv:1901.03407, 2019. Available: https://arxiv.org/pdf/1901.03407
[6] Ian Goodfellow et al., “Deep Learning,” MIT Press, 2016. Available: https://synapse.koreamed.org/pdf/10.4258/hir.2016.22.4.351
[7] Jacob Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019. Available: https://aclanthology.org/N19-1423.pdf
[8] Yinhan Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019. Available: https://arxiv.org/pdf/1907.11692
[9] Anoop R. Katti et al., "Chargrid: Towards understanding 2D documents," Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018. Available: https://aclanthology.org/D18-1476.pdf
[10] Basel Committee, "Principles for the sound management of operational risk," Bank for International Settlements, 2011. Available: https://www.bis.org/publ/bcbs195.pdf
[11] Financial Action Task Force (FATF), "Guidance for a risk-based approach: The banking sector," FATF/OECD, 2014. Available: https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Guidance-Financial-Inclusion%20-Anti-Money-Laundering-Terrorist-Financing-Measures.pdf.coredownload.pdf
[12] Tianqi Chen and Carlos Guestrin, "XGBoost: A scalable tree boosting system," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. Available: https://dl.acm.org/doi/pdf/10.1145/2939672.2939785
[13] Ivan P. Fellegi and Alan B. Sunter, "A theory for record linkage," Journal of the American Statistical Association, 1969. Available: https://www2.stat.duke.edu/~rcs46/linkage/presentations/01-baiLi_FelleigSunter1969.pdf
[14] Shaoxiong Ji et al., "A survey on knowledge graphs: Representation, acquisition, and applications," IEEE Transactions on Neural Networks and Learning Systems, 2021. Available: https://www.researchgate.net/publication/351115157
[15] David Savage et al., "Detection of money laundering groups using supervised learning in networks," arXiv preprint arXiv:1608.00708, 2016. Available: https://arxiv.org/pdf/1608.00708
[16] Douglas W. Arner et al., "FinTech, RegTech, and the reconceptualization of financial regulation," Northwestern Journal of International Law and Business, 2016. Available: https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1817&context=njilb
[17] Scott M. Lundberg and Su-In Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, 2017. Available: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
[18] Brendan McMahan et al., "Communication-efficient learning of deep networks from decentralized data," Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017. Available: https://proceedings.mlr.press/v54/mcmahan17a/mcmahan17a.pdf
[19] Véronique Van Vlasselaer et al., "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," Decision Support Systems, 2015. Available: https://www.sciencedirect.com/science/article/pii/S0167923615000846
[20] Yang Liu et al., "Pick and choose: A GNN-based imbalanced learning approach for fraud detection," Proceedings of the Web Conference 2021, 2021. Available: https://dl.acm.org/doi/pdf/10.1145/3442381.3449989
[21] Adetunji Adejumo Paul and Chinonso Ogburie, "The role of AI in preventing financial fraud and enhancing compliance," GSC Advanced Research and Reviews, 2025. Available: https://www.researchgate.net/profile/Chinonso-Ogburie/publication/390300143
[22] Samia Ara Chowdhury et al., "Next generation financial security: Leveraging AI for fraud detection, compliance, and adaptive risk management," Well Testing Journal, 2025. Available: https://www.researchgate.net/publication/394087142