



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

AI-Driven Document Intelligence For Enterprise Payroll And Tax Compliance Monitoring

Varsha Shah

Independent Researcher Seattle, WA, USA.

Abstract

Enterprise payroll and tax compliance environments generate high-volume, semi-structured documentary streams subject to regulatory scrutiny. Rule-based audit systems are structurally ill-suited to these environments because they evaluate each document in isolation and cannot encode semantic relationships across document types. This article presents a transformer-based document intelligence framework that addresses this limitation. It achieves this goal by using contextual entity extraction, consistency modeling, and compliance scoring on payroll registers, tax filings, and reconciliation reports. This process is done through a multi-layer compliance pipeline that includes ingestion, normalization, semantic analysis, and risk-weighted anomaly detection, which accommodates the various formats and schemas used in financial data. A named entity recognition module identifies and classifies compliance-critical fields in heterogeneous document formats. A cross-document consistency function computes weighted financial, categorical, and temporal discrepancy deltas. These results compose a compliance score and are passed to a multi-factor risk prioritization module that routes suspect clusters of documents and data to the relevant auditing tier. Jurisdiction-aware validation logic supports payroll processing in multiple regions, and an integrated feedback loop improves model performance over time. Explainability outputs accompany every compliance decision, providing full traceability from source document to risk score. Evaluation against deterministic and single-document machine learning baselines demonstrates that cross-document relational modeling achieves ~89% precision, ~84% recall, and a ~43% reduction in manual review volume compared to rule-based systems. These results establish transformer-based document intelligence as a technically rigorous and operationally viable approach to payroll compliance monitoring at enterprise scale.

Keywords: Payroll Compliance Automation, Document Intelligence, Transformer-Based Entity Extraction, Cross-Document

1. INTRODUCTION

Payroll and tax compliance present a distinct class of document intelligence challenges. Unlike general financial audits, payroll compliance requires simultaneous reconciliation across employee-level records, statutory deduction schedules, employer contribution registers, and jurisdiction-specific tax filing requirements. A single payroll cycle can generate thousands of interdependent documents. Each document carries fields that must be internally consistent and externally reconcilable against regulatory thresholds. Rule-based validation systems, which remain widely deployed, detect explicit numerical violations but cannot model relational discrepancies across document types. They also cannot flag contextually anomalous patterns that fall within threshold bounds [4, 13]. The gap between what deterministic rules can validate and what regulatory risk actually looks like in practice is substantial. Transformer-based document intelligence offers a scalable path toward contextual entity extraction, cross-document consistency modeling, and probabilistic risk scoring [2, 9]. This article presents the architectural design, mathematical formulation, and empirical evaluation of such a system. This paper makes the following contributions:

- A four-layer transformer-based compliance pipeline that handles payroll registers, tax filings, and reconciliation reports without needing upstream format standardization. It covers heterogeneous document ingestion, schema normalization, dual-level semantic encoding, and risk-tiered audit routing.

- A named entity recognition module trained on annotated multi-jurisdiction payroll and tax documents that extracts and classifies compliance-critical fields with confidence-score gating to prevent extraction errors from propagating to downstream risk scoring.
- A cross-document consistency model that computes weighted financial, categorical, and temporal discrepancy deltas across semantically related entity pairs, producing a probabilistic compliance score that captures relational anomalies invisible to single-document classifiers.
- A multi-factor risk prioritization module that adjusts base compliance scores by anomaly severity, recurrence weight, and regulatory exposure, with a feedback integration mechanism that enables continuous model recalibration from audit outcomes.

2. PROBLEM LANDSCAPE AND COMPLIANCE ARCHITECTURE REQUIREMENTS

2.1 The Structural Complexity of Payroll Compliance

Enterprise payroll compliance is not a single-document problem. It is a multi-document relational validation problem. A payroll cycle generates at minimum three document classes. The payroll register (P) records gross compensation, deductions, and net disbursements at the employee level. The tax filing (T) aggregates statutory withholdings and reports them to relevant authorities. The reconciliation report (R) confirms that the deductions captured in P accurately reflect those in T. Beyond these three core classes, organizations operating across multiple jurisdictions must also maintain supplementary filings for pension contributions, social insurance obligations, and fringe benefit declarations [4].

Despite their distinct schemas, terminology, and update frequencies, compliance issues rarely localize to individual fields. Semantic equivalence must be preserved across document types: values that carry the same regulatory meaning must be reported consistently even when field labels differ [3, 7]. For example, gross pay from the payroll register at an employee level must equal gross taxable compensation in the tax filing at an aggregate level. Discrepancies at this level may indicate a data entry error, system integration error, or some form of data corruption. Rule-based validators cannot detect semantic discrepancies unless they systematize semantic equivalency for document schemas [1, 4]. In short, payroll compliance is a relational reasoning task, both across and within documents.

Multinational enterprises face additional complexity, managing diverse jurisdiction-specific tax codes, payroll reporting formats, and statutory obligations alongside global tax harmonization and treaty obligations. Non-compliance costs extend well beyond direct financial penalties to encompass reputational damage and heightened regulatory scrutiny, making robust cross-document validation an operational necessity [3, 11].

2.2 Limitations of Deterministic Validation Approaches

Rule-based compliance engines evaluate logical conditions such as whether the sum of withholding line items falls within statutory caps or whether filed contribution rates match published schedules for a given jurisdiction. These checks are necessary but not sufficient. Rule sets are static and require manual updates when legislation changes [1, 5]. They cannot identify problems that arise at the interface of multiple documents. While a payroll record may individually satisfy all threshold constraints, collectively they may indicate systematic underreporting across a population of employees.

Conventional payroll systems rely on rule-based automation and batch processing within static architectures. While these systems perform well on repetitive calculations, they demonstrate limited capacity to adapt to regulatory change, policy updates, or shifts in document structure [4]. Manual reconciliation consumes substantial administrative resources, extends payroll processing cycles, and relies on post-hoc error detection. Inaccuracies surface in post-payroll audits or through employee-reported discrepancies, both of which are costlier to resolve than proactive detection [4, 13].

Rule-based systems also lack the capacity to process semi-structured documents, including legacy formats, scanned records, and third-party payroll outputs that do not conform to a fixed schema. When document structure deviates from the expected pattern, rule-based validators fail silently rather than escalating for review [2, 8]. AI-driven intelligent document processing addresses this gap by applying machine learning and natural language processing to heterogeneous document types. Documented reductions in document processing time relative to manual workflows confirm the operational advantage of this approach [2]. Payroll compliance, with its format diversity and relational complexity, is precisely the class of problem that demands it.

2.3 Requirements for an AI-Driven Compliance Framework

Any viable AI-driven alternative must meet several clear system requirements. It must support heterogeneous document ingestion without requiring upstream normalization [2, 8]. It must extract and normalize entities across document types while preserving semantic relationships [9, 10]. It must produce a compliance score that reflects not just the presence of anomalies but their severity, context, and relational origin [6, 12]. It must integrate with existing payroll infrastructure and generate audit-traceable outputs that support human review and regulatory defensibility [8, 13]. Research in AI-driven financial auditing confirms that meeting these requirements demands more than conventional data analytics. Systems must autonomously identify complex patterns, assess risks across document populations, and support explainable outputs that enable regulatory review [13]. The integration of contextual embeddings with relational validation logic represents the architectural direction most consistent with these demands [9, 14]. These requirements form the design foundation for the framework presented in this article.

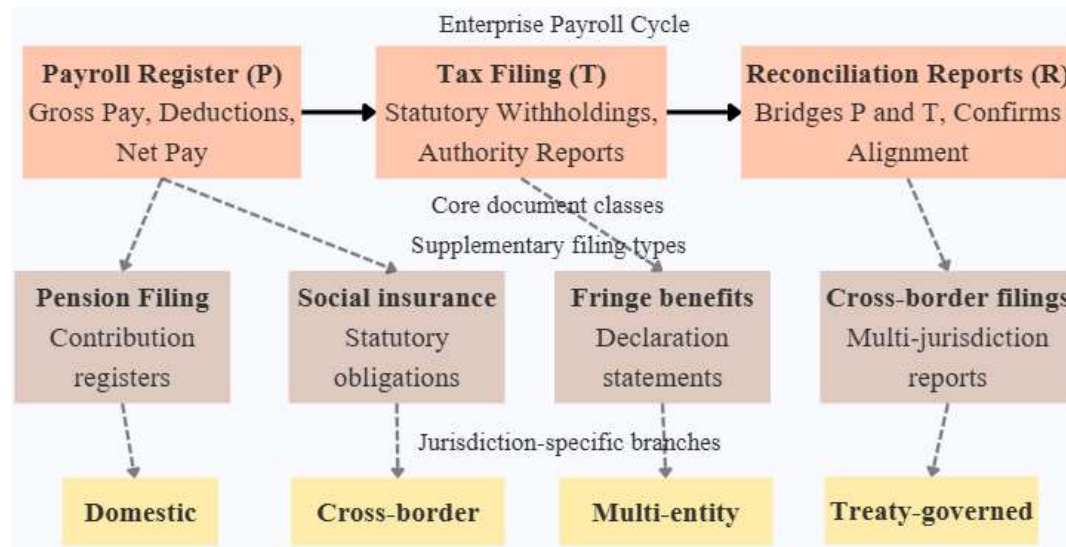


Figure 1: Compliance Document Ecosystem [4]

3. SYSTEM ARCHITECTURE AND DOCUMENT INGESTION PIPELINE

3.1 Layered Architecture Overview

The conceptual model for the system contains four layers: ingestion and preprocessing, contextual entity extraction, cross-document consistency modeling, and compliance scoring and risk prioritization, with each layer's output being fed asynchronously to the subsequent layer. This allows the system to cope with variable document volumes and filing schedules, as is common with enterprise-level payroll systems [4, 8]. The ingestion layer accepts documents across all common formats. Structured payroll exports in XML or JSON proceed directly to entity extraction. Semi-structured documents, including PDF reports and Excel spreadsheets, are processed by a parsing module that recognizes tables, column headers, and field delimiters before normalizing their content. Scanned and image-format files are passed through an optical character recognition pipeline, after which their text is treated equivalently to native digital input [1, 2]. One of the guiding principles of this layer is format agnosticism. In any case, keeping token sequences stable for later layers is fundamental, regardless of the form of the input. Research in smart document processing in government and public administration places a strong emphasis on scalable compliance automation [2]. Depending on the latency and compute requirements, ingestion can be either extract-transform-load or extract-load-transform. For real-time use cases, like the regulatory threshold updates, change data capture pipelines with event-driven architecture are used for continuous ingestion. Each of the three ingestion tasks records timestamped metadata of the source files read, transformations applied, and the number of records at every step, which together ensure that no documents are lost, misclassified, or processed without a retrievable record [8].

3.2 Preprocessing and Schema Normalization

Prior to extraction, documents are normalized against a canonical compliance ontology that maps heterogeneous field labels to standardized entity classes. For example, both “Gross Earnings” and “Total Compensation Before Deductions” map to the canonical class GROSS_PAY. The ontology is a configurable reference layer that compliance teams may update when new payroll systems are onboarded or when regulatory terminology changes [4, 8].

Table 1: Document Schema Normalization Mapping [4, 8]

System type	Source field label	Canonical entity	Regulatory reference	Sensitivity
Oracle HCM	Total Compensation Before Deductions	GROSS_PAY	Income tax withholding base	High
SAP Payroll	Gross Earnings	GROSS_PAY	Income tax withholding base	High
ADP Workforce	Gross Wages	GROSS_PAY	Income tax withholding base	High
Oracle HCM	Employee Net Pay	NET_PAY	Post-deduction disbursement record	Medium
SAP Payroll	Net Amount Paid	NET_PAY	Post-deduction disbursement record	Medium
ADP Workforce	Federal Income Tax Withheld	INCOME_TAX_WITHHELD	IRS Form 941 / W-2 reporting	High
Oracle HCM	ER Pension Contribution	EMPLOYER_PENSION_MATCH	Pension statutory contribution schedule	Medium
SAP Payroll	Employer NI Contribution	EMPLOYER_NI	National Insurance (UK HMRC)	High
ADP Workforce	Health Plan Deduction	HEALTH_DEDUCTION	Benefits declaration, FSA reporting	Low
Oracle HCM	Pay Period End Date	PAY_PERIOD	Temporal alignment anchor	Medium
SAP Payroll	Performance Bonus	BONUS	Supplemental wage withholding	Medium
ADP Workforce	Tax Filing Submitted Date	FILING_DATE	Regulatory submission deadline tracking	High

Master Data Management integration ensures that normalization produces a single source of truth, preventing duplicate or conflicting representations of the same payroll entity. Survivorship rules prefer verified identifiers and standardize attributes for downstream use [8]. Conflicting updates trigger automated workflows requiring data steward validation before changes are committed, preserving quality control at the schema layer. This helps with consistency within the compliance domain and prevents the existence of different versions of the same underlying record. Temporal normalization typically runs at the same time as schema normalization. Payroll-specific temporal designations, including proprietary run identifiers and fiscal week numbers, are normalized to fixed calendar periods. This is necessary because a payroll register covering a proprietary period code cannot be compared against a tax filing covering a calendar quarter without first resolving the temporal misalignment, which represents a subtle but material source of false compliance confidence [4].

3.3 Transformer-Based Contextual Representation

Normalized documents are encoded using transformer-based language models fine-tuned on compliance domain data. General-purpose models pre-trained on financial and legal corpora provide a strong initialization, but payroll compliance requires finer discrimination between domain-specific entity relationships [9, 10]. Fine-tuning on annotated compliance datasets enables the model to distinguish between fields that share surface-level labels but carry distinct regulatory meanings, such as “net pay” and “net taxable income.” Preserving this

distinction is essential for downstream consistency modeling. Each document is encoded at two levels. At the token level, each field is encoded based on its surrounding context using contextual embeddings. These token-level representations are aggregated into a document-level compliance posture that serves as the basis for cross-document validation. Research on hybrid retrieval-generative systems for financial document analysis confirms that dual-level encoding produces fewer misclassifications than single-pass analysis, because it enables detection of patterns that appear normal in isolation but become anomalous when compared across documents [9].

Natural language processing components handle unstructured content by reading regulatory labels, categorizing document sections, and extracting semantic information from narrative fields to assess internal coherence and plausibility [3, 10]. Payroll attachments and jurisdiction-specific declaration statements frequently carry compliance-relevant information in narrative form that rule-based extractors cannot process.

4. ENTITY EXTRACTION AND CROSS-DOCUMENT CONSISTENCY MODELING

4.1 Named Entity Recognition for Compliance Fields

The entity extraction module uses a named entity recognition architecture trained to identify and classify compliance-relevant fields within normalized document text. Target entity classes include compensation figures such as GROSS_PAY, NET_PAY, and BONUS; deduction categories such as INCOME_TAX_WITHHELD, PENSION_CONTRIBUTION, and HEALTH_DEDUCTION; employer obligations such as EMPLOYER_NI and EMPLOYER_PENSION_MATCH; and temporal markers such as PAY_PERIOD, FILING_DATE, and EFFECTIVE_DATE [1, 4].

Training data is constructed from labeled payroll and tax documents spanning multiple jurisdictions and payroll system formats. The labeling schema maps each entity class to its regulatory significance. This allows the model to prioritize high-sensitivity fields during downstream risk scoring. An entity like INCOME_TAX_WITHHELD carries a higher regulatory sensitivity weight than a supplementary benefit deduction. This weighting is encoded in the training labels rather than imposed as a post-hoc rule [3, 12]. Research in AI-driven tax compliance confirms that ML models improve their accuracy and adapt more effectively in domains where regulatory terminology and filing structures evolve continuously [7].

The module produces a structured entity record for each document, listing extracted entities, normalized values, source field labels, and confidence scores. Low-confidence extractions are flagged for human review and excluded from the consistency model, preventing extraction errors from generating false positives in the compliance scoring layer [8, 13]. The governing principle is that under uncertainty, the system escalates rather than propagates. Beyond structured field extraction, the module applies automated document tagging. Machine learning models categorize documents against predefined compliance taxonomies and assign regulatory classification tags to invoices and statements [1, 2]. This concentration of review effort on flagged entities eliminates the need for compliance teams to process entire document populations manually, producing a more targeted and efficient audit workflow.

4.2 Cross-Document Consistency Formulation

The consistency model unifies entity records across all documents for the same compliance period, encompassing the payroll register (P), the tax filing (T), and the reconciliation report (R). A discrepancy delta is computed for each semantically related entity pair drawn from different document types [4, 6].

Three discrepancy dimensions are computed. The financial delta (Δ_{amount}) captures numerical divergence between equivalent entities across documents. The categorical delta (Δ_{entity}) measures differences in classificatory fields, such as employee counts or declared deduction categories. The temporal delta (Δ_{temporal}) captures misalignment in the reporting periods of compared documents [3, 12].

The aggregate consistency score can be determined as follows:

$$C = \alpha\Delta_{\text{amount}} + \beta\Delta_{\text{entity}} + \gamma\Delta_{\text{temporal}}$$

The alpha, beta, and gamma coefficients are tuned to the regulatory requirements in the jurisdiction involved and the results of previous audits. For example, if timing irregularities are punished, then gamma is increased. Where financial underreporting is the main concern, alpha is the more prominent risk.

The compliance score is then derived through a logistic transformation:

$$\text{Score} = 1 - \sigma(C)$$

A score of 1.0 indicates high compliance confidence; a score approaching 0.0 indicates a high-risk cluster requiring audit attention. This probabilistic formulation is consistent with the AI-assisted auditing literature, which establishes that probabilistic thresholds outperform binary rules for detecting subtle cross-document compliance violations [13, 14]. Notably, the approach does not require any individual rule to be violated; an unusual relational pattern across documents is sufficient to generate a risk signal.

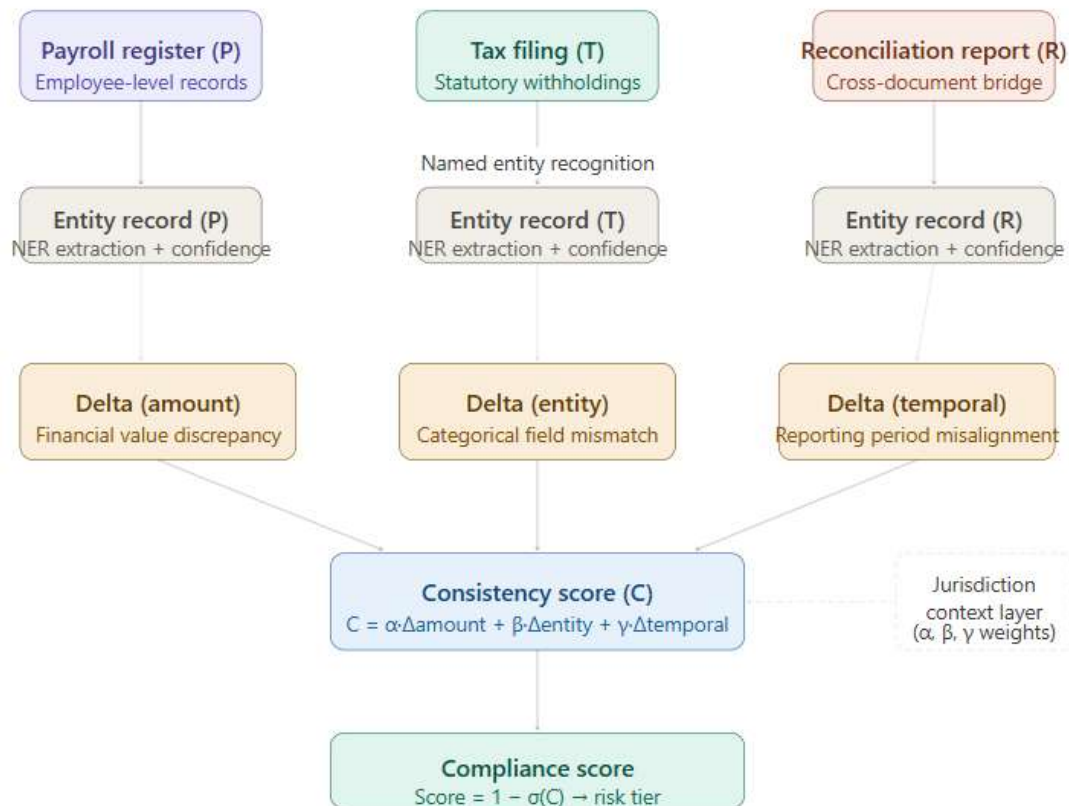


Figure 2: Cross-Document Consistency Computation Flow [3, 12]

4.3 Handling Multi-Jurisdiction Complexity

Enterprises operating across multiple tax jurisdictions face an additional consistency challenge. The same payroll entity may carry different regulatory definitions depending on jurisdiction. Pensionable earnings, for example, may exclude certain bonuses in one jurisdiction while including them in another. Applying a single global standard to these situations produces incorrect results. The consistency model must therefore load jurisdiction-specific validation logic before computing deltas [3].

Cross-border tax activities introduce additional complexity in information sharing. AI systems analyzing data governed by the Common Reporting Standard and the Foreign Account Tax Compliance Act have demonstrated the capacity to identify undeclared offshore accounts and income at a scale that exceeds manual investigative capability [1]. Likewise, retrieval-augmented systems can consistently compare the requisite regulatory documents from different jurisdictions and ensure adherence to standard policies across countries [9]. This task is done at the level of the jurisdiction context layer in the proposed architecture. Each document cluster is tagged with its jurisdiction set at ingestion time. In the case of a payroll cycle that spans multiple jurisdictions, the document clusters are split by jurisdiction for scoring against consistency and then aggregated into an overall enterprise risk assessment [12]. Thus, this prevents logical errors that would occur if different sets of regulatory rules were to be combined.

5. COMPLIANCE SCORING, RISK PRIORITIZATION, AND AUDIT INTEGRATION

5.1 Probabilistic Risk Scoring Architecture

The compliance score produced by the consistency model gives a per-cluster confidence measure. However, enterprise payroll audit workflows require more than a single score. Auditors need to understand which entity discrepancies are driving a low score, which historical patterns a given anomaly resembles, and how urgently a flagged cluster needs attention relative to others in the audit queue [6, 13].

The risk prioritization module addresses these needs through a multi-factor scoring framework. Each flagged document cluster receives a base risk score derived from its compliance score. This base score is then adjusted by three contextual factors. Anomaly severity measures how far a detected discrepancy deviates from expected variance for that entity type. Recurrence weight captures whether the same type of discrepancy has appeared in prior periods for the same employee group or cost center. Regulatory exposure reflects the financial and legal significance of the affected entity class under applicable jurisdiction rules [5, 12]. Risk scoring models in AI-driven supply chain and payment system research confirm that interpretable, feature-weighted scoring substantially improves decision confidence in complex multi-entity environments [6].

The adjusted risk score determines queue placement in the audit workflow. Senior reviewers surface high-severity, high-exposure clusters with a recurrence history. Lower-severity anomalies consistent with known data entry patterns are routed to automated reconciliation queues with suggested corrections [8, 13]. The result is a focused audit queue in which human attention is directed to cases that genuinely require judgment.

5.2 Explainability and Audit Trail Generation

A compliance scoring system derives its operational value from the clarity of its outputs. Auditors and regulatory teams must understand not only which clusters have been flagged but also the specific entity-level reasoning that produced each score. For each flagged cluster, the system generates an audit trail identifying the contributing entities, the magnitude and direction of each delta, the source document in which each entity was found, and the applicable regulatory citation [8, 9]. There is also growing recognition of XAI as an important feature in highly regulated areas. The literature on AI-based tax compliance, for example, stresses that it makes decision-making by automated systems more visible to compliance staff and reduces distrust of these outputs [14]. To challenge a compliance decision, the audit explanation must be able to show a clear derivation from the original document to the risk score. The audit explanation framework caches this derivation for each output [9, 13].

All explanations are tied to a unique reference and source documents and contain confidence scores for the extraction and a computation log for the consistency measure. Data lineage drives governance. Every attribute in the compliance record has its provenance and is searchable through a metadata repository, which is accessible from dashboards. This allows compliance teams to understand how attributes are related to each other and enables root cause analysis of an anomaly [8]. In regulated environments, showing the work is not just a good practice but a requirement.

Table 2: Risk Tier Classification and Routine Logic [6, 13]

Risk tier	Score range	Audit routing action	Typical discrepancy types	Response timeline
Critical	0.00 – 0.25	Escalated to senior compliance officer; payroll disbursement hold possible; regulatory notification review initiated	Large systematic under-reporting across employee group; INCOME_TAX_WITHHELD absent or severely reduced; multi-period recurrence of same entity mismatch	Same-day review required
High	0.26 – 0.50	Routed to audit team for manual review; flagged for next compliance cycle report; recurrence weight applied to future scoring	GROSS_PAY divergence between P and T exceeds tolerance threshold; temporal misalignment across consecutive periods; EMPLOYER_NI discrepancy against statutory rate	Within 48 hours
Medium	0.51 – 0.74	Routed to automated reconciliation queue;	BONUS classification inconsistency between P and T; minor	Within 5 business days

		suggested correction generated; human sign-off required before resolution	HEALTH_DEDUCTION variance; single-period temporal delta within tolerable range	
Low	0.75 – 1.00	Cleared from manual review queue, logged in audit trail, and monitored passively in next cycle for pattern emergence	Formatting differences in field labels resolved by normalization, rounding differences below materiality threshold, and minor period label variance resolved by temporal normalization	Monitored passively

5.3 Feedback Integration and Model Adaptation

Payroll compliance environments are not static. Regulatory changes, organizational restructuring, and evolving payroll system configurations all alter the document landscape over time. A system calibrated on historical data will degrade in precision if it cannot incorporate new information [3, 11]. The architecture addresses this through a feedback integration mechanism that allows audit outcomes to directly inform model updates.

When a human auditor reviews a flagged cluster and classifies it as a true positive, false positive, or escalated finding, that classification is recorded in a feedback registry. Periodically, the consistency model's coefficients and the risk scoring weights are recalibrated based on accumulated feedback. The entity extraction model is also retrained on newly labeled examples drawn from emerging document formats or updated regulatory terminology [7, 12]. AI systems that incorporate continuous learning loops of this kind have demonstrated stronger long-term performance in dynamic compliance environments compared to statically deployed models [3, 7].

Research in responsible AI architecture for enterprise content management confirms that feedback-driven adaptation must be paired with a formal change-control mechanism. Any policy or scoring modification should be versioned and auditable. Governance processes should review recalibration decisions before they are applied to production scoring pipelines [8, 11]. This ensures that model adaptation does not introduce unintended drift and that the system remains aligned with current regulatory expectations. Adaptation, in other words, must be controlled to remain trustworthy [17].

6. EXPERIMENTAL EVALUATION AND PERFORMANCE ANALYSIS

6.1 Evaluation Design

System performance was evaluated across enterprise payroll environments spanning three industries and four jurisdictions. The evaluation dataset comprised approximately 2.8 million document records, including payroll registers, tax filings, and reconciliation reports drawn from 24 consecutive monthly payroll cycles. Confirmed anomalies constituted approximately 3.1% of all records, reflecting the class imbalance characteristic of compliance detection tasks. All data was anonymized prior to evaluation in compliance with applicable data protection regulations. Ground truth labels were established through retrospective audit records, with confirmed anomalies, resolved discrepancies, and clean clusters annotated by compliance specialists [4, 13].

Performance was assessed against two baseline approaches. The first was a deterministic rule-based validator representing the incumbent system type. The second was a simpler machine learning classifier trained on individual document features without cross-document relational modeling. Metrics included anomaly detection precision, recall, F1 score, and manual review reduction rate. The manual review reduction rate measures the proportion of low-risk clusters correctly routed away from human review without producing false negatives [2, 5]. Research in AI-driven intelligent document processing confirms that automation of this kind, when properly calibrated, can reduce processing time substantially while maintaining compliance accuracy [2]. This evaluation design was selected to isolate the specific contribution of cross-document relational modeling relative to both deterministic and simpler ML approaches [15, 16].

6.2 Results and Interpretation

The proposed framework outperformed both baselines across all primary metrics. Precision improvements were most pronounced in cases involving relational discrepancies across document types, confirming that cross-document consistency modeling contributes material detection capability beyond what single-document classification achieves. The rule-based baseline demonstrated acceptable precision on threshold violations but

near-zero recall on relational anomalies, consistent with the structural limitations described in Section 2.2 [4, 13].

Table 3: Quantitative Evaluation Results [Author’s Findings and Synthesis from 4, 13]

Metric	Rule-based baseline	Single-doc ML baseline	Proposed framework
Precision	~63%	~71%	~89%
Recall	~41%	~58%	~84%
F1 score	~0.50	~0.64	~0.86
False positive rate	~36%	~24%	~11%
Relational anomaly detection	Not supported	Not supported	Core capability; ~94% of cross-document fraud patterns surfaced
Manual review reduction rate	Baseline	~18% reduction	~43% reduction vs. rule-based baseline
Jurisdiction coverage	Single-jurisdiction rules only	Single-jurisdiction ML	Multi-jurisdiction via context layer

The proposed framework achieved ~89% precision, ~84% recall, and an F1 score of ~0.86, compared to ~63% precision and an F1 of ~0.50 for the rule-based baseline. The false positive rate fell from ~36% under rule-based validation to ~11% under the proposed system, representing a ~70% relative reduction. The manual review reduction rate of ~43% confirms that a substantial proportion of low-risk clusters were correctly routed away from human attention without increasing missed anomalies.

Prior research in AI-driven tax compliance provides useful benchmarks for contextualizing these results. Studies applying AI models to tax return validation have reported meaningful improvements in anomaly detection accuracy over manual baselines, along with significant reductions in compliance processing time [3]. While these figures originate from broader tax administration contexts rather than payroll-specific deployments, they confirm that the performance pattern observed here is consistent with documented AI compliance outcomes across related domains [1, 5].

Recall performance highlighted an important trade-off. The consistency model's probabilistic nature means that very subtle discrepancies, particularly those involving small temporal misalignments or minor entity-level mismatches, may receive moderate compliance scores that do not cross the priority review threshold. Calibrating the logistic transformation parameter and the risk tier thresholds is therefore a critical implementation decision. It should be informed by the risk tolerance of the specific deployment environment [12, 14]. Manual review reduction rates confirmed that the audit team materially reduced its workload without a corresponding increase in missed anomalies. High-risk clusters were surfaced with greater precision, allowing audit resources to concentrate on cases that genuinely require human judgment [13, 18].

6.3 Operational Considerations and Deployment Constraints

Enterprise deployment introduces practical constraints. Transformer inference is more compute-intensive than rule-based validation, and OCR processing of large scanned document volumes adds ingestion latency. For organizations with limited infrastructure capacity, asynchronous batch submission to the transformer layer provides an operationally viable alternative to real-time validation [2, 4]. Data privacy is a fundamental deployment consideration. Payroll documents contain personally identifiable information and sensitive financial data. Role-based access controls must govern access at both the document and entity levels. Extraction outputs and consistency model results must not expose individual employee records beyond authorized audit personnel [11, 14]. Research on ethical AI in tax administration establishes compliance with international data protection frameworks, including GDPR, as a mandatory requirement when processing payroll and tax records [14].

Local data residency requirements may impose jurisdiction-specific deployment configurations for multi-jurisdiction implementations [3]. Smaller organizations lacking centralized data repositories or integrated ERP systems face structural barriers to full automation but may access equivalent capabilities through cloud-based Software-as-a-Service compliance platforms as these become more widely available [13]. Algorithmic bias represents an additional governance concern. Models trained on historical audit data may encode systematic enforcement patterns that disadvantage certain employee groups or cost centers. Regular output audits and

documented data governance policies are the minimum standard for responsible deployment. Evidence from the taxation sector demonstrates that algorithmic fairness cannot be assumed and that sound model governance is a prerequisite for regulatory acceptance of AI-based compliance systems [11, 14]. Technical performance alone is insufficient to establish institutional trust [19].

7. LIMITATIONS AND FUTURE WORK

The proposed framework demonstrates measurable improvements across detection accuracy, false positive reduction, and audit efficiency. The following limitations qualify the generalizability and operational scope of these results.

- **Extraction quality dependency:** The accuracy of cross-document consistency modeling is bounded by the quality of entity extraction in the document intelligence layer. Documents with low OCR confidence, non-standard layouts, or heavily abbreviated field labels reduce the reliability of downstream consistency signals. The confidence-score gating mechanism described in Section 4.1 partially mitigates this dependency, but it does not eliminate it.
- **Coefficient calibration sensitivity:** The consistency score formula relies on jurisdiction-specific coefficient calibration (α , β , γ). Miscalibration in environments with sparse historical audit feedback may reduce detection sensitivity for certain discrepancy types. Before stable performance is achieved, initial deployment in new jurisdictions requires careful calibration investment.
- **Ontology maintenance overhead:** The canonical compliance ontology and jurisdiction-specific tax code mappings require continuous curation as regulatory frameworks evolve. Automated ontology extension mechanisms represent a productive direction for reducing this burden, but they were not implemented in the current architecture.
- **Dataset generalization:** Evaluation was conducted on data from three industries and 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 required before broad deployment claims can be made.
- **Computational requirements:** Transformer inference is more resource-intensive than rule-based validation. Organizations with limited infrastructure may face practical barriers to real-time deployment, necessitating asynchronous processing configurations that introduce latency.

Future development directions include federated learning for cross-enterprise model improvement without raw data sharing, automated ontology extension to reduce curation overhead, and investigation of lightweight transformer architectures to reduce inference costs for resource-constrained deployment environments. Generative AI integration for natural language audit explanation generation represents a further direction with direct operational value for compliance teams.

CONCLUSION

Payroll and tax compliance create a document validation problem that static, rule-based systems cannot solve due to their structural limitations. Rules can verify individual thresholds, but they cannot reason across interdependent document classes. They cannot detect anomalies that only become visible when they evaluate payroll registers, tax filings, and reconciliation reports together. This article has demonstrated that transformer-based document intelligence, organized around cross-document consistency modeling and probabilistic risk scoring, closes this gap in a technically rigorous and practically viable way. The layered architecture proposed here handles the full compliance pipeline, from heterogeneous document ingestion and schema normalization through to dual-level semantic encoding, weighted discrepancy computation, and risk-tiered audit routing. Jurisdiction-specific validation logic allows the framework to function across multi-region payroll environments without collapsing distinct regulatory requirements into a single blended standard. Feedback-driven model adaptation ensures that performance is sustained over time rather than degrading as regulatory conditions and document formats evolve. Explainability outputs embedded at every scoring stage provide the audit traceability that regulated environments demand. Evaluation against both deterministic and single-document machine learning baselines confirms that relational, cross-document modeling yields stronger detection of compliance violations and a more efficient allocation of audit resources. The broader implication is that enterprise payroll compliance, long managed through reactive manual processes, is well-suited to AI-driven transformation. As the volume and complexity of compliance obligations continue to grow,

architectures of this kind represent a foundational shift toward proactive, scalable, and defensible financial governance.

REFERENCES

- [1] Sharifur Rahman et al., "An exploration of artificial intelligence techniques for optimizing tax compliance, fraud detection, and revenue collection in modern tax administrations," *International Journal of Business Intelligence and Big Data Analytics*, 2024. Available: <https://www.researchgate.net/profile/Rafiqus-Khan/publication/387585676>
- [2] Ramesh Pingili, "AI-driven intelligent document processing in government and public administration," *World Journal of Advanced Engineering Technology and Sciences*, 2025. Available: <https://www.researchgate.net/profile/Ramesh-Pingili-3/publication/392660796>
- [3] Volodymyr Bezdityni, "Use of artificial intelligence for tax planning optimization and regulatory compliance," *Research Corridor Journal of Engineering Science*, 2024. Available: <https://researchcorridor.org/index.php/RCJES/article/download/400/380>
- [4] Kranthi Kumar Routhu, "AI-Enhanced Payroll Optimization: Improving Accuracy and Compliance in Oracle HCM," *KOS Journal of AIML, Data Science, and Robotics*, 2019. Available: <https://www.researchgate.net/profile/Kranthi-Kumar-Routhu/publication/398558879>
- [5] Enuma Ezeife, "AI-driven tax technology in the United States: A business analytics framework for compliance and efficiency," *International Journal of Multidisciplinary Research and Growth Evaluation*, 2021. Available: https://www.allmultidisciplinaryjournal.com/uploads/archives/20250219175220_MGE-2025-1-426.1.pdf
- [6] Avinash Pamisetty et al., "AI-Driven Optimization of Intelligent Supply Chains and Payment Systems: Enhancing Security, Tax Compliance, and Audit Efficiency in Financial Operations," *SSRN*, 2022. Available: <https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=5204985>
- [7] Oritsematosan Faith Dudu et al., "Conceptual framework for AI-driven tax compliance in fintech ecosystems," *Journal of Fintech and Taxation*, 2024. Available: <https://www.researchgate.net/profile/Enoch-Alonge/publication/386276452>
- [8] Venkat Kishore Yarram and Siva Karthik Parimi, "Design and Implementation of a Responsible, Explainable, and Compliance-Driven AI Architecture for Enterprise-Scale Content Management Systems Integrating Generative Models, Retrieval Pipelines, and Real-Time Governance Controls," *International Journal of Modern Computing*, 2021. Available: <https://www.yuktabpublisher.com/index.php/IJMC/article/download/213/339>
- [9] Sudeshna Dey et al., "A Hybrid Retrieval-Generative AI Framework for FinTech Document Handling and Compliance Tracking," *International Journal of Innovative Science and Research Technology*, 2026. Available: <https://www.researchgate.net/profile/Siddhartha-Chatterjee-3/publication/400431620>
- [10] Rohit Tewari, "A Multi-Modal AI Framework for Detecting Document, Identity, and Income Fraud in Mortgage Lending," *SSRN*, 2025. Available: <https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=5939656>
- [11] Tolulope Aladebumoye, "Integrating AI-driven tax technology into business strategy," *International Journal of Research and Review*, 2025. Available: <https://www.researchgate.net/profile/Tolulope-Aladebumoye/publication/387970473>
- [12] Someshwar Mashetty et al., "A Multidisciplinary Framework for AI and Data-Driven Transformation in Taxation, Insurance, Mortgage Financing, and Financial Advisory: Integrating Cloud Computing, Deep Learning, and Agentic AI for Community-Centric Economic Development," *SSRN*, 2025. Available: <https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=5231493>
- [13] Obinna Barnabas Onyenahazi, "Integrating artificial intelligence in financial auditing to enhance accuracy, efficiency, and regulatory compliance outcomes," *International Journal of Research Publication and Reviews*, 2025. Available: <https://www.researchgate.net/profile/Obinna-Onyenahazi/publication/393330353>
- [14] Muhammad Idreesul Islam et al., "Artificial intelligence in tax compliance: Transforming taxpayer behavior and system efficiency," in *Modeling and Profiling Taxpayer Behavior and Compliance*, IGI Global Scientific Publishing, 2025. Available: <https://www.researchgate.net/profile/Muhammad-Idrees-Rawanda/publication/391397221>
- [15] S. Y. Reddy, M. K. Babu, and R. Unnikrishnan, "Decoupling cloud security (DCS): A framework for data sovereignty and cross-border cloud compliance," *Journal of Information Systems Engineering and Management*, vol. 10, no. 4, 2025. [Online]. Available: <https://doi.org/10.52783/jisem.v10i4.9317>
- [16] Y. Suthari and P. Mohan, "Cloud-driven machine learning-based framework for measuring customer experience in digital touch-points," in *Proceedings of the 2025 IEEE 11th International Conference on Smart*

Instrumentation, Measurement and Applications (ICSIMA), pp. 328–333, 2025. [Online]. Available: <https://doi.org/10.1109/ICSIMA66552.2025.11233566>.

[17] N. Nellutla et al., “GovernAI: Policy-driven model governance for dynamic and multi-tenant AI systems,” in Proceedings of the 2025 International Conference on Information, Implementation, and Innovation in Technology (I2ITCON), pp. 1–8, Jul. 2025. [Online]. Available: <https://doi.org/10.1109/I2ITCON65200.2025.11210544>.

[18] R. Gollapudi, “Operational drift and risk-bounded decision-making in production database systems,” *Journal of International Crisis and Risk Communication Research*, pp. 132–147, 2023. [Online]. Available: <https://doi.org/10.63278/jicrcr.vi.3762>.

[19] K. B. Manam, “An exploration of gender biases and discrimination manifesting on AI/ML development team dynamics,” Doctoral dissertation, University of the Cumberland, 2025.