LLM-Enabled Multi-Document Correlation for Financial Compliance and Fraud Detection

Conf42 Large Language Models (LLMs) 2026

Varsha Shah, Enterprise Technical Architect

Agenda

The Compliance Challenge

Enterprise financial compliance faces compounding pressures that legacy systems were never designed to handle.

Multi-Jurisdictional Regulations

Overlapping and conflicting regulatory frameworks across geographies create inconsistencies that are difficult to reconcile manually.

Growing Data Volumes

Payroll, tax, procurement, and transactional records scale faster than audit capacity, creating blind spots.

Sophisticated Fraud Patterns

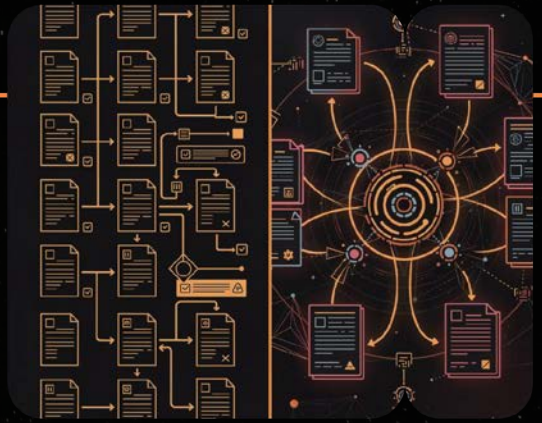
Modern fraud deliberately exploits gaps *between* datasets — invisible to any single-document analysis.

Why Rule-Based Systems Fall Short

Traditional compliance engines operate on fixed logic applied to individual documents. They cannot reason across datasets, adapt to novel patterns, or interpret unstructured content.

Before

Rule-based, document-level, reactive checks



After

AI-driven, cross-document, predictive analysis

No Cross-Document Reasoning

Anomalies that only emerge when linking payroll to procurement to tax filings go undetected.

High False Positive Rate

Rigid thresholds flag legitimate transactions, creating unsustainable manual review burdens.

Static Logic

Rules do not learn. New fraud typologies require manual rule updates, always lagging the threat.

FRAMEWORK OVERVIEW

A Three-Component AI Architecture

The architecture transforms heterogeneous enterprise records into standardized compliance intelligence through entity correlation, risk assessment, and jurisdiction-aware normalization.

Enterprise Data Sources

🗂️ Payroll Records	🧾 Tax Filings	📦 Procurement Data	💳 Transactional Data

↓ Data Ingestion & Parsing Layer

Document Processing · Extraction · Validation · Standardization → Structured Enterprise Data

COMPONENT 1

🔗 Graph-Based Entity Correlation

- Cross-document entity linking & relationship graph construction
- Hidden connection discovery & anomaly pattern detection
- Output: Connected Entity Network & Relationship Intelligence

Identifies relationships between individuals, organizations, vendors, transactions, and financial records.

COMPONENT 2

🎯 Probabilistic Risk Scoring

- Multi-signal aggregation & confidence-based scoring
- Risk prioritization with continuous learning from audit feedback
- Output: Risk-Ranked Entities, Transactions & Investigation Priorities

Transforms detected patterns and anomalies into measurable compliance risk.

COMPONENT 3

🌐 Jurisdictional Normalization Layer

- Currency normalization & tax code mapping
- Regulatory framework alignment & cross-border comparability
- Output: Standardized Compliance Context & Jurisdiction-Aware Assessments

Normalizes risk findings across diverse regulatory environments.

→ Entity Relationships & Anomaly Signals

→ Risk Scores & Investigation Priorities

→ Normalized Compliance Intelligence

Compliance Intelligence Outputs

🚨 Compliance Alerts	📋 Audit-Ready Reports	🔍 Investigation Dashboard	📊 Regulatory Insights
📡 Compliance Monitoring			

🔄 *Continuous Learning & Audit Feedback loop refines risk models and entity correlations over time.*

SOLUTION FLOW

End-to-End Solution Workflow

How enterprise financial data is transformed into actionable compliance intelligence — from raw records to regulatory-ready outputs.

🗂️ Data Ingestion

Payroll, Tax, Procurement & Transactional records ingested, parsed, validated, and standardized into structured enterprise data.

🔗 Entity Correlation

Cross-document entity linking builds a unified relationship graph — surfacing hidden connections across vendors, individuals, and transactions.

🎯 Risk Scoring

Multi-signal anomaly patterns aggregated into calibrated, confidence-weighted risk scores. High-risk cases prioritized for investigation.

🌐 Jurisdictional Normalization

Risk findings normalized across currencies, tax codes, and regulatory frameworks — enabling consistent cross-border compliance assessment.

🧠 LLM Intelligence

Large Language Models generate natural-language explanations, surface contextual insights, and produce audit-ready narrative summaries.

📊 Compliance Outputs

Compliance Alerts, Audit-Ready Reports, Investigation Dashboard, Regulatory Insights, and Continuous Monitoring delivered to stakeholders.

Each stage produces structured outputs consumed by the next — creating a tightly coupled, end-to-end compliance intelligence pipeline.

🔄 *Audit outcomes feed back into risk models, enabling continuous learning and model refinement over time.*

ARCHITECTURE DETAIL

Component Deep Dive: How the Three Layers Work Together

Each component is purpose-built for a distinct intelligence function — together they form a closed-loop compliance reasoning system.

🔗 Graph-Based Entity Correlation

Purpose: Create a unified cross-document view of all entities and their relationships.
Key Responsibilities:

- Cross-document entity linking
- Relationship graph construction
- Hidden connection discovery
- Anomaly pattern detection

Inputs: Structured Enterprise Data (Payroll, Tax, Procurement, Transactional)
Outputs: Connected Entity Network · Relationship Intelligence
Business Value: Surfaces fraud rings, shell company networks, and multi-party collusion invisible to single-document systems.

🎯 Probabilistic Risk Scoring

Purpose: Quantify compliance risk using multiple weak signals rather than binary rule triggers.
Key Responsibilities:

- Multi-signal aggregation
- Confidence-based risk scoring
- Investigation prioritization
- Continuous learning from audit feedback

Inputs: Entity Relationship Intelligence & Anomaly Signals
Outputs: Risk-Ranked Entities · Risk-Ranked Transactions · Investigation Priorities
Business Value: Reduces false positives, surfaces high-confidence risks, and continuously improves through audit feedback loops.

🌐 Jurisdictional Normalization Layer

Purpose: Ensure consistent compliance assessment across diverse regulatory environments.
Key Responsibilities:

- Currency normalization
- Tax code mapping & harmonization
- Regulatory framework alignment
- Cross-border comparability

Inputs: Risk Intelligence & Entity Assessments
Outputs: Standardized Compliance Context · Jurisdiction-Aware Assessments
Business Value: Enables global enterprises to apply consistent compliance standards across 50+ jurisdictions without manual reconciliation.

The three components are tightly coupled — entity intelligence feeds risk scoring, risk scoring informs normalization, and audit outcomes continuously refine all three layers.

EXECUTIVE SUMMARY

Business Value & Compliance Impact

The architecture delivers measurable compliance outcomes across detection, monitoring, reporting, and investigation — at enterprise scale.

🔍 Early Anomaly Detection

Identifies compliance risks weeks before they escalate — by correlating signals across payroll, tax, procurement, and transactional records simultaneously.

Business outcome: Reduced regulatory exposure & penalty risk

📄 Cross-Document Compliance Monitoring

Continuously monitors relationships between documents that legacy systems treat in isolation — closing the gap where fraud and non-compliance hide.

Business outcome: End-to-end compliance coverage across document boundaries

🌐 Cross-Border Risk Assessment

Normalizes risk findings across currencies, tax codes, and regulatory frameworks — enabling consistent global compliance without manual reconciliation.

Business outcome: Consistent compliance standards across 50+ jurisdictions

✅ Audit Readiness

Generates structured, narrative audit reports with LLM-powered explanations — reducing audit preparation time and improving regulator confidence.

Business outcome: Faster audit cycles & stronger regulatory relationships

📊 Regulatory Reporting

Produces jurisdiction-aware compliance reports aligned to local regulatory requirements — reducing manual reporting effort and error risk.

Business outcome: Automated, accurate regulatory submissions

🎯 Investigation Prioritization

Risk-ranked entity and transaction lists focus investigator attention on highest-confidence cases — maximizing team efficiency and detection ROI.

Business outcome: Higher investigation success rates with fewer resources

~3M

Anonymized records evaluated

Across payroll, tax, procurement, and transactional datasets

3

Tightly coupled AI components

Delivering end-to-end compliance intelligence

50+

Jurisdictions supported

Through normalization and regulatory framework alignment

By enabling continuous cross-document analysis and learning from audit feedback, this architecture transforms compliance from a reactive obligation into a proactive, intelligence-driven capability.

Example: Employment Tax Compliance Correlation

Input Records

Payroll System

Employee payroll processed for March payroll cycle

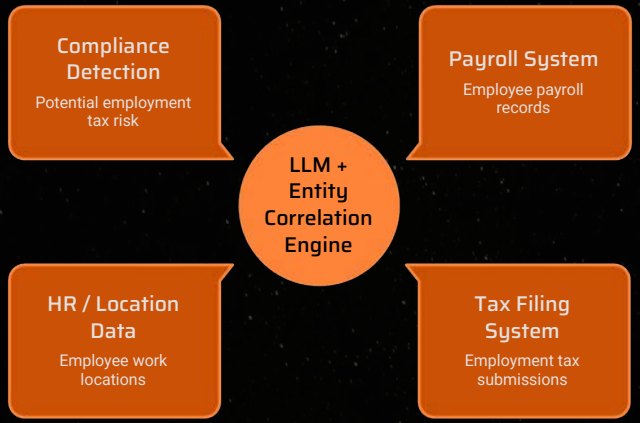
Tax Filing System

Employment tax filing submitted after jurisdictional due date

HR / Work Location Data

Employee location classified under different tax jurisdiction

AI Correlation Layer



LLM + Entity Correlation Engine

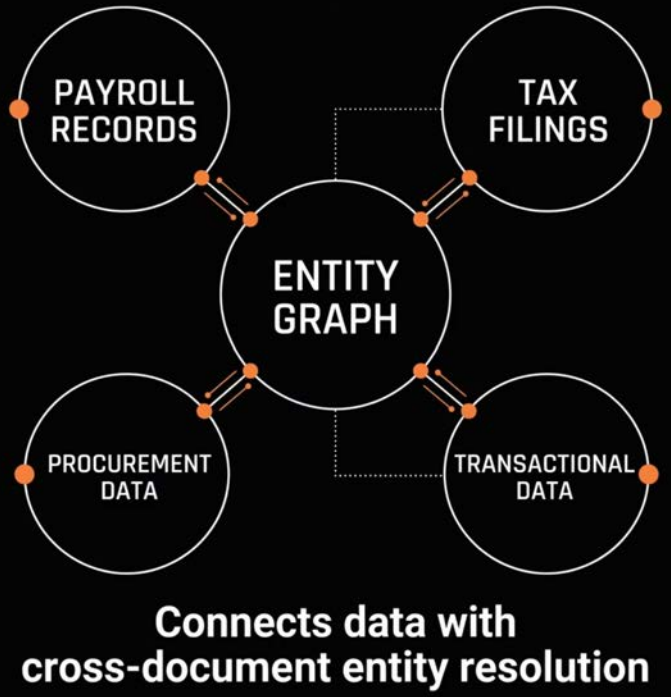
Detection

⚠️ Compliance Risk Detected

Late tax filing identified due to cross-jurisdictional mismatch between employee location, payroll records, and filing timelines.

"Individually valid records appeared compliant in isolation, but cross-document correlation identified delayed tax filing exposure across jurisdiction-specific reporting requirements."

Component 1: Graph-Based Entity Correlation Engine



The correlation engine constructs a unified entity graph by resolving identities and relationships across heterogeneous financial data sources.

- Entities (individuals, vendors, accounts) are linked across all document types
- Graph traversal surfaces hidden relationships invisible at the document level
- LLMs assist in resolving ambiguous entity references in unstructured text

Component 2: Adaptive Probabilistic Risk Model

Rather than binary rule triggers, the risk model aggregates multiple weak anomaly signals into calibrated, interpretable risk scores.

Signal Aggregation

Anomaly indicators from each data domain are weighted and combined into a composite score.

Probabilistic Calibration

Scores reflect true likelihood of fraud, reducing false positives from threshold-based triggers.

Audit Feedback Loop

Confirmed audit outcomes continuously retrain the model, improving precision over time.

Component 3: Jurisdictional Normalization Layer

The Problem

Cross-border financial analysis is undermined by inconsistent currencies, tax structures, and reporting standards. Direct comparison without normalization produces unreliable signals.

The Solution

A dedicated normalization layer standardizes all incoming records to a canonical representation before correlation and risk scoring, enabling reliable cross-jurisdictional analysis across all four supported jurisdictions.

Normalization Dimensions

- Currency conversion with period-accurate exchange rates
- Tax code harmonization across jurisdictions
- Reporting standard alignment (GAAP, IFRS, local frameworks)
- Date, entity classification, and transaction type standardization

LLM INTEGRATION

Role of Large Language Models

LLMs are not a wrapper — they are deeply integrated to handle the tasks that structured systems cannot.

Semantic Understanding

Interpreting intent and meaning in unstructured financial documents, contracts, and audit notes beyond keyword matching.

Contextual Linking

Connecting references across documents where entity names, structures, or formats differ.

Interpretation Layer

Generating human-readable explanations for flagged anomalies, supporting auditor review and regulatory documentation.

EVALUATION

Experimental Setup

~3M

Anonymized Records

Evaluated across payroll, tax, procurement, and transactional datasets

4

Jurisdictions

Reflecting real-world multi-jurisdictional enterprise conditions

5

Years of Data

Longitudinal dataset enabling temporal pattern analysis

All records were anonymized prior to evaluation. The dataset was designed to reflect realistic enterprise conditions, including cross-border transactions, varying reporting standards, and mixed structured and unstructured document types.

Key Results

High Precision & Recall

The framework achieved strong detection performance, correctly identifying fraud patterns across document boundaries with low miss rates.

Reduced False Positives

Significant reduction in false positive rate compared to rule-based baselines, directly lowering unnecessary manual review volume.

Lower Audit Workload

Measurable decrease in manual audit effort, enabling compliance teams to focus on high-confidence risk cases.

From Reactive to Predictive Compliance

Reactive Validation

Continuous Monitoring

Predictive Intelligence

By enabling continuous cross-document analysis and learning from audit feedback, the framework shifts compliance from a periodic, reactive process to a persistent, intelligence-driven capability. Each audit cycle improves model calibration for future detection.

Architectural Considerations for Enterprise Deployment

Scalability

Graph and LLM components must be designed for horizontal scaling to handle enterprise data volumes without degradation in latency or accuracy.

Data Privacy & Governance

LLM inference on financial records requires strict data residency, access control, and audit trail requirements — particularly under GDPR, CCPA, and sector-specific mandates.

Model Explainability

Regulators require interpretable outputs. LLM-generated rationales for flagged entities support examiner review and documentation.

Feedback Integration

The system must support structured feedback from auditors to continuously improve risk model calibration without retraining from scratch.

Practical Limitations & Open Challenges

LLM Hallucination Risk

In high-stakes compliance contexts, LLM outputs must be constrained, validated, and audited — not consumed as authoritative without verification.

Schema Heterogeneity

Real enterprise environments include legacy systems, inconsistent schemas, and undocumented data formats that challenge normalization pipelines.

Regulatory Lag

Compliance frameworks evolve. The system's normalization and rule layers must accommodate frequent regulatory updates without full redeployment.

Key Takeaways

1

Cross-Document Reasoning is the Gap

Fraud and compliance risk live at the intersection of datasets. Single-document systems cannot close this gap.

2

LLMs Unlock Unstructured Data

Semantic understanding and contextual linking bring unstructured financial documents into the correlation pipeline for the first time.

3

Continuous Learning Compounds Value

Audit feedback loops mean the system improves with every cycle, shifting compliance from static rules to adaptive intelligence.

Thank You

Varsha Shah — Enterprise Technical Architect

linkedin.com/in/varsha-shah-7b5111247

varsha.shah.tech@gamil.com

Conf42 Large Language Models (LLMs) 2026

"By enabling continuous cross-document analysis and learning from audit feedback, this approach shifts compliance from a reactive validation process to a predictive and intelligence-driven model."