Intelligent Regulatory Insight System for Indian Public Companies: An AI Driven Financial Forensics and Compliance Platform

Swapna Lokande
Department of Computer Science and Business Systems
St. Vincent Pallotti College of Engineering & Technology
Nagpur, India
Email: Swap9890

Rishikesh Koli
Department of Computer Science and Business Systems
St. Vincent Pallotti College of Engineering & Technology
Nagpur, India
Email: 1983rishikesh@gmail.com

Tejaswini Patil
Department of Computer Science and Business Systems
St. Vincent Pallotti College of Engineering & Technology
Nagpur, India
Email: tejaswinipatil875@gmail.com

Isha Sonkusare
Department of Computer Science and Business Systems
St. Vincent Pallotti College of Engineering & Technology
Nagpur, India
Email: ishasonkusare.006@gmail.com

Yashraj Kulkrani
Department of Computer Science and Business Systems
St. Vincent Pallotti College of Engineering & Technology
Nagpur, India
Email: yashrajkulkarni0405@gmail.com

***Abstract*—The Intelligent Regulatory Insight System for Indian Public Companies is a comprehensive multi agent financial forensics platform designed to support fraud detection, risk assessment, and regulatory compliance for publicly listed firms operating in India. The platform brings together structured financial statements, real time market data, news articles, social media sentiment, and regulatory enforcement information inside a unified architecture that can operate with minimal human intervention and near real time responsiveness. The underlying design philosophy centers on a modular ten agent system where each agent specializes in distinct tasks such as multi source data ingestion, forensic metric computation, composite risk scoring, compliance checking, market sentiment analysis, peer comparison, and continuous regulatory monitoring. Modern transformer based language models power an interactive question answering layer that enables analysts and regulators to query the system using natural language and receive contextual answers accompanied by supporting evidence and clear explanations. Initial experiments conducted on a representative sample of large capitalization Indian companies listed on major stock exchanges suggest that the Intelligent Regulatory Insight System for Indian Public Companies can maintain consistent and reliable metric calculations across heterogeneous data sources, successfully surface unusual financial patterns that may indicate manipulation or distress, and generate human readable risk narratives that can effectively support investigators, forensic auditors, and institutional investors in their decision making processes. Beyond demonstrating strong technical performance, the system represents a meaningful step toward practical regulatory technology that can be integrated into existing operational workflows at securities regulators, stock exchanges, and financial institutions.**

Keywords: financial forensics, Indian capital markets, SEBI, fraud detection, FinBERT, regulatory technology, multi agent systems, risk assessment

I. Introduction

Public equity markets function as critical mechanisms for capital allocation in modern economies, and their effectiveness fundamentally depends on the availability of reliable financial reporting and transparent corporate disclosure practices. When publicly traded companies engage in earnings manipulation, deliberately conceal liabilities, or selectively disclose material information to favored parties, the resulting information asymmetries create avoidable risks for ordinary investors, institutional lenders, and other market participants. Regulators are then forced to intervene after economic damage has already been inflicted on stakeholders, which undermines confidence in the fairness and integrity of the capital markets.

In the Indian context, the Securities and Exchange Board of India has recognized these challenges and has begun to systematically expand its capabilities in several related domains including digital forensics, forensic accounting, cyber security investigations, and advanced data analytics.[web:18][web:24] These efforts have included partnerships with specialized academic and research institutions that focus on forensic science methodologies and computational analysis techniques. Such

collaborative initiatives signal a clear strategic shift toward technology driven supervision and evidence based enforcement, moving beyond traditional manual audit and inspection processes that are inherently limited in their ability to cover thousands of listed entities and millions of transactions in real time.

At the same time, the sheer volume, velocity, and variety of data related to listed companies have grown exponentially. Structured data now includes not only quarterly and annual financial statements but also real time market prices, trading volumes, and granular transactional records. Unstructured data encompasses news articles, social media discussions, analyst research reports, management commentary transcripts, and enforcement orders issued by regulators. Manual or spreadsheet based forensic analysis methods cannot scale to handle this data and often miss cross links between accounting signals, market behaviour, and narrative sentiment.

This convergence of regulatory need and data abundance creates a strong case for automated systems that can systematically read, clean, normalize, and synthesise financial statements, market data, textual information, and regulatory documents in a consistent way. Such systems should augment rather than replace human expertise. They must provide traceable data lineage, interpretable metrics, and clear explanations that can withstand scrutiny in legal proceedings, board reviews, or academic work. They also need to be deployable on realistic infrastructure.

The Intelligent Regulatory Insight System for Indian Public Companies aims to address this challenge by offering an end to end platform architecture tailored to Indian equity markets, regulatory frameworks, and disclosure practices. While the system focuses on Indian public companies, it leverages both global and domestic data sources to achieve broad coverage. Data sources include standardised financial statements, market prices and trading volumes, news content, and official regulatory documents from the securities regulator and related authorities.

The rest of the paper explains the design and evaluation of the system. Section II reviews related work in financial forensics, financial natural language processing, and regulatory technology. Section III describes the system methodology including architecture, data, metrics, and risk scoring. Section IV presents results from experiments on Indian companies. Section V concludes and outlines directions for future work.

II. Literature Review

A. Financial Forensics and Red Flag Approaches

Traditional approaches to financial forensics rely on ratio analysis, common size statements, and heuristic red flag indicators. Ratios such as current ratio, quick ratio, debt to equity, and interest coverage are compared with thresholds or sector norms to spot potential liquidity or solvency issues. Common size statements express each item as a percentage of revenue or total assets, which helps reveal structural changes over time. Trend analysis examines how revenue, margins, and cash flows evolve across periods.

These methods are still widely used because they are transparent and easy to communicate. However, they are usually implemented by hand or with small spreadsheets, which restricts the number of companies and years that can be analysed. They also tend to treat each metric independently and rarely incorporate external signals such as trading behaviour or news sentiment.

Structured models like manipulation scores and distress models combine several ratios into a single index. While they offer more formal structure, most of them were developed for specific markets and time periods and may not transfer cleanly to an Indian context. The Intelligent Regulatory Insight System for Indian Public Companies therefore keeps the idea of computing a wide set of ratios and vertical and horizontal metrics but arranges them into broader risk dimensions that can be tuned over time.

B. Financial Language Models

Advances in language modelling have made it possible to process financial text with more nuance than before. FinBERT and similar models fine tune a transformer architecture on financial corpora and achieve better performance on sentiment and classification tasks than general purpose models.[web:17][web:37] Such models can identify whether a sentence carries positive, negative, or neutral information for a firm and can be adapted for tasks like topic classification or extraction of specific details.

For the system described here, financial language models are primarily used to derive sentiment scores from news and media coverage and to support the question answering component. Embeddings from domain tuned models are stored in a vector database, and retrieval is used to select context passages for question answering. This helps tie textual evidence directly to metrics and risk scores.

C. Regulatory Technology and Supervisory Technology

Regulatory technology covers tools and systems that help regulated firms comply with rules and help regulators supervise markets. Supervisory technology usually refers to systems used inside regulatory bodies and exchanges. Examples include transaction monitoring, surveillance rules on order books, and tools that check whether mandatory disclosures have been filed.

In India, regulatory bodies have started to invest in digital forensics, data analytics, and formal collaborations with forensic science institutions.[web:21][web:24] These initiatives acknowledge that supervision cannot rely only on manual reviews. However, many existing systems focus on trading patterns or basic compliance checks. Fewer tools combine detailed financial statement analysis, text understanding, and peer benchmarking into a single environment. The Intelligent Regulatory Insight System for Indian Public Companies is intended to move toward such an integrated view.

III. Methodology

A. System Architecture and Agents

The Intelligent Regulatory Insight System for Indian Public Companies uses a microservice architecture built around ten agents. FastAPI serves as the external interface for all requests. Celery and Redis handle asynchronous tasks, and PostgreSQL stores structured numeric data and metadata. A separate vector store using ChromaDB holds document embeddings for retrieval.

The ten agents and their roles are:

- Agent 1: Data Ingestion
- Agent 2: Forensic Analysis
- Agent 3: Risk Scoring
- Agent 4: Compliance Validation
- Agent 5: Report Generation
- Agent 6: Orchestration
- Agent 7: Question and Answer Interface
- Agent 8: Market Intelligence
- Agent 9: Peer Benchmarking
- Agent 10: Regulatory Monitoring

Agent 6 coordinates the others. When a user triggers an analysis, the orchestrator schedules an ingestion job for Agent 1, waits for normalized financial data to be available in the database, and then calls Agents 2 and 3 to compute metrics and risk scores. Agent 4 runs compliance checks. Agent 5 prepares human readable outputs. Agents 8, 9, and 10 maintain background data for sentiment, peer benchmarks, and regulatory events. Agent 7 answers questions using all available information.

Table I summarises the agents, inputs, and outputs.

Figure 1 shows the high level processing flow from external data sources through agents to end users. The figure is based on the diagram you supplied and has been integrated into the LaTeX file.

TABLE I
Agents in the Intelligent Regulatory Insight System for Indian Public Companies

Agent	Role	Main Inputs and Outputs
1	Data Ingestion	Inputs: Yahoo Finance, NSE, BSE, FMP, filings, news. Output: normalized tables in PostgreSQL and text in ChromaDB.
2	Forensic Analysis	Input: normalized statements. Output: twenty nine forensic metrics with quality flags.
3	Risk Scoring	Input: metrics. Output: six dimension risk scores and contributing factors.
4	Compliance Validation	Inputs: metrics, regulatory text. Output: potential non compliance indicators.
5	Report Generation	Inputs: metrics, scores, flags. Output: PDF reports, dashboards, CSV exports.
6	Orchestration	Inputs: analysis requests. Output: job schedules, statuses, logs.
7	Q&A Interface	Inputs: user queries, embeddings. Output: answers with citations to documents.
8	Market Intelligence	Inputs: news, Google Trends, prices. Output: sentiment and attention time series.
9	Peer Benchmarking	Inputs: metrics, sector data. Output: peer distributions, z scores, and outlier lists.
10	Regulatory Monitoring	Inputs: enforcement orders, circulars. Output: company to event links and timelines.

B. Data Sources and Normalisation

The system uses a mixture of structured and unstructured data. Structured data covers quarterly and annual income statements, balance sheets, cash flow statements, and daily market data. Yahoo Finance serves as the primary source for these fields because it provides a consistent interface for both Indian and global companies. Financial Modeling Prep offers extended history and additional breakdowns and is used as a backup when needed. The National Stock Exchange and Bombay Stock Exchange portals provide official reference data for tickers, listing status, and corporate actions.

Unstructured data includes company filings in portable document format or XBRL, enforcement orders and circulars from the securities regulator, and news articles from financial media outlets. These documents are fetched using scripts and converted to clean text. Each document is tagged with metadata such as company identifier, date, document type, and source.

All structured data is mapped into an internal schema with standard field names. For each company and quarter the system stores a consolidated record linking all statement lines and market data. Each field records its source and whether it was cross checked. Where sources disagree beyond a tolerance, both values are kept along with a flag indicating a conflict.

Table II summarises the main data sources.

Embeddings for text documents are computed using finance tuned models and stored in ChromaDB. These embeddings support semantic search for question answering and regulatory checks.

C. Forensic Metrics

Agent 2 computes twenty nine forensic metrics divided into three categories: vertical analysis, horizontal analysis, and classic ratios. The goal is to give a broad but interpretable picture of how a company is reporting its performance and how that reporting evolves.

Vertical analysis converts financial statements into common size form. For the income statement, each line is expressed as a share of revenue. For the balance sheet, each asset item is expressed as a share of total assets and each liability or equity item as a share of total liabilities and equity. This makes it easy to see whether, for example, the share of receivables

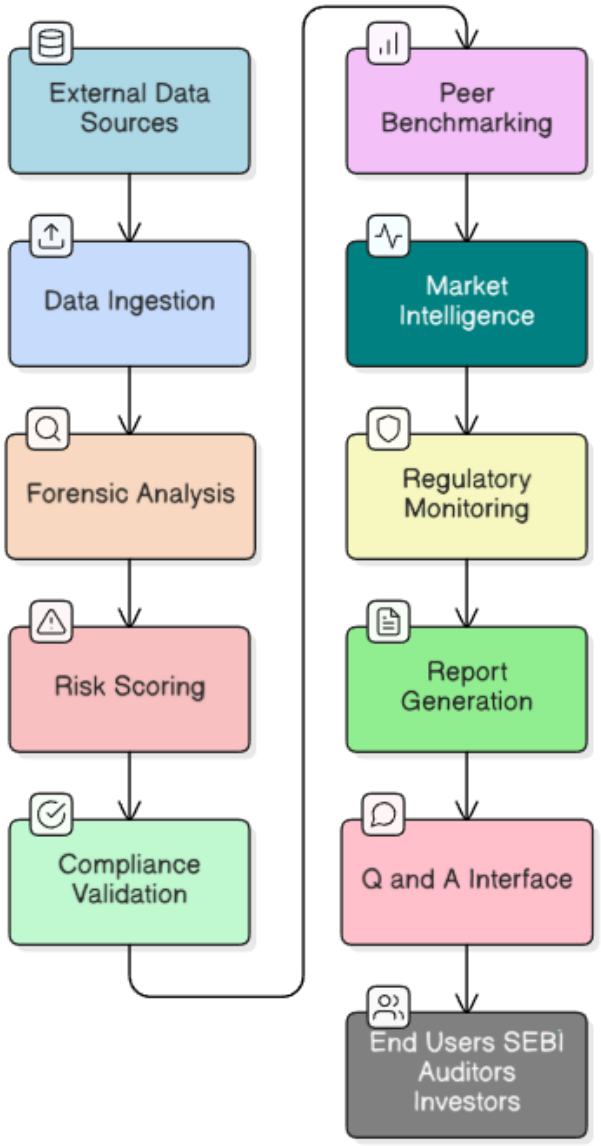

Fig. 1. Processing pipeline of the Intelligent Regulatory Insight System for Indian Public Companies, from external data sources through analysis agents to end users such as SEBI analysts, auditors, and investors.

is creeping up over time or whether the firm is increasingly financed by long term debt.

Horizontal analysis focuses on movement from one period to the next. For a given item X_t, the growth rate is

$$g_t = \frac{X_t - X_{t-1}}{X_{t-1}} \times 100\%$$

assuming the previous value is non zero. Growth rates are computed for revenue, operating income, net income, key asset and liability items, and operating cash flow. The system also calculates simple rolling summaries to capture broader trends.

Ratio analysis links the three main statements. Liquidity ratios capture the ability to meet short term obligations.

TABLE II
DATA SOURCES AND CHARACTERISTICS

Source	Coverage	Frequency	Format
Yahoo Finance	Indian and global equities	Intraday and daily	JSON and CSV feeds
NSE Portal	Indian listed firms	Daily	HTML and PDF pages
BSE Portal	Indian listed firms	Daily	HTML and PDF pages
FMP API	Global coverage	Daily	JSON responses
Company Filings	Indian listed firms	Quarterly and annual	PDF and XBRL
News Feeds	Financial media	Continuous	HTML and RSS feeds
SEBI Website	Enforcement actions and circulars	Event driven	PDF and HTML
Google Trends	Search interest for tickers and terms	Weekly	CSV time series

Profitability ratios measure how efficiently assets and equity are converted into earnings. Leverage ratios track debt levels and the capacity to service interest. Each ratio is computed using the normalized fields and is associated with a quality flag indicating whether all inputs came from primary data or whether any imputation was used.

Table III lists the metrics grouped by category.

D. Risk Dimensions and Compliance Checks

Agent 3 organises the metrics into six risk dimensions: earnings quality, leverage and solvency, liquidity pressure, revenue stability, governance and disclosure, and market behaviour anomalies. For each dimension it selects a subset of metrics and indicators and combines them into a score.

Continuous metrics are standardised within peer groups by computing z scores. Binary indicators capture events such as restatements, auditor changes, and mentions in enforcement documents. A general scoring form is

$$R = \sum_{i=1}^{n} w_i z_i + \sum_{j=1}^{m} v_j I_j$$

where w_i and v_j are weights. Weights are set to reflect domain views on importance and can be adjusted as more data becomes available. Scores are binned into qualitative levels such as low, moderate, and high risk, and the system keeps track of which metrics contribute most.

Agent 4 runs compliance checks. Some checks are arithmetic (for example, reconciliation of cash balances across statements). Others rely on retrieving relevant parts of regulations and comparing them to disclosure text. For instance, if rules require disclosure of certain related party transactions, the system checks filing text for that language and flags missing or vague wording.

TABLE III
FORENSIC METRICS IN THE INTELLIGENT REGULATORY INSIGHT SYSTEM FOR INDIAN PUBLIC COMPANIES

Category	Metrics
Vertical analysis (11)	Revenue share, cost of goods sold share, operating expense share, operating income share, net income share, current assets share, fixed assets share, intangible assets share, current liabilities share, long term debt share, equity share.
Horizontal analysis (10)	Quarter on quarter revenue growth, operating income growth, net income growth, total assets growth, current assets growth, total liabilities growth, operating cash flow growth, receivables growth, inventory growth, cash and equivalents growth.
Financial ratios (8)	Current ratio, quick ratio, debt to equity, interest coverage, return on assets, return on equity, gross margin, operating margin.

IV. RESULTS AND DISCUSSION

A. Data Quality and Coverage

The system was evaluated on a group of large Indian companies from several sectors over multiple years. For the majority of company quarters the ingestion layer was able to obtain all fields required to compute the full set of metrics. In many observations the same fields were available from more than one source. In those cases, values usually agreed within a narrow band, giving confidence that the normalisation was correct.

Where discrepancies did occur, they were usually linked to timing, minor restatements, or small definitional differences. The ingestion agent marks these discrepancies and preserves both values. Analysts can then inspect them if a particular case becomes important for investigation.

B. Patterns, Peer Context, and Narrative Support

The metrics and risk scores highlighted patterns that align with intuitive expectations. Companies that later reported stress or underwent restructuring often showed early signals in the liquidity and leverage dimensions, such as falling current ratios, higher debt to equity, and reduced interest coverage. Firms with very strong headline revenue growth sometimes exhibited fast growth in receivables and weaker cash.

V. CONCLUSION

REFERENCES

[1] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre trained Language Models," arXiv:1908.10063, 2019.[web:17]

[2] A. H. Huang et al., "FinBERT: A Large Language Model for Extracting Information from Financial Text," 2023.[web:37]

[3] Securities and Exchange Board of India, "Memorandum of Understanding between SEBI and National Forensic Sciences University," 2025.[web:24][web:42]

[4] Securities and Exchange Board of India, "Empanelment of Agencies for Providing Digital Forensics Services," 2025.[web:21]