YELLOWSENSE TECHNOLOGIES

Brief description of our idea:

Fraud in the securities market is increasingly sophisticated — ranging from fake trading apps and Ponzi schemes to pump-and-dump groups on WhatsApp/Telegram and deepfake corporate announcements. These activities erode retail investor trust and undermine market integrity. Detecting such fraud requires analyzing data across **multiple parties** — banks (suspicious fund transfers), brokers (unusual trades), social media monitors (fake tips/videos), and exchanges (abnormal price movements). However, no single entity is willing or able to share raw client data openly due to privacy, compliance, and competition concerns.

Our solution introduces **Confidential Clean Rooms (CCR)** for fraud detection. A CCR is a secure computing environment based on **Confidential Computing and Trusted Execution Environments (TEEs)** where encrypted data from multiple sources can be analyzed collaboratively without ever exposing the raw inputs. For example, a bank can upload encrypted fund transfer records, a broker can contribute encrypted trading data, and a monitoring tool can provide flagged social media posts. Within the CCR, AI/ML models and rule-based systems run on the combined data to detect patterns such as:

- The same phone number/email used across fraudulent bank transfers and suspicious trades.
- Correlation between spikes in stock activity & coordinated WhatsApp group tips.
- Detection of manipulated videos or fake IPO allotment messages.

The CCR outputs only **alerts and risk scores**, never raw data. Regulators like SEBI or stock exchanges receive a **dashboard view** showing red flags (e.g., “Client ID masked123 correlated with pump-and-dump group activity”) without access to sensitive client-level data.

This approach achieves three goals simultaneously:

1. **Investor Protection** – Early detection of fraud saves investors from losses.
2. **Market Development** – Restores confidence by ensuring safer participation.
3. **Supervision** – Provides SEBI with real-time insights at scale, without privacy compromises.

Our hackathon prototype demonstrates this flow with encrypted broker, bank, and social datasets, analyzed in a CCR to produce actionable fraud alerts. In production, this will scale on TEEs (Intel SGX, AMD SEV, AWS Nitro), enabling India to leapfrog global standards in market integrity.

Technology stack we intend to use:

Our solution combines **Confidential Computing infrastructure** with **AI-driven fraud detection models** to deliver a secure and scalable system for SEBI's fraud prevention mandate.

1. **Confidential Computing & Trusted Execution Environments (TEE):**
    - Core deployment will run on hardware-backed enclaves (Intel SGX, AMD SEV-SNP, AWS Nitro Enclaves, or GCP Confidential VMs).
    - This ensures all computations on sensitive data (bank transfers, broker trades, social media signals) happen in encrypted memory, inaccessible even to cloud providers or system admins.
    - Remote attestation mechanisms will prove to regulators and participants that the computation environment is tamper-proof.
2. **Encryption & Data Privacy:**
    - End-to-end encryption of inputs using AES-256/Fernet before upload.
    - Differential privacy and tokenization to mask PII in outputs.
    - Audit logs and cryptographic proofs to ensure compliance.
3. **Fraud Analytics Layer (AI/ML + Rule Engines):**
    - NLP pipelines (Transformers/BERT) to analyze WhatsApp/Telegram/social posts for fraud signals (e.g., "guaranteed returns," fake IPO promises).
    - Time-series anomaly detection (Prophet, PyTorch, Scikit-learn) to flag sudden trading or transfer spikes.
    - Correlation algorithms across encrypted datasets to detect coordinated fraud (e.g., same phone/email across bank + broker).
    - Deepfake detection (CV models) for video/audio fraud.
4. **Data Ingestion & Orchestration:**
    - Brokers, banks, and monitoring tools provide encrypted CSV/JSON via secure APIs.
    - Apache Kafka / Pulsar (future) for real-time ingestion.
    - Airflow for pipeline orchestration.
5. **User Interface (for SEBI & Exchanges):**

- Streamlit/Dash-based dashboards for prototype.
- Production: React.js + FastAPI backend for scalable alert dashboards.
- Role-based access control for regulators vs brokers.

This stack ensures **robust cybersecurity**, **real-world feasibility**, and **scalability** to millions of trades, messages, and transactions daily, while strictly preserving privacy.

Prototype title / name:

"NIRIKSHAK – a secure AI-powered system for real-time fraud detection and compliance monitoring in financial markets." "Nirikshak" (निरीक्षक) already means inspector / overseer in Hindi-Sanskrit. This is our prototype solution that uses **Confidential Clean Rooms (CCR)** to detect and prevent securities market fraud while ensuring absolute data privacy. It enables multiple stakeholders — banks, brokers, exchanges, and monitoring tools — to contribute encrypted data into a **Trusted Execution Environment (TEE)**, where AI/ML models analyze patterns of fraud such as pump-and-dump schemes, fake IPO allotments, or deepfake corporate announcements. The CCR produces **fraud alerts and risk scores** without exposing raw client-level data, aligning with SEBI's mandate of **investor protection, market integrity, and secure supervision**.

Brief description of the prototype:

Nirikshak is a secure, AI-powered prototype designed to detect and prevent frauds in securities markets by leveraging advanced Confidential Computing, Natural Language Processing (NLP), and Machine Learning (ML) models. The system focuses on protecting investors from deceptive practices such as fake investment apps, fraudulent advisors, pump-and-dump stock tips, deepfake announcements, and misleading corporate disclosures.

The prototype continuously scans multiple digital channels—websites, mobile apps, social media platforms (Twitter, WhatsApp, Telegram), and public disclosures—using AI models trained on linguistic, behavioral, and network patterns. It identifies anomalies such as unusual stock tips, sudden price movements linked to online chatter, impersonation attempts, or announcements inconsistent with historical company data.

To ensure trust and compliance, *Nirikshak* is built on **Confidential Computing infrastructure** (Trusted Execution Environments), enabling sensitive investor and broker data to be analyzed securely without exposure to external parties. This makes the solution privacy-preserving while still enabling regulators, exchanges, and intermediaries to gain actionable intelligence.

The output is delivered through an interactive **dashboard for regulators**, which provides:

- Real-time fraud alerts with confidence scores.
- Linkages between suspicious content and market activity.
- Advisor/broker verification against SEBI's regulatory database.
- AI-generated credibility scores for corporate announcements.

In its future versions, *Nirikshak* can be extended to provide a **mobile-facing investor app**, enabling retail investors to verify the legitimacy of advisors, stock tips, or investment apps instantly.

By combining **cutting-edge AI/ML, secure TEEs, and regulatory integration**, *Nirikshak* aligns directly with SEBI's mandate of investor protection and market integrity, offering a scalable and deployable tool to safeguard India's growing retail investor base.

Key features, intended users and functionality:

Nirikshak is a next-generation fraud detection prototype that safeguards India's securities markets by combining **AI/ML, NLP, and Confidential Computing (TEE)**. It has three core dimensions: fraud detection, secure data processing, and regulatory intelligence delivery.

Key Features:

1. **Fraud Content Detection** – Uses NLP and ML to scan websites, social media (WhatsApp, Telegram, Twitter), and investment apps for suspicious content such as fake advisors, misleading stock tips, pump-and-dump schemes, deepfakes, and false corporate announcements.
2. **Entity Verification** – Cross-verifies investment advisors, brokers, or apps against SEBI's registered intermediary's database to flag impersonators.

3. **Market–Social Signal Linkage** – Connects anomalies in online chatter with unusual trading/price activity to detect manipulation attempts in near real time.
4. **Credibility Scoring** – Assigns AI-driven credibility scores to corporate announcements by comparing with historical disclosures, counter-party statements, and market fundamentals.
5. **Confidential Processing** – Uses Trusted Execution Environments (TEEs) to ensure all sensitive data (investor, broker, regulator) is analyzed securely without exposure, meeting stringent privacy and cybersecurity needs.
6. **Regulatory Dashboard** – Provides actionable insights through a unified dashboard with alerts, heatmaps, and investigation trails.

Intended Users:

- **SEBI & Regulators**: Monitor fraudulent activities, take enforcement actions.
- **Stock Exchanges**: Track and verify corporate announcements, unusual price movements.
- **Brokers & Intermediaries**: Prevent impersonation of their brand and protect clients.
- **Retail Investors (future app extension)**: Verify legitimacy of advisors, stock tips, or apps instantly before making decisions.

Functionality:

- Ingests data streams (social media, public disclosures, exchange filings).
- Runs fraud detection pipelines inside TEEs to ensure privacy.
- Flags anomalies with confidence scores and links to market activity.
- Displays alerts and verification results in an interactive dashboard.

Together, these features empower regulators and investors to stay ahead of fraudsters while maintaining trust and integrity in India's capital markets.

Key innovation / differentiator:

Nirikshak goes beyond existing fraud detection tools by introducing a **privacy-preserving, AI-powered surveillance system** purpose-built for India's securities markets. Its differentiators lie in both **technological depth** and **regulatory alignment**.

1. **Confidential Computing (TEE Integration):** Unlike traditional fraud analytics systems that centralize sensitive investor and broker data, *Nirikshak* runs fraud detection pipelines within **Trusted Execution Environments (TEEs)**. This ensures that raw data from regulators, exchanges, and brokers remains protected even during processing. No other solution in India's financial ecosystem offers this level of data security and trust.
2. **End-to-End Market Signal Fusion:** While most tools only track either market activity or social media, *Nirikshak* integrates **online chatter (tips, deepfakes, fake websites)** with **exchange data (trades, filings, price moves)**. This fusion enables early detection of fraud patterns such as pump-and-dump or misinformation-led manipulation.
3. **Regulatory-Grade Identity Verification:** By directly cross-referencing advisors, brokers, and platforms against **SEBI's registered intermediaries database**, *Nirikshak* instantly filters impersonators and fake apps—something general fraud detectors do not address.
4. **AI-Powered Credibility Scoring of Corporate Announcements:** Instead of just flagging anomalies, *Nirikshak* generates **contextual credibility scores** by comparing disclosures with historical filings, peer company actions, and financial performance. This helps regulators prioritize investigations.
5. **Scalable, Low-Cost Deployment:** Built with modular AI microservices and TEEs, *Nirikshak* can be scaled across exchanges, brokers, and eventually to **retail investor-facing apps**, enabling both top-down (regulator-driven) and bottom-up (investor-driven) protection.
6. **Alignment with SEBI's Mandate:** While global fraud detection systems exist, *Nirikshak* is **custom-tailored for Indian markets**, vernacular-ready, and aligned with SEBI's core pillars—**investor protection, market integrity, and supervision efficiency**.

By combining **advanced security (TEE), AI-driven fraud detection, and regulatory alignment**, *Nirikshak* becomes not just another monitoring tool, but a **LeapTech innovation** capable of revolutionizing fraud prevention in capital markets.

What makes your solution unique or impactful?

Nirikshak is unique because it brings **confidential computing, AI, and regulatory intelligence** together in a way no existing fraud detection system has achieved.

1. **Data Privacy by Design:** Instead of centralizing raw investor or broker data, *Nirikshak* processes information within **Trusted Execution Environments (TEEs)**. This ensures that even during fraud analysis, sensitive financial data is shielded from misuse—a critical differentiator in building regulator and market trust.
2. **Holistic Fraud Monitoring:** Traditional fraud detection focuses only on **trade data** or **social media chatter**. *Nirikshak* fuses signals across **corporate filings, stock exchange announcements, unusual market movements, and online misinformation campaigns (deepfakes, pump-and-dump tips, fake IPO apps)**. This **multi-layered detection** enables earlier and more accurate fraud discovery.
3. **Contextual Credibility Scoring:** Beyond flagging suspicious activity, *Nirikshak* provides **explainable AI scores** on the credibility of announcements or advisories, guiding regulators and investors on which alerts demand immediate attention. This minimizes noise and improves enforcement efficiency.
4. **Vernacular-Ready & Investor-Friendly:** The prototype is designed with **multilingual interfaces**, enabling investors across India (including Tier 2/3 cities) to verify advisor identities, announcements, and suspicious offers in their local language.
5. **Scalable Market Impact:** Built as modular AI microservices, *Nirikshak* can serve multiple stakeholders—**regulators, brokers, and retail investors**—with the same core technology. Over time, this can dramatically reduce fraudulent activities, restore **investor confidence**, and strengthen the **credibility of Indian capital markets globally**.

By blending **cutting-edge confidential computing with localized AI-driven fraud detection**, *Nirikshak* is not just a compliance tool—it's a **market safeguard** designed to protect millions of Indian investors and uphold SEBI's mission of fair, transparent markets.

Demo Link / GitHub / Video:

https://github.com/ARox2005/Yellowsense_fraud_detection_Demo_Video

---

Potential impact & scalability:

Nirikshak can significantly strengthen **market integrity and investor trust** by addressing frauds at scale. Today, millions of retail investors face risks from fake IPO apps, deepfakes, pump-and-dump schemes, and fraudulent advisors. By embedding **confidential computing with AI/ML fraud monitoring**, Nirikshak creates a **privacy-preserving surveillance layer** that regulators, brokers, and investors can rely on. The impact is twofold: investors gain confidence to participate more actively, and regulators get a **scalable compliance shield**.

The system is **cloud-ready, modular, and API-driven**, enabling seamless integration with stock exchanges, broker back-offices, and investor apps without overhauling existing infrastructure. Built on a **distributed microservices model**, it can scale from **thousands to millions of data points per second**, ensuring real-time detection even during peak market activity. With **multilingual interfaces**, the solution can penetrate Tier-2/3 markets, democratizing safe investing. Long-term, Nirikshak can expand globally as a **benchmark model for AI-powered, confidential market surveillance**.

How will this prototype address the respective problem?

Nirikshak directly addresses SEBI's challenge of **fraud prevention** by providing a unified, AI-powered system that can detect, verify, and flag fraudulent market activities across multiple channels. It ingests data from **social media, stock exchange announcements, corporate filings, broker records, and advisor credentials**, then applies **NLP, credibility scoring, and anomaly detection** to identify potential frauds such as pump-and-dump tips, fake IPO apps, misleading advisories, or deepfake content.

Unlike existing point solutions, *Nirikshak* is built on **confidential computing**, ensuring raw investor or broker data remains secure during analysis—solving both **fraud detection and data privacy** concerns in one system. It provides **real-time alerts** to regulators and brokers, along with **multilingual investor tools** for verifying announcements and advisors. By creating a **trustworthy, scalable, and privacy-preserving fraud detection layer**, the prototype not only prevents investor losses but also strengthens **confidence in Indian capital markets**.

Regards

YellowSense Team:

prakhargoyal@iitb.ac.in

animesh@ai.yellowsense.in

Members:

1. Animesh Sharma: +91- 9835-086-449
2. Prakhar Goyal: +91-9869-397-868
3. Shashank Naik: +91 – 9008-033-789