VIHAAN
LET 8 BEGIN

QUANTIFI.AI

AGENTIC AI FOR REAL-TIME ADAPTIVE FRAUD DETECTION

A Modular, Explainable , Multi Agent Framework

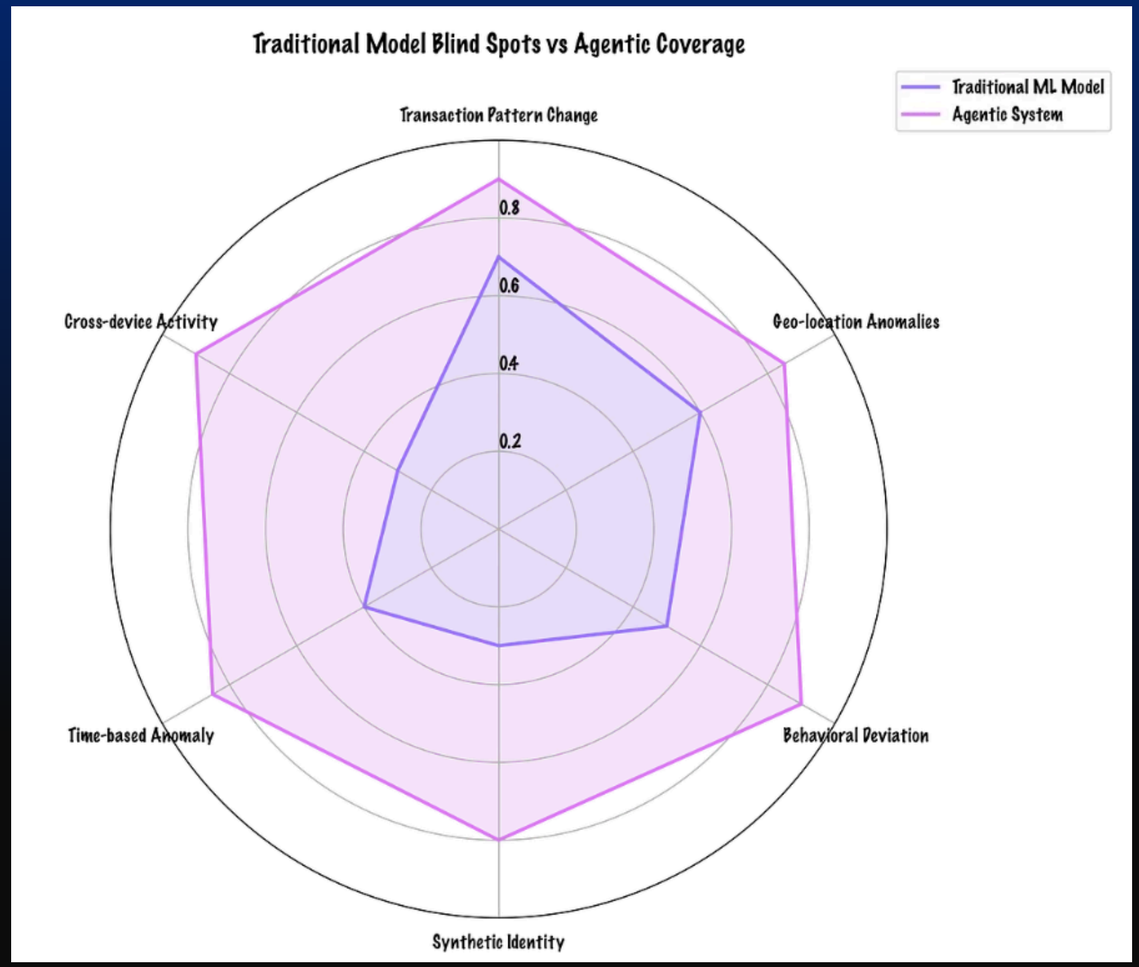
Fraud is not just a technical detection problem – it's a fast-changing, adversarial, high-regulation battlefield.
Traditional models hit a wall. Agentic AI brings the next-gen agentic toolkit to break through it

By Kushal Khemka, Mehul

PROBLEM STATEMENT

""Modern fraud is non-stationary and evolving in real time, driven by bots, social engineering, synthetic identities, and adversarial behaviors. This is where agentic AI shines.""

- Fraud evolves faster than model retraining cycles.
- High False positives (= legitimate transaction that is incorrectly flagged as fraudulent) hurt the customer experience and drain time.
- Interpretability is critical to comply with audits and regulations.
- Systems must react in near real-time.
- Non-compliance with **GDPR**, PSD2, CCPA.



- **Heuristics**: e.g. manually defined rules like "flag any transaction over $10,000 made abroad at midnight." — fast, but rigid and easy to game.
- **Signature detection:** e.g. known patterns of fraud=previously seen attacks: fail to detect *new* fraud patterns (**zero-day frauds**).
- **Ensemble ML**: combines multiple models (e.g., decision trees, logistic regression) to improve accuracy: but works with large labeled datasets and hard to interpret

THE SOLUTION

Proposed Solution – Multi Agents & Graph Neural Networks for Fraud Detection System

Contextual Feature Extractor

This agent uses prompt engineering and vector search on prior labeled transactions to extract semantically similar transaction clusters. It enriches transaction metadata with contextual signals like merchant behavior, device fingerprint anomalies, and cross-session irregularities.

Pattern Divergence Analyst

This agent compares the transaction to a dynamic behavioral profile of the user, which is built through prior embeddings and time-series forecasting using Multi-Agent Divergence Policy Optimization (MADPO). It evaluates:

- Deviations in transaction size/time/geo
- New devices or merchant IDs
- Sudden frequency bursts

Each deviation is scored.

Risk Synthesizer Agent

This agent fuses the pattern scores and flags from Agent 2 with industry-accepted risk signals (e.g., MCC code risk scores, BIN lookup history, geolocation risk tiers). It applies an LLM-driven reasoning template to synthesize the signals into human-readable rationales.

Explanation Generation Agent

To meet audit requirements, this agent generates a plain-language justification for the risk classification. It cites the rationale from Agent 3, the transaction history, and known fraud trends. These justifications are cached and indexed for compliance audits.

Decision Recommender Agent

This agent performs weighted decisioning based on:

- Risk score
- Confidence thresholds
- Customer tier (e.g., high-value clients may bypass soft declines)
- Historical false positive rate for similar profiles

It chooses from options: approve, soft decline (require OTP), hard block, or route to manual review.

Finally, this last 6th core agent learns from feedback loops:

- **Analyst overrides**
- **Post-event fraud labeling**
- **Customer dispute resolutions**

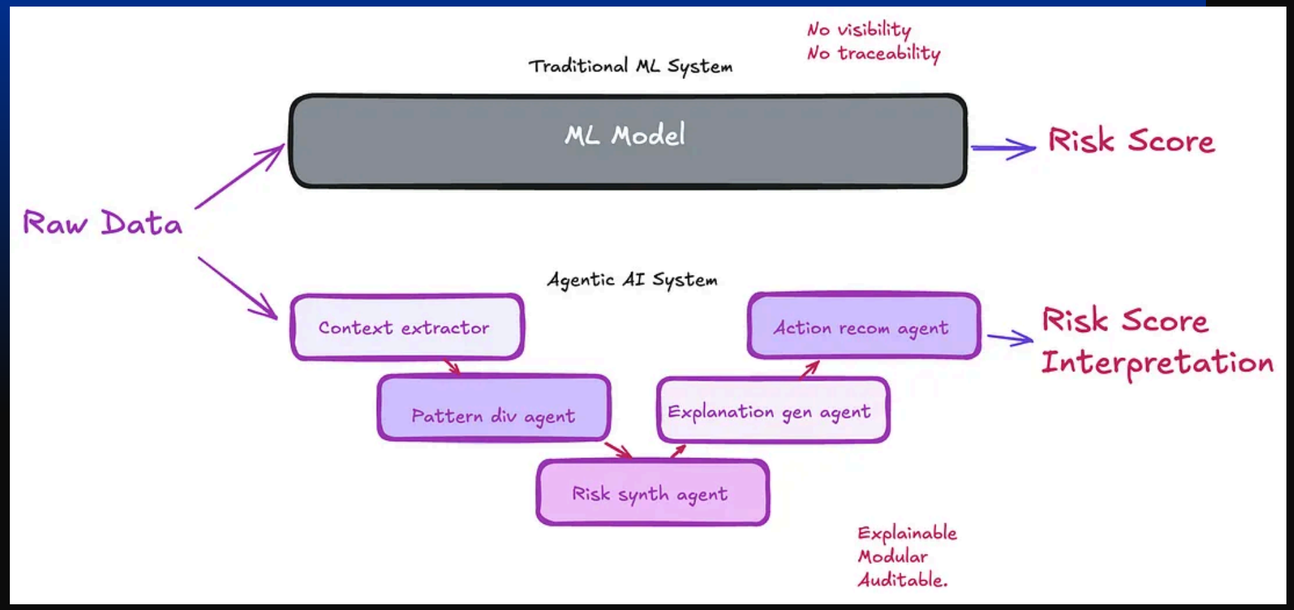
It fine-tunes the agent-specific prompting and weights based on this feedback, ensuring long-term improvement without full model retraining.

Why Multi Agent instead of Traditional ML ?

Unlike monolithic fraud models, this architecture has key benefits:

- Explainability: Every decision is narratively justified and traceable.
- Scalability: Each agent can scale independently.
- Domain Adaptability: You can swap or fine-tune agents per region or risk category.
- Resilience: If one agent fails, others can still carry the signal forward.
- Human-in-the-loop Ready: Designed for seamless analyst intervention.

Core Workflow



MULTI AGENTS

AUXILIARY AGENTS

Compliance Office Agent

Monitors the entire fraud detection pipeline to ensure strict adherence to internal policies and external regulations such as AML, KYC, and CFT. It audits agent activities, enforces alert thresholds, detects compliance breaches, and maintains traceable logs for internal reviews or external audits.

Multi-Channel / Social-Engineering Detection Agent

Analyzes customer communication channels—chat, email, voice, and SMS—for phishing, vishing, and social engineering attempts. It uses NLP and speech models to detect coercion, impersonation, scam indicators, and correlates risky messages with downstream transaction patterns.

Regulatory Filing Agent

Automatically generates and submits jurisdiction-specific reports like SARs, CTRs, and CIP logs by extracting flagged activities and narrative context from other agents. It ensures timely, accurate submissions to regulators via secure APIs and maintains a filing history with audit trails.

Synthetic-Identity & Identity-Graph Agent

Detects fabricated identities and account networks by building and analyzing user-device-address relationship graphs using Graph Neural Networks (GNNs). It flags anomalous clusters, overlapping identity markers, and co-located behavioral patterns that indicate synthetic or fake profiles.

USE CASES

Banking & Financial Services

Detect:

- Synthetic identity fraud through identity graph analysis.
- [IMP] Fraud Detection and Suspicious Transactions , suspicious merchants , Mule Accounts (Future Plan)
- Account takeovers and unusual access using device fingerprinting and behavioral biometrics.
- Money laundering patterns via anomaly detection on transaction flows and velocity.
- Regulatory alerts through automated filing agents (SAR, CTR) and KYC due diligence bots.

E-Commerce & Digital Marketplaces

Protect online transactions and customer accounts from fraud using:

- Multi-channel fraud analysis (voice/chat/email scams).
- Bot behavior and fake account detection using device and behavior biometrics.
- Real-time scoring of purchases and logins via pattern divergence and contextual profiles.

Healthcare & Insurance (NOT FOCUSED NOW)

Secure medical claims and insurance workflows by:

- Detecting duplicate claims and upcoding with historical pattern analysis.
- Flagging anomalous billing behavior by correlating across provider networks.
- Identifying fabricated patients or claims via synthetic identity graphing.

Fraud Investigation & Compliance

Streamline fraud operations with agents dedicated to:

- Audit-ready reporting and automated regulatory filings (e.g., SAR, CTR, CIP).
- Regulation-specific checks (GDPR, KYC/AML) using dedicated compliance agents.
- Privacy-preserving collaborative analysis across consortiums using federated learning.

TECHSTACK

crewai

LangGraph

Streamlit

OpenAI

PyTorch geometric

FUTURE OF FRAUD DETECTION AI AGENTS

The future of AI-driven fraud detection looks promising, with continuous advancements expected in the coming years. As AI technologies improve, our fraud detection agent will evolve to meet the changing landscape of fraud tactics. Some of the key advancements we anticipate include:

- **Improved Explainability**: As AI models become more transparent, our agent will offer even clearer explanations for its decisions. This will help build trust between fraud analysts and regulatory bodies.
- **Enhanced Behavioral Analytics:** Future versions of the agent will leverage even more granular behavioral data, such as emotional analytics and psychological profiling, to better predict fraudulent intent and behaviors.
- **Integration with Blockchain:** Blockchain technology offers exciting possibilities for fraud prevention. In the future, our AI agent could integrate with blockchain networks to provide real-time, immutable transaction verification, further enhancing fraud detection capabilities.
- **Collaboration Across Organizations:** As AI models become more connected and share insights in real time, our fraud detection agent could be part of a larger ecosystem, enabling information sharing between institutions and improving the detection of cross-platform fraud.

FUTURE PLANS

Federated Learning / Privacy-Preserving Agent

Enables secure, collaborative model training across institutions by using federated learning, **differential privacy, and encrypted model updates**. It preserves data locality, ensures regulatory compliance, and helps build robust, generalized fraud detection models without sharing raw data.

With open-source LLMs agents , data will too remain protected

Device & Behavioral-Biometrics Agent

Continuously profiles users based on behavioral patterns like typing rhythm, mouse dynamics, swipe behavior, and device posture. **It also scores device integrity (e.g., rooted, emulated, jailbroken devices), helping to detect account takeovers, bots, and anomalous user sessions.**

THANKS FOR YOUR ATTENTION!