Team DeepPeek

Hackathon Presentation : AI Driven Entity Intelligence & Risk analysis

Samarth Mishra,
Akshat Tripathi
Ayush Bhatt,
Akash Kumar,
Raju Nelluri

Architecture

- Architecture of end to end landscape

Implementation

- Implementation of respective modules

Tech Stack, Data Sources & Models we tried

- Implementation of respective modules

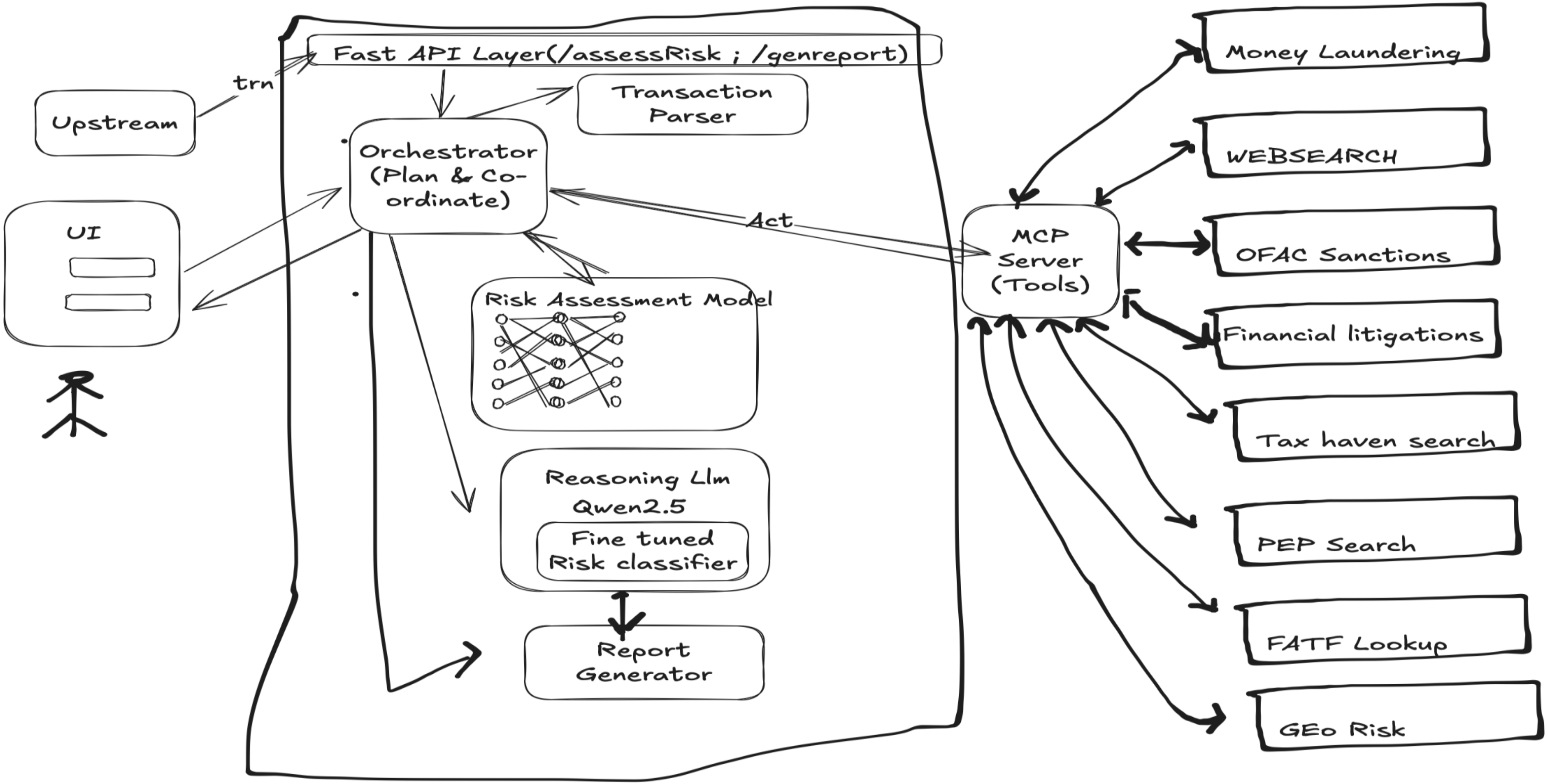
Results

- Output artefacts and details

Summary & future forward

- Lessons learnt
- Future developments

DeepPeek - AI Driven Entity Intelligence & Risk Analysis



Implementation Details

DeepPeek - FastAPI Layer

The FastAPI layer serves as the main entry point for the system, handling API requests from users or integrated applications..

DeepPeek - Risk Assessment Model

The core intelligence engine evaluates transaction risk using machine learning and rule-based techniques.

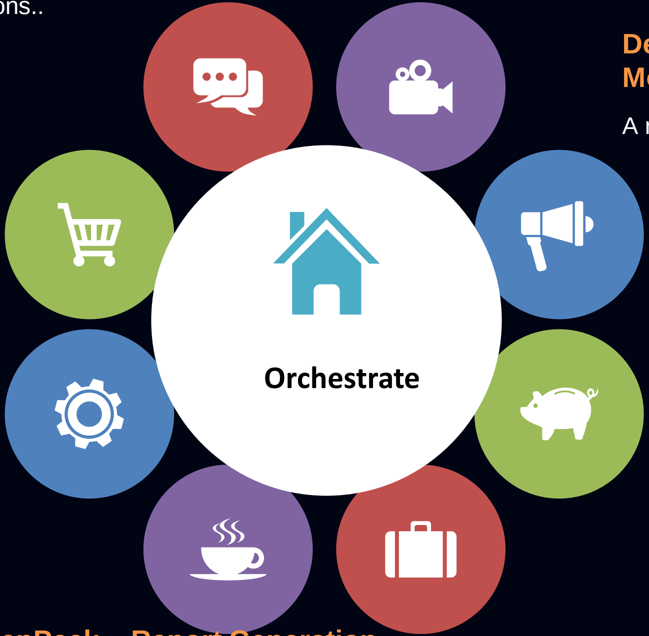
DeepPeek - Transaction Parser

This module extracts key transaction details to aid risk analysis. It:

- Parses structured and unstructured transaction data (e.g., bank transactions, wire transfer logs etc).

DeepPeek - Web Search Module

A real-time adverse media screening



DeepPeek - Politically Exposed Persons (PEP) Search

This module verifies if an entity is classified as a PEP, which includes government officials and their associates

DeepPeek - Money Laundering Detection Module

This module identifies potential money laundering activities using advanced heuristics and ML models.

DeepPeek - OFAC Sanctions Lookup

This module checks individuals, organizations, and financial institutions against the Office of Foreign Assets Control (OFAC) sanctions lists

DeepPeek - Tax Haven Search Module

Identifies financial transactions and entities linked to tax havens.

DeepPeek - Report Generation with Qwen2.5 LLM

Comprehensive risk intelligence report using Qwen2.5 LLM

DeepPeek - Financial Action Task Force (FATF) Lookup

This module ensures compliance with FATF high-risk jurisdictions and recommendations.

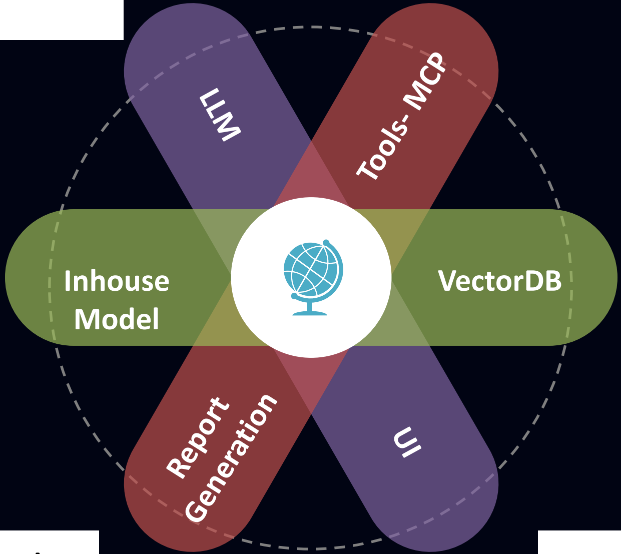
Tech Stack, Data Sources & Models we tried..

Qwen2.5

ANTHROP\C

Model Context Protocol

OFAC List	PEP List	FATF List
News & Senti	Tax Haven	AML



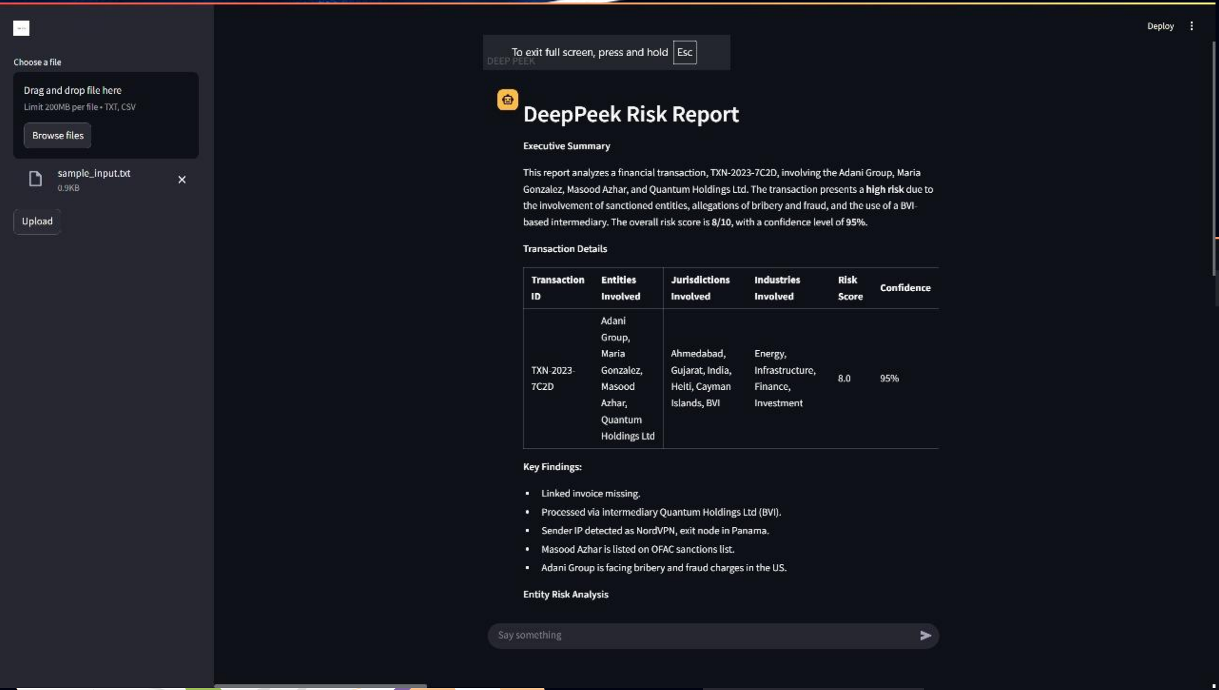
qdrant

Models tried: XGBoost, TreeSearch, RandomForest, Support Vector Machine,

Streamlit

Agents / LLMS tried: smolagents, Owl agents, Gemini, Qwen

Results



DeepPeek Risk Report

Executive Summary

This report analyzes a financial transaction, TXN-2023-7C2D, involving the Adani Group, Maria Gonzalez, Masood Azhar, and Quantum Holdings Ltd. The transaction presents a **high risk** due to the involvement of sanctioned entities, allegations of bribery and fraud, and the use of a BVI-based intermediary. The overall risk score is **8/10**, with a confidence level of **95%**.

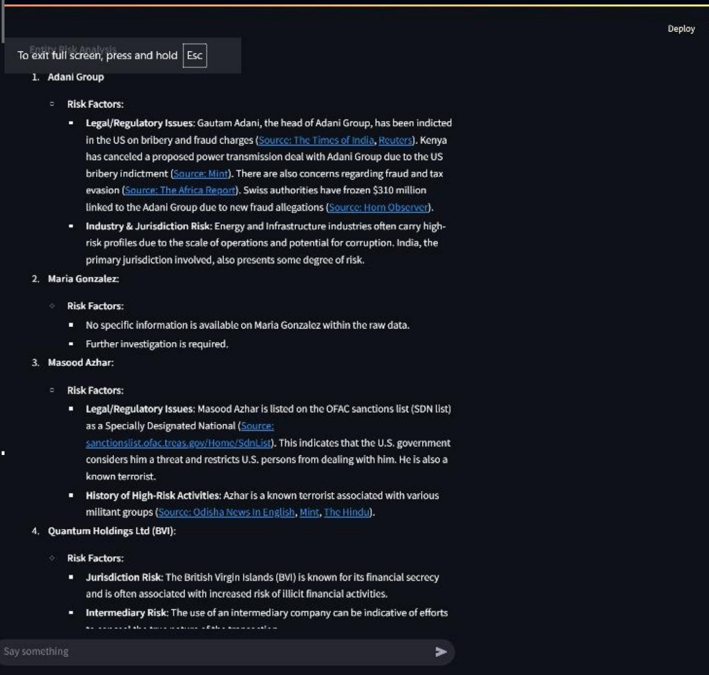
Transaction Details

Transaction ID	Entities Involved	Jurisdictions Involved	Industries Involved	Risk Score	Confidence
TXN-2023-7C2D	Adani Group, Maria Gonzalez, Masood Azhar, Quantum Holdings Ltd	Ahmedabad, Gujarat, India, Heiti, Cayman Islands, BVI	Energy, Infrastructure, Finance, Investment	8.0	95%

Key Findings:

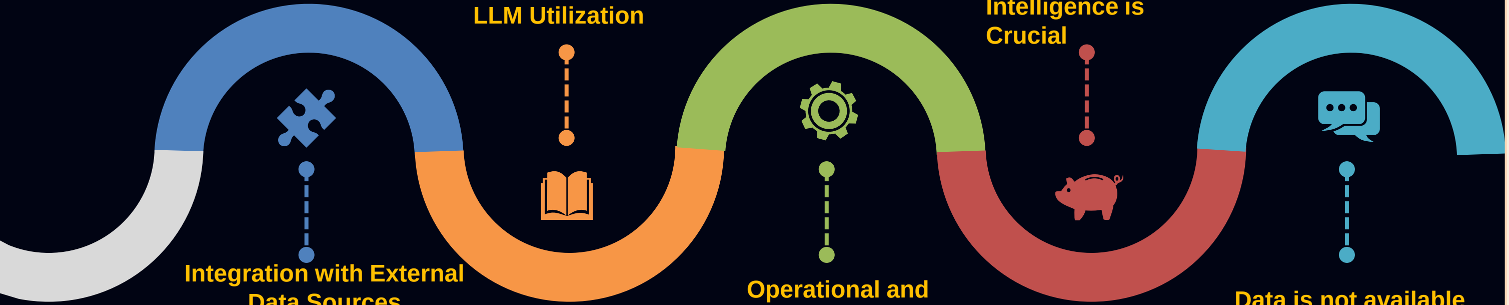
- Linked invoice missing.
- Processed via intermediary Quantum Holdings Ltd (BVI).
- Sender IP detected as NordVPN, exit node in Panama.
- Masood Azhar is listed on OFAC sanctions list.
- Adani Group is facing bribery and fraud charges in the US.

Entity Risk Analysis



1. **Adani Group**
   - **Risk Factors:**
     - **Legal/Regulatory Issues**: Gautam Adani, the head of Adani Group, has been indicted in the US on bribery and fraud charges (Source: The Times of India, Reuters). Kenya has canceled a proposed power transmission deal with Adani Group due to the US bribery indictment (Source: Mint). There are also concerns regarding fraud and tax evasion (Source: The Africa Report). Swiss authorities have frozen $310 million linked to the Adani Group due to new fraud allegations (Source: Horn Observer).
     - **Industry & Jurisdiction Risk**: Energy and Infrastructure industries often carry high-risk profiles due to the scale of operations and potential for corruption. India, the primary jurisdiction involved, also presents some degree of risk.
2. **Maria Gonzalez:**
   - **Risk Factors:**
     - No specific information is available on Maria Gonzalez within the raw data.
     - Further investigation is required.
3. **Masood Azhar:**
   - **Risk Factors:**
     - **Legal/Regulatory Issues**: Masood Azhar is listed on the OFAC sanctions list (SDN list) as a Specially Designated National (Source: sanctionslist.ofac.treas.gov/Home/SdnList). This indicates that the U.S. government considers him a threat and restricts U.S. persons from dealing with him. He is also a known terrorist.
     - **History of High-Risk Activities**: Azhar is a known terrorist associated with various militant groups (Source: Odisha News In English, Mint, The Hindu).
4. **Quantum Holdings Ltd (BVI):**
   - **Risk Factors:**
     - **Jurisdiction Risk**: The British Virgin Islands (BVI) is known for its financial secrecy and is often associated with increased risk of illicit financial activities.
     - **Intermediary Risk**: The use of an intermediary company can be indicative of efforts to conceal the true nature of the transaction.

Summary & lessons learnt..



Integration with External Data Sources

OFAC, FATF, and PEP databases are frequently updated, requiring a real-time data pipeline.

- Cached versions of sanction lists led to false negatives, making real-time updates essential.

Lessons on AI and LLM Utilization

Qwen2.5 LLM Performed Well but Needed Guardrails
The LLM struggled with financial domain-specific jargon, requiring custom prompt engineering and fine-tuning.

Operational and Compliance Lessons

Explainability is Key for Financial Compliance
The Qwen2.5 LLM-generated reports helped, but human review was still required for high-risk cases.

Fine-Tuning NLP Models for Risk Intelligence is Crucial

Pre-training on banking transaction data and compliance reports improved accuracy.

Data is not available on Internet

Majority of the Data is not available on internet so collecting the data and normalizing it has been difficult to train our models.

Final Summary

Balance automation with human oversight: While AI improved efficiency, human intervention was still required for high-risk cases.

- Iterative improvements are key: The system required continuous retraining and fine-tuning to remain effective

- Compliance is non-negotiable: Adhering to global regulatory frameworks from day one prevented legal roadblocks.

- Explainability builds trust: Providing transparent AI-driven decisions helped gain regulator and stakeholder confidence.

Appendix

Implementation – in Full Details

DeepPeek - FastAPI Layer

The FastAPI layer serves as the main entry point for the system, handling API requests from users or integrated applications. It provides:

- High-performance request handling using Python’s FastAPI framework, ensuring low latency.
- Authentication and authorization to restrict access based on user roles.
- Request validation and preprocessing, ensuring input data is sanitized before processing.
- Response formatting to standardize output for downstream consumption.

The FastAPI layer calls the orchestrator module to execute the risk assessment workflow.

DeepPeek - Orchestrator

The orchestrator acts as the central coordination engine, managing the flow of data and decision-making across the different risk assessment components. It:

- Receives input data from the FastAPI layer and routes it to the Transaction Parser.
- Calls the Risk Assessment Model to compute a preliminary risk score.
- Triggers various risk intelligence tools such as money laundering detection, OFAC sanctions lookup, and geo-risk analysis.
- Aggregates outputs from all modules and forwards them to Qwen2.5 LLM for report generation.
- Handles error management and retries to ensure smooth operation

DeepPeek - Transaction Parser

This module extracts key transaction details to aid risk analysis. It:

- Parses structured and unstructured transaction data (e.g., bank transactions, wire transfer logs etc).
- Identifies senders, receivers, intermediary banks, and payment channels.
- Detects anomalous patterns (e.g., countries, unusual counterparties).
- Converts extracted data into a normalized format for the Risk Assessment Model.

DeepPeek - Risk Assessment Model

The core intelligence engine evaluates transaction risk using machine learning and rule-based techniques. It:

- Computes an initial risk score based on transaction attributes.
- Leverages historical fraud patterns and anomaly detection algorithms.
- Cross-checks transactions with customer risk profiles and behavior trends.
- Integrates with external data sources (e.g., sanctions lists, adverse media) to refine risk scores.

DeepPeek - Money Laundering Detection Module

This module identifies potential money laundering activities using advanced heuristics and ML models. It:

- Detects layering and structuring techniques (e.g., multiple small transactions below reporting thresholds).
- Identifies unusual transaction chains involving offshore accounts.
- Matches transactions with known money laundering typologies.
- Assigns a laundering probability score to each transaction.

DeepPeek - Web Search Module

A real-time adverse media screening tool that:

- Searches public sources for negative news, legal cases, or fraud reports.
- Uses web scraping and NLP models to extract sentiment and relevance.
- Flags entities linked to criminal activities, bankruptcies, or regulatory actions.

Implementation – in Full Details contd..

DeepPeek - OFAC Sanctions Lookup

This module checks individuals, organizations, and financial institutions against the Office of Foreign Assets Control (OFAC) sanctions lists. It:

- Uses fuzzy matching to detect name variations and aliases.
- Flags transactions involving sanctioned entities or embargoed countries.
- Integrates with global sanctions databases for broader coverage.

DeepPeek - Tax Haven Search Module

Identifies financial transactions and entities linked to tax havens. It:

- Cross-checks counterparties against known tax haven jurisdictions.
- Flags shell companies and anonymous financial vehicles.
- Detects unusual transaction flows into low-tax or no-tax regions

DeepPeek - Politically Exposed Persons (PEP) Search

This module verifies if an entity is classified as a PEP, which includes government officials and their associates. It:

- Matches individuals and entities against global PEP databases.
- Assigns a PEP risk level based on their role and influence.
- Flags potential conflicts of interest or corruption risks.

DeepPeek - Financial Action Task Force (FATF) Lookup

This module ensures compliance with FATF high-risk jurisdictions and recommendations. It:

- Flags transactions involving countries with weak AML regulations.
- Checks compliance with FATF’s blacklist and graylist.
- Identifies non-cooperative financial institutions.

DeepPeek - Geo-Risk Analysis Module

This module assesses geopolitical and location-based financial risks. It:

- Assigns risk scores based on country stability, regulatory frameworks, and crime rates.
- Detects cross-border transactions involving high-risk regions.
- Uses geospatial analytics to flag suspicious patterns.

DeepPeek - Report Generation with Qwen2.5 LLM

The final step involves generating a comprehensive risk intelligence report using Qwen2.5 LLM. It:

- Synthesizes findings from all risk modules.
- Generates a human-readable risk assessment summary.
- Provides recommendations for compliance officers and risk teams.
- Uses explainable AI techniques to justify risk scores and flag

Lessons learnt 0n full details

Architectural and System Design Lessons

Modular Design is Essential

- The modular approach allowed flexibility in integrating new risk tools without disrupting the core system.
- Independent microservices enabled parallel processing, improving performance and scalability.

Orchestration Complexity Needs to be Managed

- Managing multiple risk intelligence tools required an efficient orchestration mechanism.
- Asynchronous processing and event-driven architecture helped optimize execution but added complexity in debugging and error handling.

API Gateway and Security Considerations

- The FastAPI layer performed well, but security concerns like rate limiting, authentication, and request validation were crucial.
- Ensuring zero-trust architecture (e.g., API tokenization and role-based access) helped prevent data leaks.

.

Data Handling and Risk Model Improvements

Data Normalization is Critical

- Transaction data formats varied significantly across sources, requiring robust ETL (Extract, Transform, Load) pipelines for example the PEP and OFAC have different dimensions and need to be normalized.
- Handling missing data, duplicate records, and inconsistent formats was a major challenge.

Machine Learning Models Need More Context

- The risk assessment model worked well but sometimes lacked contextual awareness, leading to false positives.
- Incorporating historical transaction patterns and behavioral analysis improved accuracy.

Bias in Risk Models Needs Constant Monitoring

- Certain regions or customer segments were flagged at higher rates due to inherent biases in training data.
- Regular auditing and explainability techniques helped improve fairness and compliance

Lessons Learnt

Integration with External Data Sources

Sanctions and PEP List Data is Dynamic

- OFAC, FATF, and PEP databases are frequently updated, requiring a real-time data pipeline.
- Cached versions of sanction lists led to false negatives, making real-time updates essential.

Challenges in Web Search and Adverse Media Analysis

- Web scraping for negative media screening had issues with spam filtering, duplicate articles, and fake news detection.
- Sentiment analysis models struggled with sarcasm and legal jargon, requiring NLP fine-tuning.

Operational and Compliance Lessons

Regulatory Constraints Must Be Considered Early

- Certain jurisdictions have strict data residency and privacy regulations (e.g., GDPR, CCPA).
- Ensuring audit logs, traceability, and data minimization was crucial for compliance.

Explainability is Key for Financial Compliance

- Regulators and compliance officers needed clear explanations of AI-driven risk scores.
- The Qwen2.5 LLM-generated reports helped, but human review was still required for high-risk cases.

Lessons on AI and LLM Utilization

Qwen2.5 LLM Performed Well but Needed Guardrails

- The LLM provided detailed risk summaries, but at times generated overly cautious reports with unnecessary flags.
- Implementing fact-checking and confidence scoring helped reduce misleading outputs.

Fine-Tuning NLP Models for Risk Intelligence is Crucial

- The LLM struggled with financial domain-specific jargon, requiring custom prompt engineering and fine-tuning.
- Pre-training on banking transaction data and compliance reports improved accuracy.

Final Takeaways

- Balance automation with human oversight: While AI improved efficiency, human intervention was still required for high-risk cases.
- Iterative improvements are key: The system required continuous retraining and fine-tuning to remain effective.
- Compliance is non-negotiable: Adhering to global regulatory frameworks from day one prevented legal roadblocks.
- Explainability builds trust: Providing transparent AI-driven decisions helped gain regulator and stakeholder confidence.

Thanks a lot & Make us Win ☺