RISK UNLOCKED

TEAM:
ARCHIT LALL
ATHARVA A. MUGLIKAR
KAUSHAL BAID
RAJAT RAJORIA
RITABRATA DAS

INDEX

PROBLEM STATEMENT

- Currently, data analysts spend significant manual effort analyzing party names from transaction details to determine the correct entities involved.
- This process becomes even more challenging when dealing with corporations, non-profits, and potential shell companies due to naming inconsistencies, abbreviations, fraudulent entities, and lack of structured information.

SOLUTION PROPOSED

AI Chatbot:

Provides an interactive risk assessment chatbot using Ollama (Mistral-7B) and Gemini LLM, enhanced with text-to-speech (TTS) support for voice-based responses.
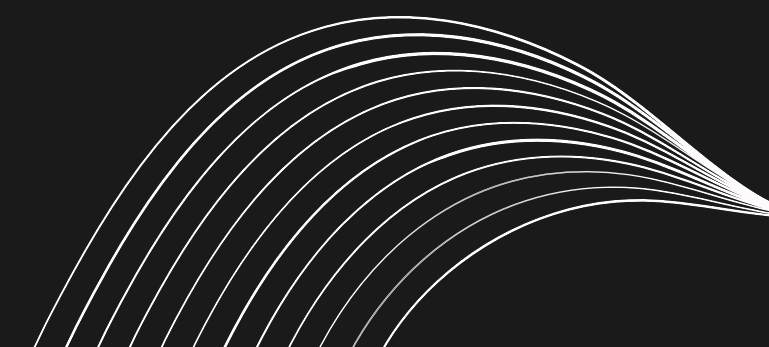
Risk Score Calculation:

Analyzes entities and transactions for sanction violations and geopolitical risk.

Sanctions & Compliance:

Cross-checks entities with OFAC, UN, FATF, EU, and OpenSanctions API.

News Sentiment Analysis:

- Uses Google News API and FinBERT to detect negative press coverage.

Entity Recognition:

- Extracts company details using NER-based classification.

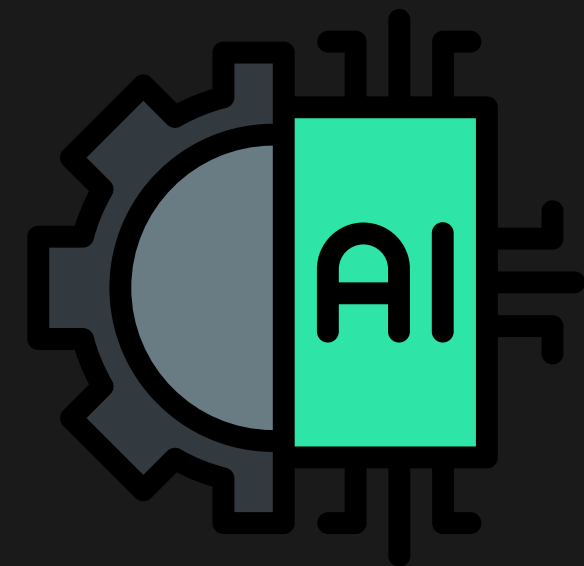# TECH STACK

Public APIS

Python

Large language Models

DATA SOURCES

Google News

Gleif

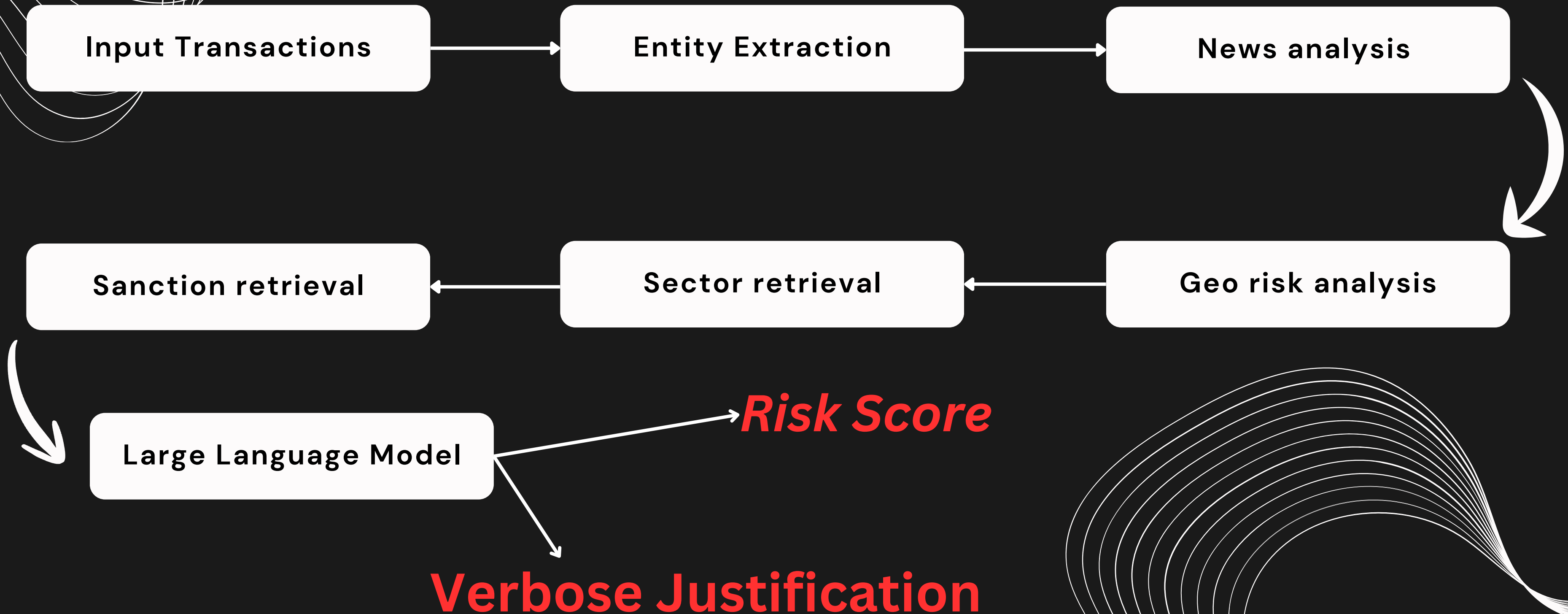Open sanction

OFAC

Sec Edgar

ARCHITECTURE

Input Transactions → Entity Extraction → News analysis

Sanction retrieval ← Sector retrieval ← Geo risk analysis

Large Language Model → *Risk Score*

Large Language Model → **Verbose Justification**

DATA PROCESSING

Life cycle of data:

1. Input data → Entity extraction :
   Name of entities involved is extracted along with their type
   and the country of jurisdiction(also checks for PEP and individual)
   Output :
   ```
   { "Transaction ID": txn_id,
     "Extracted Entity": [item["Extracted Entity"] for item in classified_entities],
     "Entity Type": [item["Entity Type"] for item in classified_entities],
     "Supporting Evidence": [item["Supporting Evidence"] for item in classified_entities],
     "Confidence Score": round(overall_confidence, 2),
     "Countries": countries
     }
   ```

DATA PROCESSING

2. News Analysis :

Checks the current news of recent 2 months to check if the extracted entities are in the news. If yes, then capture sentiment around the entities and a score is assigned. This score is then added along with the output of the previous step.

Output:

The score can range between 0 - 1, 0 being not bad at all and 1 being very bad

DATA PROCESSING

3. Geo risk analysis:
   Countries are analysed based on metrics such as CPI , AML, FATF and GTI. Then a risk score specific to this is factor alone is calculated along with a small risk analysis note. Both of this is attached to the output to the previous step.
   Output:
   The score can range between 0 - 1 and the note will contain description and reasoning for the score

DATA PROCESSING

4. Sector retrieval:
   Sectors of both entities (if the entities are not individuals) is retrieved using Sec Edgar API. This data is later used to check the relevance of the 2 sectors using LLM.
   Output:
   The sectors of each entity is attached to the output of the previous step

DATA PROCESSING

5. Sanction analysis:

The entities are verified against a sanction list using OFAC and open_sanctions APIs. Then this data is fed to an LLM to makes sense of the data. The data has notes as well as confidence score. The makes sense of the data and the output of this is again attached to the output of the previous step

Output:

Key notes found

DATA PROCESSING

6. Verdict:

The cumulative data built from the previous steps is fed to an LLM which is given a well articulated prompt, which includes a risk scoring formula as a base(weights change as seen fit). The LLM then produces a risk score taking into consideration everything and thinking beyond just straight facts. A justification is provided with evidences which lead to the score.

Output:

Risk Score and justification

Output UI
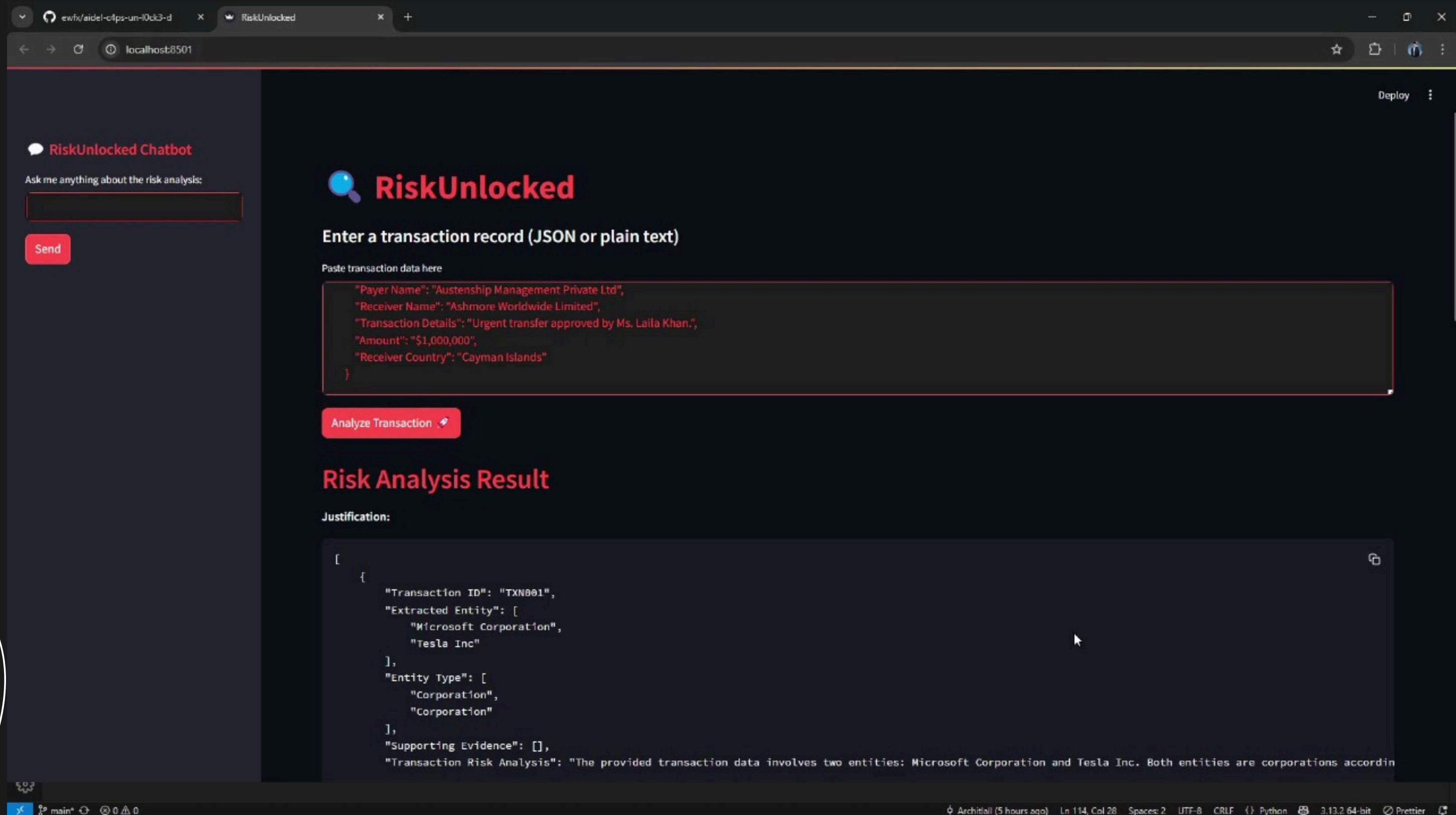
IMPACT

Reduces Manual Effort & Improves Efficiency

Enhances Fraud Detection & Compliance

Provides Explainable & Data-Driven Risk Assessment

Prevents Financial Crime & Protects Reputation