

Technology Hackathon AI-Driven Entity Intelligence Risk Analysis

1. Challenge Overview

Develop a sophisticated Generative AI/ML powered system that automates the research and evidence-gathering process for identifying, verifying, and risk-scoring entities (corporations, non-profits, shell companies, and financial intermediaries) from complex, multi-source transaction data. The solution should significantly reduce manual effort for data analysts, improve accuracy, and provide a robust risk evaluation mechanism.

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. The goal of this challenge is to build an AI-driven system that:

- Extracts entity names from unstructured and structured transaction data.
- Enriches the extracted names with publicly available data (e.g., company registries, online sources, financial news, regulatory filings, and legal databases).
- Identifies potential fraudulent or high-risk entities through anomaly detection.
- Classifies entities into categories (corporation, non-profit, shell company, government agency, etc.).
- Assigns a risk score based on entity attributes, and associated networks (business/sectors associated with the entities).
- Provides supporting evidence and confidence scores to assist analysts in decision-making.

2. Expectations from Participants

Teams are expected to:

1. **Develop an AI Model** - Implement a sophisticated Generative AI/ML-based approach to extract, research, classify, and risk-score entities.
2. **Use Open Datasets & APIs** - Integrate freely available sources such as:
 - a. OpenCorporates API
 - b. Wikipedia/Wikidata
 - c. SEC EDGAR filings

- d. Financial crime databases (e.g., OFAC sanctions list, World Bank's PEP list)
- e. News articles and public repositories

Technology Hackathon AI-Driven Entity Intelligence Risk Analysis

3. **Demonstrate Real-Time Data Handling (Optional)** - Implement real-time or near-real-time entity lookup and enrichment.
4. **Implement Anomaly Detection** - Use AI/ML techniques to identify fraudulent or shell companies.
5. **Provide an Evidence Trail** - Generate a structured output justifying entity classification, risk scoring, and supporting references.
6. **Provide a Justification** - Provide text-based reasoning for classification (refer sample output for an example).
7. **Build a Usable Interface (Optional but Recommended)** - A sophisticated dashboard for entity search, validation, and risk monitoring.

3. Technical Constraints & Guidelines

- **Use only Free Tools & APIs:** No paid services should be used.
- **Leverage NLP & LLMs:** Utilize open-source LLMs (Hugging Face models, OpenAI API free tier, GPT-J, LLaMA).
- **Data Processing Pipelines:** Implement efficient data parsing, anomaly detection, and enrichment mechanisms.
- **Risk Evaluation Mechanism:** Define clear parameters for calculating risk scores based on financial data, entity reputation, and transaction anomalies.
- **Output Format:** JSON/CSV format with extracted entities, classifications, risk scores, and supporting evidence.
- **API Endpoint:** Expose API endpoint to invoke suitable capabilities.

4. Sample Test Data Set

Participants shall consider a complex dataset of structured and unstructured transaction records. Example:

Structured data -

Transaction ID	Payer Name	Receiver Name	Transaction Details	Amount	Receiver Country
TXN001	Acme Corp	SovCo Capital Partners	Payment for services rendered	\$500,000	USA

TXN002	Global Health Foundation	Save the Children	Grant disbursement	\$2,000,000	UK
TXN003	XYZ Ltd	ABC GmbH	Purchase of office supplies	\$15,000	Germany
TXN004	Green Earth Org	CCMI	Environmental project funding	\$750,000	Cayman Islands
TXN005	Oceanic Holdings LLC	Alas Chiricanas	Offshore Investment	\$5,000,000	Panama

Participants can add additional features, if required.

Unstructured data -

Example 1:

Transaction ID: TXN-2023-5A98 Date: 2023-08-15 14:22:00

Sender: Name: "Global Horizons Consulting LLC" Account: IBAN CH56 0483 5012 3456 7800 9 (Swiss bank) Address: Rue du Marché 17, Geneva, Switzerland Notes: "Consulting fees for project Aurora"

Receiver: Name: "Bright Future Nonprofit Inc" Account: 987654321 (Cayman National Bank, KY) Address: P.O. Box 1234, George Town, Cayman Islands Tax ID: KY-45678 Amount: \$49,850.00 (USD) Currency Exchange: N/A Transaction Type: Wire Transfer Reference: "Charitable Donation - Ref #DR-2023-0815" Additional Notes: "Urgent transfer approved by Mr. Ali Al-Mansoori (Director)." "Linked invoice missing. Processed via intermediary Quantum Holdings Ltd (BVI)." Sender IP: 192.168.89.123 (VPN detected: NordVPN, exit node in Panama)

Example 2:

Transaction ID: TXN-2023-7C2D Date: 2023-08-15 14:25:00

Sender: Name: "Quantum Holdings Ltd" Account: VGB2BVZR024987654321 (British Virgin Islands) Beneficiary Owner: "Maria Gonzalez" (DOB: 1975-04-12, Spanish passport E5789012)

Receiver: Name: "Golden Sands Trading FZE" Account: AE450330000012345678901 (Dubai, UAE) Registration: UAE Free Zone License #789-F2 Amount: \$950,000.00 (USD) Currency Exchange: EUR > USD (Rate: 1.12) Transaction Type: SWIFT Reference: "Commodity Trade Settlement - Contract #DX-889" Additional Notes: "Funds routed via Deutsche Bank Frankfurt (DEUTDEFF) + Emirates NBD Dubai (EBILAEAD)." "Approver:

Mr. Viktor Petrov (Linked to OFAC SDN List entry #9876, 2022)." "Invoice attached: Oil Equipment Procurement (PDF missing metadata)."

Example 3:

Blocked Person X owns 50 percent of Entity A, and Entity A owns 50 percent of Entity B. Entity B is blocked as person X owns, indirectly, significant part of Entity B. In addition, Blocked Person X's 50 percent ownership of Entity A makes Entity A, a blocked entity. Entity A's 50 percent ownership of Entity B in turn makes Entity B too, a blocked entity.

Other unstructured data may include publicly available news sources.

Expected Output Format

JSON

```
{
  "Transaction ID": "TXN001",
  "Extracted Entity": ["Acme Corporation", "SovCo Capital Partners"],
  "Entity Type": ["Corporation", "Corporation"],
  "Risk Score": 0.65,
  "Supporting Evidence": ["OpenCorporates", "Company Website"],
  "Confidence Score": 0.95,
  "Reason": "SovCo Capital Partners is not on sanctions list but an entity of interest. It is owned by Russian businessmen and related to Soccombank PJSC, a sanctioned entity"
}
```

JSON

```
{
  "Transaction ID": "TXN005",
  "Extracted Entity": ["Oceanic Holdings LLC", "Alas Chiricanas"],
  "Entity Type": ["Shell Company", "Corporation"],
  "Risk Score": 0.9,
  "Supporting Evidence": ["Panama Papers Database", "Sanctions List"],
  "Confidence Score": 0.85,
  "Reason": "Oceanic Holdings LLC is a shell company and finds numerous mentions in Panama Papers database. It is a sanctioned entity listed on Specially Designated Nationals and Blocked Persons list"
}
```

JSON

```
{
  "Transaction ID": "TXN-2023-5A98",
  "Extracted Entity": ["Oceanic Holdings LLC", "Bright Future Nonprofit Inc", "Ali Al-Mansoori"],
  "Entity Type": ["Shell Company", "NGO", "PEP"],
  "Risk Score": 0.95,
  "Supporting Evidence": ["Panama Papers Database", "Sanctions List"],
  "Confidence Score": 0.87,
  "Reason": "Transaction involves payment from Swiss-based Global Horizons to Cayman Islands nonprofit, approved by PEP-linked Ali Al-Mansoori"
}
```

5. Submission Guidelines

Participants must submit:

- A GitHub repository with the code, including a README file with setup instructions.
- A presentation (PDF/Slides) explaining the approach, challenges, risk scoring methodology, data sources used and results.
- A demo video (optional but recommended) showcasing the solution in action.

Git template folder structure for Hackathon

- Repository information: <Team Repo>
- Code templates and folder structure

/<Team Repo>

--artifacts/

--arch/ #Architecture documents

--demo/ #presentation or Demo to support

--code

--src/ #source code

--test/ #test cases

Resources & References

- OpenCorporates: <https://api.opencorporates.com>
- Wikidata API: https://www.wikidata.org/wiki/Wikidata:Data_access
- SEC EDGAR: <https://www.sec.gov/edgar.shtml>
- OFAC Sanctions List: <https://www.treasury.gov/resource-center/sanctions>

- Hugging Face Transformers: <https://huggingface.co/docs/transformers>