**Title: Generative AI-Powered Data Profiling Solution for Regulatory Reporting**

**1. Introduction**

Regulatory reporting in the banking sector involves compiling vast amounts of data to meet compliance requirements. Ensuring data accuracy and adherence to regulations is critical to avoiding penalties and improving reporting efficiency.

This project focuses on developing a Generative AI-powered data profiling solution to automate the extraction, validation, and risk assessment of regulatory data. The solution leverages OpenAI models and machine learning techniques to enhance accuracy, efficiency, and compliance.

**2. Objective**

* Extract, interpret, and refine regulatory reporting instructions to identify key data validation requirements.
* Automatically generate profiling rules based on regulatory guidelines.
* Validate transactions against profiling rules and detect anomalies.
* Suggest remediation actions and generate explanations for flagged transactions.

**3. Solution Architecture**

**3.1 System Components**

* **Data Ingestion Module**: Reads regulatory instructions and transactional data (CSV format).
* **AI-Powered Rule Extraction**: Uses OpenAI’s GPT-4 to generate profiling rules from regulatory guidelines.
* **Data Validation Engine**: Applies rules to transactions to identify discrepancies and anomalies.
* **Risk Scoring Mechanism**: Assigns risk scores based on transaction history and compliance violations.
* **Remediation Module**: Suggests corrective actions for flagged transactions.

**3.2 Workflow**

1. Load regulatory instructions (PDF/TXT format) and financial transactions (CSV format).
2. Use OpenAI’s GPT to generate data profiling rules.
3. Apply validation rules to transactional data.
4. Identify high-risk transactions and assign risk scores.
5. Generate a report with flagged transactions and remediation suggestions.

**4. Implementation Details**

**4.1 Technology Stack**

* Programming Language: Python
* AI Models: OpenAI GPT-40
* Machine Learning: Scikit-learn
* Data Processing: Pandas, NumPy
* Deployment: Streamlit (for interactive UI), Flask API

**4.2 Folder Structure**

├── project\_root/

│ ├── data/

│ │ ├── transactions.csv

│ │ ├── regulatory\_instructions.pdf

│ ├── src/

│ │ ├── data\_loader.py

│ │ ├── profiler.py

│ │ ├── validator.py

│ │ ├── risk\_scoring.py

│ ├── reports/

│ ├── results/

│ │ ├── transactions\_with\_risk.csv

│ ├── app.py

│ ├── README.md

**4.3 Sample Output Format (transactions\_with\_risk.csv)**

Transaction ID, Account Balance, Transaction Amount, Reported Amount, Currency, Country, Transaction Date, Risk Score, Flagged Reason, Suggested Remediation

1001, 15000, 500, 500, USD, US, 2025-02-25, Low, - , -

1002, 32000, 1200, 1200, LUA, UK, 2025-02-20, Medium, Unrecognized Currency, Verify currency code

1003, -5000, 300, 300, USD, UK, 2025-02-18, High, Negative Balance, Verify OD account status

1004, 70000, 2000, 1800, USD, US, 2025-02-28, Medium, Amount Mismatch, Verify cross-currency conversion

**5. Test Cases & Validation**

| **Test Case** | **Input** | **Expected Output** |
| --- | --- | --- |
| Valid Transaction | Transaction amount = Reported amount | No flag |
| Cross-currency deviation | Deviation within 1% | No flag |
| Cross-currency deviation > 1% | Deviation exceeds 1% | Flagged transaction |
| Negative balance | Account balance < 0 (Non-OD) | Flagged transaction |
| High-risk country | Transaction amount > 50,000 in high-risk country | Flagged transaction |

**6. Future Enhancements**

* Incorporate real-time transaction monitoring.
* Enhance AI model with fine-tuned regulatory compliance datasets.
* Integrate with banking systems for real-time alerts.

**7. Conclusion**

This AI-powered data profiling solution automates compliance checks, reduces manual efforts, and enhances accuracy in regulatory reporting. It provides real-time insights into high-risk transactions, helping auditors and compliance teams take proactive actions.