Automated Regulatory Data Profiling System

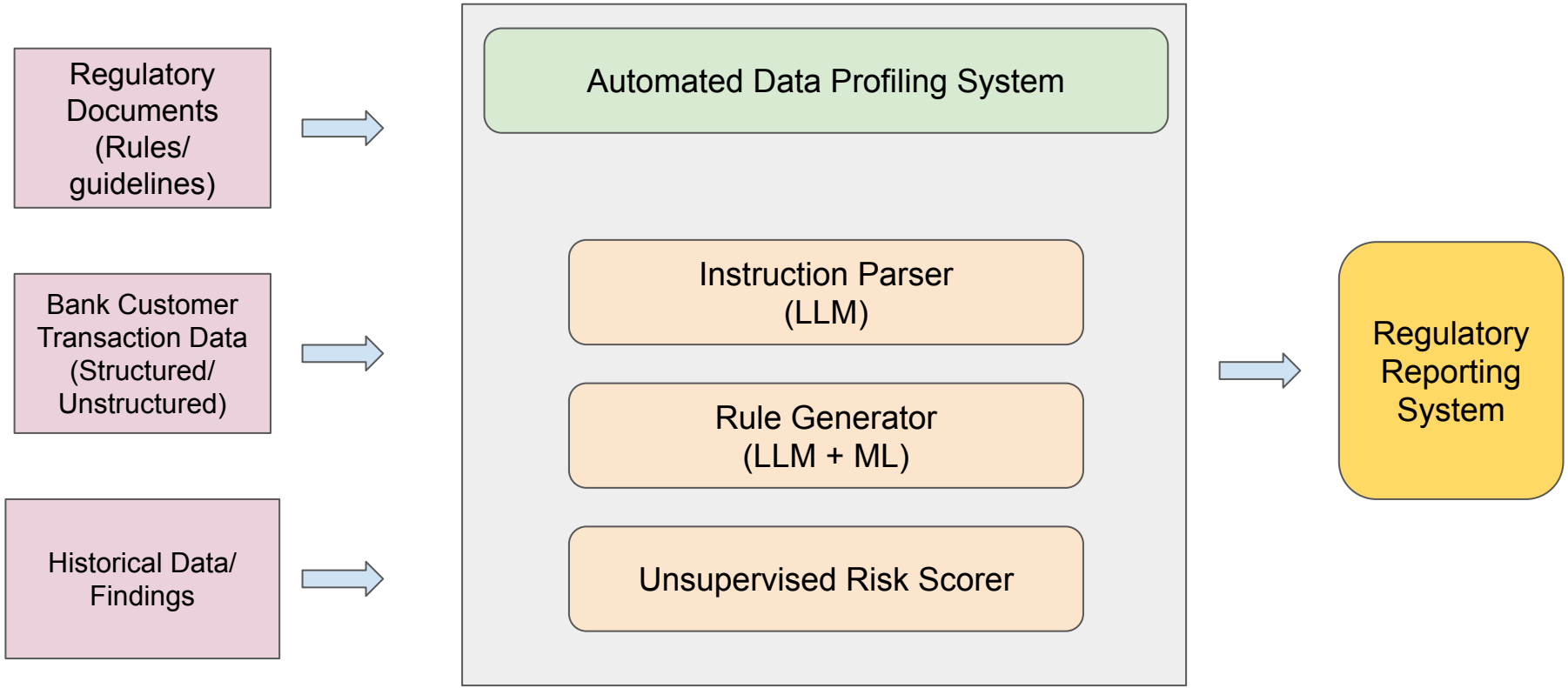
Team: AI_DEEP_LOOP

- Venubabu
- Murali
- Salar Syed
- Balaram
- Sashank

Problem statement

Regulatory reporting in the banking sector involves compiling vast amounts of data to meet compliance requirements. A critical aspect of this process is data profiling, which ensures that the reported data aligns with regulatory reporting instructions. Traditionally, this involves manually defining profiling rules based on the underlying data and regulatory requirements. This challenge aims to automate data profiling using Generative AI (LLMs) and unsupervised machine learning techniques. Participants will develop a solution that can generate data profiling rules, perform adaptive risk scoring, and suggest remediation actions based on regulatory reporting instructions.

Solution Approach



Components

- Regulation Parser
- Automated Rule generator
- Adaptive risk scoring engine
- Remediation Action Advisor
- Continuous Learner module

Regulation Parser

- Use a sequence-to-sequence language model to extract reporting requirements
- The reporting requirements can be PDFs/ spreadsheets or any text data with rules or guidelines defined.
- Convert the data to text.
- Construct prompt asking model for requirements
- Process text through model (LLM)

Automated Rule Generator

- Consumes LLM model
- Gets rule suggestions from LLM based on requirements
- Discovers Additional rules through data pattern analysis
- Constructs a detailed prompt
- Results in comprehensive data profiling rules that address both documented requirements and inherent data characteristics

Adaptive Risk Scoring Engine

- Feature Creation: Converts raw risk scoring data to numerical features
- Feature scaling: Normalize features for better model performances
- Anomaly detection: Creates feature vector for each violation as
    - Severity flag/score
    - Reason for anomaly
    - Regulatory recommendation
    - Frequency

Remediation Actions Advisor

Constructs detailed prompt consisting of:

- Current rule violation
- Historical violations
- Instructions to generate remediation actions

- Uses LLM to generate suggestions
- Option to add new remediations to knowledge base

Continuous Learner Module

- Implements continuous learning and implements improvements based on feedback
- Considers following components as input:
    - Rule generator
    - Risk Scoring engine
    - Remediation Action advisor
- Prepares training data from feedback

Implementation Considerations

- Data Security
- Model Explainability (maintain audit trails of automated decisions)
- Performance monitoring

Conclusion

This solution represents a significant advancement over traditional manual approaches to regulatory data profiling, combining pattern recognition capabilities of supervised ML with contextual understanding of LLMs to create more robust, adaptive compliance system.