Flipkart

GRiD

5.0

Problem Statement Title: Compliance Monitoring and Enforcement through Log Analysis

Team Name: Loophole

Team members details

Team Name	Loophole		
Institute Name/Names	Indian Institute Of Information Technology, Allahabad		
Team Members >	1 (Leader)	2	3
Name	Shubhendra Gautam	Subham Panda	Saikat Shadhukhan
Batch	2025	2025	2025

Deliverables/Expectations for Level 2 (Idea + Code Submission)

Project Documentation: *Documentation Link*

Project Code: *Code link* / *Google Colab Link*

Data Files: *Files link*

Explanation Video: *Video Link*

All In One: *Folder Link*

Use-cases

P0 (Critical):

- Real-time Compliance Monitoring: Detect non-compliant activities instantly to prevent potential breaches.

P1 (Important):

- Historical Audit and Analysis: Examine past logs to identify vulnerabilities and enhance policies.
- Automated Remediation Recommendations: Provide actionable solutions for detected non-compliance.

P2 (Beneficial but not urgent):

- User Behavior Analysis: Refine security policies based on user behaviors.
- Compliance Reporting: Generate visual reports showcasing compliance metrics.

P3 (Supplementary Features):

- Feedback Loop for Improvement: Refine the system based on user feedback.
- Anomaly Detection: Identify suspicious activities not matching any known rule.

Solution statement/ Proposed approach

Breakdown of the Problem & Solutions:

1. Data Collection & Preprocessing:
    1. Problem: Acquiring and preparing relevant log data for analysis.
    2. Solution: Use synthetic data generators or collect logs from various sources, followed by preprocessing (cleaning, tokenization) to make it suitable for the model.

2. Rule Definition & Interpretation:
    1. Problem: Understanding and codifying compliance rules.
    2. Solution: Extract rules from compliance documents and convert them into a machine-readable format. Use LLMs to infer relationships between rules and log parameters.

3. Model Selection & Training:
    1. Problem: Choosing an appropriate model and training it to identify non-compliant activities.
    2. Solution: Opt for a pre-trained LLM (e.g., GPT-2, DistilBERT) and fine-tune it on our dataset comprising logs and rules.

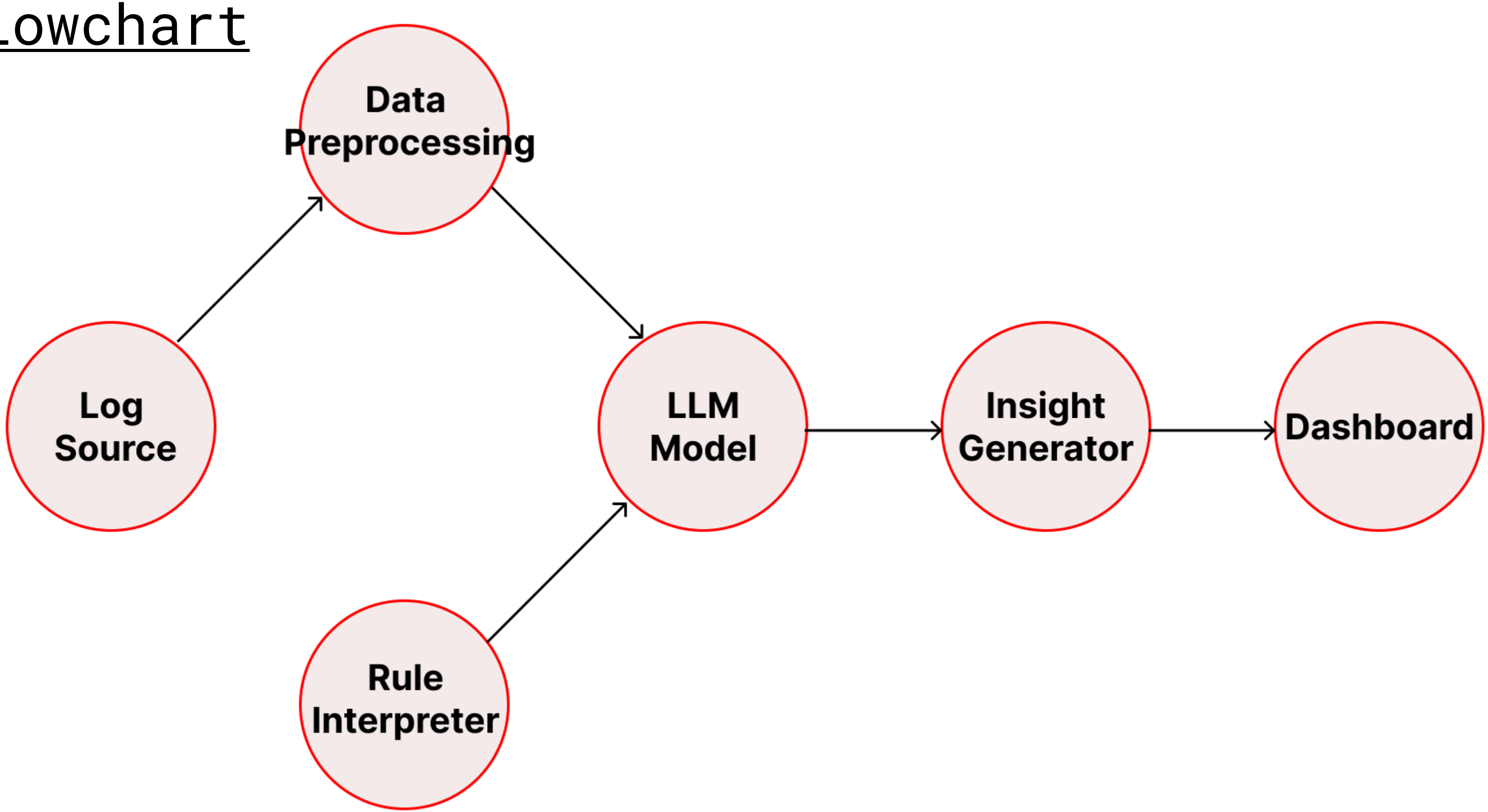
Solution statement/ Proposed approach

4. Real-time Compliance Monitoring:
    1. Problem: Analyzing logs in real-time.
    2. Solution: Implement a streaming mechanism to feed live logs into the model, ensuring immediate detection of non-compliance.

5. Insight Generation & Reporting:
    1. Problem: Interpreting the model's output to generate actionable insights.
    2. Solution: Post-process the model’s output to extract relevant information, followed by visualizing the data in a user-friendly dashboard for easy interpretation.

6. Feedback & Iteration:
    1. Problem: Continuous improvement of the system based on feedback.
    2. Solution: Incorporate feedback mechanisms allowing users to provide input on false positives/negatives. Use this feedback for iterative model refinement.

Flowchart


Data Preprocessing
Log Source
LLM Model
Insight Generator
Dashboard
Rule Interpreter

Limitations

While the design we discussed offers a robust approach to compliance monitoring using LLMs, there are inherent limitations:

1. Data Dependency:
    1. The performance of the LLM is highly dependent on the quality and quantity of the data. If the logs or rules are not comprehensive or are noisy, the model's accuracy can be compromised.

2. Model Overhead:
    1. LLMs, especially models like GPT-2 or DistilBERT, are resource-intensive. This can lead to increased computational costs and may not be suitable for real-time applications on limited hardware.

3. Generalization:
    1. The model might overfit to the training data, especially if there's a lack of diversity in the logs or rules. This can limit its ability to generalize well to new, unseen data.
4. Dependency on Pre-trained Models:
    1. The solution relies on pre-trained models which come with their own biases and limitations. If these models have not been trained on relevant data, their utility might be limited.

Limitations

5. Interpretability:
    1. Deep learning models, including LLMs, often act as black boxes. It can be challenging to understand why a particular log was flagged as non-compliant, which might be crucial for certain regulatory environments.

6. Adaptability Concerns:
    1. While the system is designed to be adaptive, constantly changing rules or log formats might require frequent retraining or model adjustments.

7. False Positives/Negatives:
    1. No model is perfect. There might be instances where compliant activities are flagged or non-compliant activities are missed, leading to potential security or operational issues.
8. Scaling Issues:
    1. As the volume of logs increases, real-time processing can become challenging, potentially leading to lags or missed logs.
9. Rule Complexity:
    1. Some compliance rules might be too intricate or nuanced for the model to fully grasp, especially if they involve multi-step logical conditions or are context-dependent.

Future Scope

Future Scope & Enhancements:

1.Model Upgrades: Transition to newer, more efficient models as they emerge.

2.Real-time Learning: Allow the system to adapt in real-time from incoming data.

3.Advanced Visualization: Develop intuitive dashboards for deeper insights.

4.Multi-modal Input: Accept various data types like audio, video, or network traffic.

5.Automated Actions: Automate corrective actions for certain non-compliant activities.

6.System Integration: Integrate with SIEM tools, IAM systems, and cloud platforms.

Future Scope

7. Reinforcement Feedback: Refine decisions using feedback loops.

8. Anomaly Detection: Identify unusual patterns indicative of potential issues.

9. Enhanced Privacy: Implement advanced data encryption and security protocols.

10. User-friendly Rule Interface: Allow easy rule definitions or modifications.

11. Improved Scalability: Architectural enhancements for handling increased data.

12. Cross-Platform Compatibility: Ensure wide-ranging platform and OS support.

These points emphasize the potential evolution of the project to cater to broader compliance management needs and proactive prevention.

Flipkart

GRiD

5.0

Thank You