Functional Requirements Document (FRD)

AI-Enhanced Data Privacy and Compliance Monitoring with AI Agents

Prepared By: Rahul Singh

Date: June 08, 2025

1 Purpose

This Functional Requirements Document (FRD) specifies the technical and functional requirements for the AI-Enhanced Data Privacy and Compliance Monitoring system, leveraging AI agents for autonomous operation. The system will use NLP, ML, and privacy-preserving techniques to classify sensitive data, detect unauthorized access, and ensure compliance with GDPR, CCPA, and HIPAA.

2 System Overview

The system comprises AI agents for:

- **Data Discovery Agent**: Identifies and classifies sensitive data using NLP and ML.
- **Monitoring Agent**: Performs real-time anomaly detection.
- **Reporting Agent**: Generates automated compliance reports.
- **Integration Agent**: Connects with SIEM, IAM, and cloud platforms.
- **Privacy-Preserving Agent**: Ensures secure data analysis with zkML or federated learning.
- **User Interface Agent**: Manages a web-based dashboard for monitoring and configuration.

3 Functional Requirements

3.1 FR1: Data Discovery and Classification

- **FR1.1**: Data Discovery Agent shall use NLP models (e.g., BERT, spaCy) to identify sensitive data (e.g., PII, health records) in unstructured formats (emails, PDFs, Word documents).
- **FR1.2**: Agent shall use ML classifiers (e.g., Random Forest, SVM) for structured data in databases (SQL, NoSQL).
- **FR1.3**: Agent shall achieve 95% classification accuracy for standard PII types.
- **FR1.4**: Users shall define custom classification rules via a web interface managed by the User Interface Agent.
- **FR1.5**: Agent shall tag data with metadata (e.g., sensitivity level, regulatory relevance) stored in PostgreSQL.

3.2 FR2: Real-Time Monitoring and Anomaly Detection

- **FR2.1**: Monitoring Agent shall track data access logs in real-time across cloud (AWS S3, Azure Blob) and on-premises environments.
- **FR2.2**: Agent shall use unsupervised ML (e.g., Isolation Forest, Autoencoders) to detect anomalies (e.g., unauthorized access, unusual transfers).

- **FR2.3**: Agent shall maintain a false positive rate below 5%.
- **FR2.4**: Agent shall prioritize alerts (low, medium, high) and display them via the User Interface Agent's dashboard.
- **FR2.5**: Alerts shall include details (user ID, data accessed, timestamp) and be sent via email, SIEM, or SMS.

3.3 FR3: Compliance Reporting

- **FR3.1**: Reporting Agent shall generate automated reports for GDPR Article 30, CCPA, and HIPAA compliance.
- **FR3.2**: Reports shall include data inventory, access logs, and incident summaries.
- **FR3.3**: Users shall schedule reports (daily, weekly) and export them in PDF/CSV via the User Interface Agent.
- **FR3.4**: Agent shall store reports for 3 years in a tamper-proof database.

3.4 FR4: Integration with Existing Systems

- **FR4.1**: Integration Agent shall connect to SIEM (Splunk, QRadar) and IAM (Okta, Azure AD) via REST APIs.
- **FR4.2**: Agent shall support cloud platforms (AWS, Azure, GCP) for data discovery and monitoring.
- **FR4.3**: Integration latency shall not exceed 500 milliseconds.
- **FR4.4**: Agent shall use OAuth 2.0 for secure authentication.

3.5 FR5: Privacy-Preserving Analysis

- **FR5.1**: Privacy-Preserving Agent shall use zkML or federated learning for secure data analysis.
- **FR5.2**: No sensitive data shall be exposed during model training or inference.
- **FR5.3**: Agent shall log operations in a tamper-proof audit trail (e.g., Hyperledger).

3.6 FR6: User Interface

- **FR6.1**: User Interface Agent shall provide a web-based dashboard for data classification, monitoring, and reporting.
- **FR6.2**: Dashboard shall display real-time alerts, data inventory, and compliance status.
- **FR6.3**: Agent shall enforce RBAC (admin, auditor, analyst).
- **FR6.4**: Dashboard shall support multi-language options for global teams.

4 Non-Functional Requirements

4.1 NFR1: Performance

- Agents shall process 1 petabyte of data with <1-second anomaly detection latency.
- System shall handle 10,000 concurrent data access events without degradation.

4.2 NFR2: Scalability

- Agents shall scale horizontally using cloud-native auto-scaling (AWS, Azure, GCP).
- Kubernetes shall orchestrate agent deployment.

4.3 NFR3: Security

- Encrypt data in transit and at rest with AES-256.
- Require MFA for dashboard access.

4.4 NFR4: Reliability

- Achieve 99.9% uptime with automated failover.
- Implement monitoring with Prometheus and Grafana.

4.5 NFR5: Compliance

- Comply with GDPR, CCPA, HIPAA, and ISO 27001.
- Store tamper-proof audit logs for 3 years.

5 Technical Specifications

- **AI/ML Frameworks**: TensorFlow, PyTorch, spaCy for NLP/ML tasks.
- **Programming Languages**: Python for agent logic, JavaScript (React) for dashboard.
- **Infrastructure**: AWS, Azure, GCP, and on-premises with Kubernetes.
- **Database**: PostgreSQL (metadata, reports), Elasticsearch (logs, unstructured data).
- **APIs**: RESTful APIs with OAuth 2.0.
- **Privacy Tech**: zkML Prover, TensorFlow Federated.
- **Communication**: Apache Kafka for event-driven agent interactions.

6 Assumptions

- Data sources provide API access or log exports.

- Sufficient computational resources (e.g., GPUs) for agent training.
- Cloud providers support required APIs.

7 Constraints

- **Budget**: $500,000 for software, hardware, and agent deployment.
- **Timeline**: 12 months, with pilot in 6 months.
- **Regulatory Compliance**: GDPR, CCPA, HIPAA adherence.
- **Legacy Systems**: Integrate without major overhauls.

8 Dependencies

- High-quality training data for agents.
- Vendor support for SIEM/IAM integrations.
- IT team training on agent management.

9 Testing Requirements

- **Unit Testing**: Validate agent accuracy (95% for classification/anomaly detection).
- **Integration Testing**: Ensure seamless SIEM, IAM, and cloud integration.
- **Performance Testing**: Verify scalability (1 petabyte, 10,000 events).
- **Security Testing**: Conduct penetration testing for adversarial attack resistance.

10 Deliverables

- AI agent-based monitoring system.
- Web-based dashboard for agent management.
- Documentation (user manuals, technical guides).
- Training for IT and compliance teams.