# Document Metadata

author: jfraze@mycomp.org

title: Model Monitoring and Event Logging for GenAI Systems

nist\_function: Manage

priority\_phase: Must

last\_reviewed: 2025-06-24

status: draft

## **Purpose**

To require real-time and post-hoc monitoring of generative AI system behavior, with event logging for traceability, drift detection, incident response, and compliance with hospital policies.

## **Scope**

Covers all GenAI systems deployed in production environments, including clinical NLP tools, LLM-powered bots, and external API integrations with content generation capabilities.

## **Policy Statement**

All deployed GenAI systems must support active behavioral monitoring and log critical events—including model output anomalies, drift indicators, and inference boundary violations—to designated governance bodies and security tools.

## **Roles and Responsibilities**

• AI System Owners: Ensure runtime telemetry, prompt/output logging, and drift detection are operational.

• IT Security Team: Integrate logs with SIEM and incident response pipelines (NIST 800-61r3).

• Compliance & Risk: Monitor logs for safety, privacy, and regulatory triggers.

• Clinical Leads: Validate alerts tied to clinical safety and patient interaction.

## **Implementation Phases**

### **Must Do**

• Log prompts and outputs under defined security classifications.

• Detect and alert on hallucination rate thresholds and unexpected behavior (AI 600-1 §2.3.1).

• Apply inference-time logging and model drift metrics (MANAGE-2.1).

### **Should Do**

• Use anomaly scoring and natural language output classifiers.

• Store logs in tamper-evident formats and retain per audit standards.

• Link logs to role-based dashboards for clinical governance review.

### **Recommended**

• Integrate logging with real-time safety triggers (e.g., “kill switches”).

• Apply federated signal sharing for jailbreaking and misuse indicators (top-AI-vuln #4, #6).

• Use synthetic output watermarking for downstream traceability (AI 600-1 §2.4.1).

## **References**

• NIST AI RMF: MANAGE-2.1, MANAGE-4.1

• AI 600-1: §2.3.1, §2.4.1

• NIST SP 800-61r3: Incident response guidelines

• HIPAA §164.308(a)(1)(ii)(D)

## **Review Cycle**

Post-deployment and quarterly review; immediately after any detected model drift or safety incident.