# Document Metadata

author: jfraze@mycomp.org

title: Incident Response and Model Escalation Procedures

nist\_function: Manage

priority\_phase: Must

last\_reviewed: 2025-06-24

status: draft

## **Purpose**

To define structured incident response and escalation procedures for failures, breaches, or anomalies in generative AI system behavior—supporting containment, remediation, and legal/compliance reporting.

## **Scope**

Applies to all GenAI systems producing patient-visible outputs, interacting with protected health data, or integrating third-party generative models within hospital systems.

## **Policy Statement**

GenAI-related incidents—including hallucinations, prompt injection, unauthorized PHI exposure, and model drift—must trigger documented escalation and response workflows coordinated with cybersecurity, clinical, and compliance stakeholders.

## **Roles and Responsibilities**

• Security Operations: Own GenAI incident detection, triage, and containment (aligned to NIST 800-61r3).

• AI Governance Board: Review root cause and corrective actions post-incident.

• Clinical Risk Managers: Escalate AI-driven misdiagnoses or unsafe recommendations.

• Legal & Compliance: Coordinate disclosure, audit logs, and external regulatory reporting.

## **Implementation Phases**

### **Must Do**

• Define GenAI-specific incident types and risk thresholds (e.g., hallucination > 5%).

• Integrate GenAI event classes into hospital-wide incident response plan (SP 800-61r3).

• Require incident reports for jailbreaking, PHI leakage, or misuse triggers.

### **Should Do**

• Conduct tabletop exercises simulating GenAI failures.

• Tag logs and incidents to affected model version and data segment.

• Escalate incidents to HHS/FDA/ONC when required by law (HIPAA, 45 CFR §164.308).

### **Recommended**

• Maintain a registry of known GenAI incident types and response playbooks.

• Share anonymized incident learnings with external safety consortia (e.g., CAIRN, CISA).

## **References**

• NIST AI RMF: MANAGE-3.1, MANAGE-4.1

• NIST AI 600-1: §2.1.3, §2.3.1

• NIST SP 800-61r3 (April 2024 IPD)

• HIPAA Security Rule: §164.308(a)(6)(ii)

## **Review Cycle**

Incident-driven, and annually during cybersecurity readiness audits.