INTERNAL REVIEW DRAFT

AI Incident Response Plan

Sample Company LLC

Prepared as a structured operating document for internal use, review, and controlled distribution.

Review this document after major operational, regulatory, or technology changes.

Version 1.0

DOCUMENT CONTROL

RECOMMENDED

PREPARED FOR Sample Company LLC	**DOCUMENT OWNER** AI Incident Response Lead
CLASSIFICATION Confidential	**DISTRIBUTION** Executive Leadership, AI Incident Response Team, IT / Security Operations
GENERATED July 21, 2026	**NEXT REVIEW DATE** 2027-07-21
STATUS Internal Review Draft	**DOCUMENT VERSION** 1.0
TEMPLATE VERSION 1.0.0	**TARGET CONTROL LEVEL** Recommended
FRAMEWORKS Core baseline only	

Table of Contents

INFO: IMPORTANT NOTICE — READ BEFORE USE

This document was generated as a structured starting template from the information provided. It does not constitute legal, regulatory, or professional compliance advice, and it does not by itself establish or prove compliance with any framework or regulation. Specific obligations vary by jurisdiction, industry, and circumstance. Before adopting or relying on it, have this document reviewed by qualified legal counsel and the appropriate internal owners, and adapt it to your organization's actual environment.

Preparation

Executive Summary

This AI Incident Response Plan defines how Sample Company LLC detects, responds to, and recovers from security incidents specific to artificial intelligence and machine-learning systems, covering data leakage via AI tools (sensitive data shared with AI vendors) and prompt injection attack on AI-powered features. Sample Company LLC is a Technology-sector organization of approximately 26-50 employees operating across 2-5 business locations.

Plan posture: detection draws on user/employee reports of AI errors or misuse and automated monitoring of AI tool usage patterns, the pre-defined containment actions are disable/suspend AI tool or feature and revoke AI API keys and access tokens, and AI incidents flow through the existing IT incident reporting system.

Conventional security controls — signature-based detection, endpoint agents, network flow analysis — do not identify model poisoning, prompt injection, training-data contamination, or AI hallucination cascades. This plan is aligned with the NIST AI Risk Management Framework (AI RMF) MANAGE function and the emerging requirements of the EU AI Act.

The sections that follow define detection mechanisms, containment actions, investigation procedures, recovery steps, and AI incident reporting. The plan integrates with the broader Incident Response Plan but maintains AI-specific runbooks where general-purpose IR procedures are insufficient.

The organization is also subject to the following frameworks, which this document acknowledges but does not map specific controls to. Local review is required to confirm alignment with each framework's requirements before distribution:

- HIPAA
- GDPR

AI Incident Categories

This plan addresses 2 AI-specific incident categories, each mapped to distinct containment options, notification requirements, and forensic evidence needs. Triage assigns a category early so the appropriate response track activates without delay.

AI Incident Categories

INCIDENT CATEGORY	DESCRIPTION
Data Leakage via AI Tools	Sensitive, proprietary, or regulated data is transmitted to an AI vendor through prompts, file uploads, or API calls, potentially exposing the data to unauthorized processing or retention.
Prompt Injection Attack	An adversary crafts malicious input to manipulate an AI system into bypassing safety controls, disclosing confidential information, or performing unauthorized actions.

Detection & Analysis

AI Incident Detection

AI incidents often lack the technical signatures of traditional IT security events — there is no packet capture for prompt injection and no endpoint agent for model drift. Detection relies on behavioral baselines, output-quality monitoring, and a culture of user reporting. The mechanisms below cover both the pipeline (API volumes, latency, output distributions) and the human layer.

- User/employee reports of AI errors or misuse
- Automated monitoring of AI tool usage patterns

AI incidents are reported through the existing IT incident reporting system. This ensures a unified triage process and consistent tracking across all incident types.

All employees who interact with AI tools are responsible for reporting suspected AI incidents immediately upon discovery. Training on recognizing AI-specific incident indicators is included in the security awareness program.

Response & Recovery

AI Incident Containment

AI containment may involve vendor coordination, model-version locks, or data-pipeline suspension that takes hours and cannot be undone without a formal re-deployment. The pre-approved actions below avoid ad-hoc decisions that create new risks during an incident. The selected actions provide a hard stop (disabling the tool or feature entirely) and credential revocation.

The following immediate containment actions are authorized upon confirmation of an AI incident.

- Disable/suspend AI tool or feature
- Revoke AI API keys and access tokens

API Key Revocation Procedure

Operating procedures cover AI API key revocation procedure; reviews of AI API key revocation procedure run on the documented cadence, with exceptions tracked to closure.

Data Leakage Containment Procedure

Operating procedures cover containment procedure for data leaked to AI tools; reviews of containment procedure for data leaked to AI tools run on the documented cadence, with exceptions tracked to closure.

AI Incident Investigation

AI investigations reconstruct causality from probabilistic model outputs and a partial telemetry surface — prompt logs, model versions, training-data provenance, and (for vendor-hosted systems) only what the vendor exposes via API logs. The evidence and root-cause categories below define the investigation surface.

- AI conversation/prompt logs
- AI tool usage audit trails

Investigators should evaluate the following root cause categories when determining the origin and contributing factors of an AI incident:

- User error (shared sensitive data accidentally)
- Policy violation (intentional misuse)

All investigation findings must be documented in the incident record. The investigation report should include a timeline of events, evidence collected, root cause determination, and recommendations for preventing recurrence.

AI Incident Recovery

AI recovery may require model rollback, data remediation, vendor coordination, or stakeholder trust rebuild. Data shared with an AI vendor often cannot be fully recovered — recovery planning accounts for that and pre-drafts stakeholder communications for the one-directional case.

AI Service Restoration Procedure

Operating procedures cover AI service restoration procedure; reviews of AI service restoration procedure run on the documented cadence, with exceptions tracked to closure.

Data Remediation After AI Data Exposure

Operating procedures cover data remediation after AI data exposure; reviews of data remediation after AI data exposure run on the documented cadence, with exceptions tracked to closure.