AI Incident Response Playbook

Type: Process Playbook | **Audience:** Incident Response Teams, Operations, AI Program Leads

1. Incident Severity Levels

Severity	Definition	Response	Examples
SEV-1 Critical	Widespread harm, safety risk, major impact	Minutes	Discriminatory outputs at scale
SEV-2 High	Significant harm, limited scope	Hours	Bias in production, compliance violation
SEV-3 Medium	Moderate impact, recoverable	24-48 hrs	Performance degradation
SEV-4 Low	Minor impact, cosmetic	1 week	Edge case failures

2. Incident Response Phases

Phase 1: Detection and Triage (0-15 min)

☐ Acknowledge alert/report and verify legitimacy
☐ Assess initial severity level and notify incident commander
☐ Begin incident documentation

Phase 2: Containment (15-60 min)

☐ Option A: Rollback - Revert to previous known-good version
☐ Option B: Fallback - Switch to simpler/safer backup system
☐ Option C: Disable - Take system offline entirely
☐ Option D: Human Review - Add manual oversight layer

Phase 3: Investigation (1-24 hours)

☐ Conduct root cause analysis
☐ Determine scope of impact and identify all affected parties
☐ Gather and preserve evidence

Phase 4: Remediation (24-72 hours)

☐ Implement and test fix in dev/staging environments
☐ Gradual rollout with monitoring
☐ Verify fix effectiveness

Phase 5: Recovery and Redress

□ Restore normal operations
□ Notify affected parties of resolution
□ Provide compensation if applicable and update stakeholders

Phase 6: Post-Incident Review

□ Conduct blameless retrospective and document lessons learned
□ Update playbooks, monitoring, and implement preventive measures

3. Communication Templates

Internal Notification: [SEV-X] AI Incident - [System] | Status: [Investigating/Mitigating/Resolved] | Impact: [Desc] | Actions: [Current] | Next Update: [Time]

External/Regulatory: Date: [Date] | System: [Name] | Nature: [Desc] | Affected: [Number] | Remediation: [Actions] | Prevention: [Safeguards]

4. Incident Documentation

Field	Details
Incident ID	
Date/Time Detected	
Severity Level	[ ] SEV-1 [ ] SEV-2 [ ] SEV-3 [ ] SEV-4
System Affected	
Incident Commander	
Root Cause	
Resolution	
Date/Time Resolved	