COURSE SYLLABUS

AI Security Monitoring and Incident Response (Self-Paced)

Eight hours of self-paced AI security monitoring and incident response: the four monitoring categories, a signals table that maps what you observe to what it means, and the containment actions specific to AI systems.

Group classes (Live Online) and onsite training are available for this course. For more information, email corporate@nobledesktop.com or visit: https://www.nobledesktop.com/classes/ai-security-monitoring-and-incident-response-self-paced



hello@nobledesktop.com • (212) 226-4149

Course Outline

Module 1

- Foundations: The AI Security Landscape and Federal Policy Context (1h50)
- Why AI systems need a different security lens
- The shared vocabulary of data poisoning, adversarial examples, prompt injection, model drift, model extraction and supply-chain risk
- What counts as an AI incident, covering malicious, accidental and environmental causes
- The specific current guidance that bears on monitoring and incident response
- Who you work with day to day

Module 2

- AI Security Monitoring Practices (1h50)
- What continuous monitoring means for high-impact AI systems
- The four monitoring categories of data quality and integrity, model performance and behavior, access and usage, and output and impact
- Baseline and anomaly detection, threshold alerting, adversarial input detection and human-in-the-loop review
- Risk-based security logging built on continuous event monitoring and threat hunting capability areas
- A signals table mapping observed patterns to likely causes

Module 3

- Identifying and Classifying AI Security Incidents (1h50)
- Recognizing an AI security incident
- The evidence and artifacts unique to AI incidents
- Classifying severity by impact on safety, legal rights, benefits, critical infrastructure and sensitive data
- The initial triage sequence
- A prompt-injection case scenario worked end to end

Module 4

- AI Incident Response, Reporting, and Post-Incident Actions (1h50)
- The four-phase lifecycle of preparation, detection and analysis, containment, eradication and recovery, and post-incident activity, each mapped to AI-specific activity
- Containment actions specific to AI systems
- Internal reporting and escalation
- Root cause analysis, updating impact assessments and monitoring plans, revalidation before return to automated operation, and when stakeholder communication is warranted