AI Ethics, Security & Governance for Leaders

Course ISI-1620 1 Day Instructor-led, Hands on

Course Description

This practical one-day course equips business and operations leaders with the knowledge and tools to implement safe, responsible AI adoption through effective governance, risk management, and security controls. Students will learn to recognize common AI risks including privacy violations, IP leakage, bias, and data exposure, and apply proven safe-use patterns and review processes. The course is tool-agnostic and emphasizes hands-on exercises including red-teaming, evaluation design, and policy creation.

Who Should Attend?

This course is designed for business and operations leaders, risk partners, program owners, compliance officers, and anyone responsible for safe AI adoption in their organization. No technical background is required—the focus is on practical risk management, governance, and organizational enablement.

Course Objectives

In this course, you will learn how to:

- Recognize and assess common AI risks: privacy violations, intellectual property leakage, bias, and data exposure
- Apply safe-use patterns and review processes to mitigate identified risks
- Design and implement evaluation frameworks using golden datasets, rubrics, and audit processes
- Conduct vendor assessments using model cards, data processing agreements (DPAs), and logging requirements
- Perform lightweight red-team exercises to identify vulnerabilities including prompt injection, over-reliance, and jailbreaks
- Create governance frameworks and policies that balance innovation with risk management and compliance
- Align AI initiatives with legal, compliance, and IT requirements without impeding delivery timelines.

Topics

- AI risk landscape: privacy, IP protection, bias, and data leakage
- Stakeholder roles and responsibilities: business, IT, legal, and vendor management

- Data protection in practice: PII handling, IP safeguards, prompt security, and output validation
- Prompt de-risking techniques and secure prompt engineering patterns
- Evaluation frameworks: golden dataset construction, rubric design, and audit processes
- Evaluation card templates and documentation standards
- Vendor procurement and assessment: model cards, DPAs, logging, and compliance requirements
- Red-teaming methodologies: prompt injection, jailbreaks, over-reliance testing
- Governance frameworks: RACI matrices, change control, and communication strategies
- Responsible AI policy development and implementation

Prerequisites

Prior to attending, students should have:

- No specific technical prerequisites required
- Understanding of organizational governance and risk management concepts (helpful but not required)
- Interest in responsible AI adoption and risk mitigation
- Modern web browser with internet access
- Optional: Organization's policy documents for tailoring exercises to specific context

Course Outline

Module 1: AI Risk Landscape and Stakeholder Roles

- Understanding AI-Specific Risks
    - Privacy violations and PII exposure
    - Intellectual property leakage and confidential data risks
    - Bias, fairness, and discrimination concerns
    - Hallucinations, accuracy, and reliability issues
- Stakeholder Ecosystem
    - Business ownership and accountability
    - IT and security responsibilities
    - Legal and compliance requirements
    - Vendor management and procurement oversight
- Risk Assessment Framework
    - Identifying high-risk use cases
    - Risk scoring and prioritization
    - Mitigation strategy selection

Module 2: Data Protection in Practice

- PII and Sensitive Data Handling
    - Identifying PII in prompts and outputs
    - Data minimization strategies
    - Anonymization and pseudonymization techniques
- Intellectual Property Protection
    - Preventing confidential data leakage in prompts
    - Output review and IP filtering
    - Model training data considerations

- Hands-On Lab 1: De-Risk This Prompt
    - Analyze problematic prompts for data exposure
    - Rewrite prompts to prevent leakage
    - Identify and mitigate bias in prompt design
    - Test improved prompts for safety

Module 3: Evaluation and Monitoring

- Building Evaluation Frameworks
    - Golden dataset construction and management
    - Rubric design for systematic assessment
    - Human evaluation vs. automated metrics
- Audit and Compliance Processes
    - Regular review cycles and cadence
    - Documentation requirements
    - Compliance reporting and audit trails
- Hands-On Lab 2: Build an Evaluation Card
    - Select a real organizational use case
    - Define evaluation criteria and rubrics
    - Create golden dataset examples
    - Design review process and documentation
    - Complete evaluation card template

Module 4: Procurement and Vendor Management

- Model Cards and Documentation Review
    - Understanding model capabilities and limitations
    - Training data composition and bias assessment
    - Performance metrics and benchmarks
- Data Processing Agreements
    - DPA requirements and key provisions
    - Data residency and sovereignty
    - Subprocessor management
- Logging and Monitoring Requirements
    - Audit log specifications
    - Retention policies and access controls
    - Incident response and breach notification

Module 5: Red-Teaming Fundamentals

- Common Attack Vectors
    - Prompt injection and manipulation
    - Jailbreaks and constraint circumvention
    - Over-reliance and automation bias
- Red-Team Exercise Design
    - Threat modeling for AI systems
    - Test scenario development

Contact ISInc for more information at 916.920.1700 or by visiting our website at http://www.isinc.com

    - Finding documentation and severity assessment
- Hands-On Lab 3: Red-Team a Simple Assistant
    - Set up test environment with sample assistant
    - Attempt prompt injection attacks
    - Test for jailbreaks and constraint violations
    - Document findings with severity ratings
    - Propose mitigations and fixes

Module 6: Governance Without Friction

- Organizational Structures
    - RACI matrix for AI initiatives
    - Centers of Excellence and champions programs
    - Review boards and approval processes
- Change Control and Communication
    - Lightweight review processes
    - Stakeholder communication strategies
    - Balancing speed with safety
- Hands-On Lab 4 - Capstone: Responsible AI Policy and Checklist
    - Draft 2-page Responsible AI policy for your organization
    - Create pre-deployment review checklist
    - Define escalation paths and approval workflows
    - Design enablement materials for business users
    - Present policy and implementation plan