

**Foundations of Responsible AI with The NIST AI Risk Management Framework**

**Course Number:** AI-146WA
**Duration:** 2 days

**Overview**

This Responsible AI course is about not just theoretical knowledge but also its practical application in real-world scenarios. It teaches the knowledge and skills needed to responsibly design, develop, and deploy artificial intelligence (AI) systems.

The training investigates the NIST (US National Institute of Standards and Technology) AI Risk Management Framework (AI RMF), the premier AI guidelines for US-based private and public sector companies. Participants gain practical experience applying the AI RMF to real-world scenarios through interactive discussions, case studies, and hands-on exercises.

**Prerequisites**

No prior experience is required.

**Materials**

All Responsible AI training students receive comprehensive courseware.

**Software Needed on Each Student PC**

Students should have Zoom installed as the conference platform.

**Objectives**

* Identify and assess risks associated with AI systems throughout their lifecycle, from design to deployment and beyond
* Implement the NIST AI Risk Management Framework (RMF) to govern, map, measure, and manage AI risks effectively
* Develop trustworthy AI systems that are safe, secure, explainable, privacy-enhanced, fair, accountable, and reliable
* Apply risk management strategies to mitigate or avoid AI risks, ensuring the responsible and ethical use of AI technologies
* Evaluate the trustworthiness of AI systems using established criteria and identify potential violations
* Design comprehensive risk assessment processes for AI systems, considering factors such as categorization, limitations, and cost-benefit analysis
* Measure AI system performance and risks using appropriate metrics and testing procedures
* Prioritize and manage AI risks effectively, implementing mitigation strategies, transfer mechanisms, or acceptance plans when necessary
* Evaluate the specific risks associated with generative AI, such as hallucinations, data privacy, and abusive content
* Conduct end-to-end AI risk management case studies to apply knowledge and skills in real-world scenarios

**Outline**

* Introduction to Risk Management
	+ What is Risk?
	+ What is Risk Management?
	+ Why a Risk Management Framework?
	+ The NIST Risk Management Framework (RMF)
	+ Frame, Assess, Respond, Monitor
	+ Risk Management Roles
	+ Prepare
	+ Categorize
	+ Select
	+ Implement
	+ Assess
	+ Authorize
	+ Monitor
	+ Legacy of the NIST RMF
	+ Case Study: Risky Business
* Introduction to AI and Its Societal Impact
	+ What is Artificial Intelligence?
		- Defining terms
		- Brief history and overview of AI technologies
	+ Understanding AI’s Potential Benefits
	+ Understanding AI’s Potential Harms
	+ The Need for Responsible AI
	+ Introduction to AI Systems
	+ The AI System Lifecycle
	+ Common AI Actors
	+ Assigning activities to AI lifecycle phases
* Understanding AI Risks and the NIST AI Risk Management Framework (RMF)
	+ What is AI Risk?
	+ AI Risks vs. Traditional Software Risks
	+ Challenges in AI Risk Management
	+ Introduction to the NIST AI RMF
	+ Attributes of the NIST AI RMF
	+ Alternatives to the NIST AI RMF
	+ Core Functions: Govern, map, measure, manage
	+ Is it an AI risk?
* Understanding Trustworthy AI with Case Studies
	+ Safe
	+ Secure & Resilient
	+ Explainable & Interpretable
	+ Privacy-Enhanced
	+ Fair
	+ Accountable & Transparent
	+ Valid & Reliable
	+ Issue Spotting: Violations of Trustworthy AI
* Govern
	+ Policies, processes, procedures, and practices
	+ Accountability structures
	+ Diverse Input
	+ Culture & Communication
	+ Engagement Processes
	+ Policies and Procedures
	+ Contingency Plans
	+ Issue Spotting: Governance Structures
* Map
	+ AI in Context: Application
	+ AI in Context: Actors
	+ AI in Context: Aspirations
	+ Risk Assessment Process
	+ Categorization of AI
	+ AI Limitations and oversight
	+ Cost—Benefit Analysis
	+ AI System Scoping
	+ AI Operator Training
	+ Lab – Finding Risks in Sample Applications
* Measure
	+ Fit-for-Purpose
	+ Testing procedures
	+ Acceptable limits
	+ Categories of metrics
	+ Data lineage
	+ Documenting errors
	+ Evaluating TEVV processes
	+ Perverse incentives
	+ Independent testing & red teaming
	+ Production monitoring & safety statistics
	+ Chaos Engineering
	+ Labeling hazards
	+ Fairness and bias—in all forms
	+ Environmental impacts
	+ Comparison tests
	+ Emergent risks
	+ Frontier risks
	+ Measuring the metrics
	+ Feedback loops
	+ Gas station bathrooms & effective communication
	+ Risk Report Design
* Manage
	+ Go / No Go – When to develop and use AI
	+ AI Risk Prioritization
	+ Mitigation
	+ Transfer
	+ Avoidance
	+ Acceptance
	+ Insurance
	+ Sustaining value
	+ Emergency Stop
	+ Supply-chain Risks
	+ Pre-trained models
	+ Sunset clauses
	+ Continual improvement
	+ Incident response
	+ Transfer/Avoid/Mitigate/Accept
* AI RMF: Generative AI Profile
	+ CBRN Information
	+ Hallucinations
	+ Dangerous output
	+ Data privacy
	+ Environmental impact
	+ Human-AI Configuration
	+ Information Integrity
	+ Cybersecurity
	+ Intellectual Property / Controlled Information
	+ Abusive Content
	+ Toxicity, Bias, and Homogenization
	+ Supply Chain
	+ Evaluating GenAI Risk
* Capstone
	+ End-to-End AI Risk Management Case Study