CONCEPT

AI Risk Assessment, Model Validation and AI Assurance



AI Risk Assessment, Model Validation and AI Assurance

Course Overview

Enterprise adoption of artificial intelligence in automated credit decisioning, fraud detection, customer engagement, and business operations has outpaced established risk governance routines. Deployed models frequently operate without formal catalogue documentation, undergo performance testing restricted solely to baseline accuracy, and face degradation when production data distributions diverge. Generative architectures introduce further operational exposures, including hallucinated outputs, malicious prompt injection, and intellectual property leakage. This programme delivers technical methods for risk professionals, validation leads, and assurance teams to rigorously evaluate machine learning pipelines against internationally recognized standards across all development and production phases. Practical case scenarios guide practitioners through impact classification, technical performance verification, bias mitigation, and ongoing monitoring, culminating in a defensible deliverable structured by Core Concept.

Course Objectives

- Establish and curate an operational artificial intelligence registry featuring defensible risk tiering aligned with ISO/IEC 42001 and ISO/IEC 5338 life cycle principles.
- Execute structured impact and risk assessments applying the analytical stages defined in ISO/IEC 23894 and ISO/IEC 42005.
- Validate machine learning model performance, algorithmic fairness, technical robustness, and explainability against defined statistical acceptance thresholds.
- Analyze generative architecture vulnerabilities, mitigating prompt injection, sensitive data leakage, and confabulation hazards.
- Construct monitoring protocols for production drift detection, stability metrics, and empirical revalidation thresholds.
- Synthesize an evidence-based assessment file to support independent evaluation and model committee scrutiny.

Target Audience

- Enterprise risk analysts and managers embedding machine learning oversight into operational risk profiles
- Model validation specialists responsible for technical verification and algorithmic challenge
- Internal assurance practitioners planning independent evaluations of automated decision systems
- Machine learning engineers and data science leads preparing model cards and validation documentation
- Compliance advisors and privacy officers evaluating algorithmic impacts on end users
- Technology procurement leads reviewing external algorithmic systems and third-party foundation models

Course Outline

Day 1: AI Risk Landscape and System Inventory

- Algorithmic system typologies and lifecycle stages within ISO/IEC 5338 engineering processes
- Systemic failure analysis covering bias propagation, confabulation, and autonomous error cascades
- Trustworthy system principles and foundational governance baselines
- Registry assembly, metadata requirements, and use-case risk classification
- Internal governance diagnostic and procedural gap analysis

Day 2: AI Standards and Risk Governance Frameworks

- Management system specifications and Annex A operational controls in ISO/IEC 42001
- Risk management guidelines and assessment workflows under ISO/IEC 23894
- Core governance functions covering operational mapping, measurement, and treatment
- Generative architecture risk profiles and domain-specific vulnerability taxonomies
- System quality characteristics and verification criteria defined in ISO/IEC 25059

Day 3: Impact Assessment and Model Validation Methodologies

- System impact assessments utilizing ISO/IEC 42005 structured methodologies
- Documentation auditing across data provenance sheets and model specification cards
- Empirical validation techniques covering holdout partitions, confusion matrices, and threshold tuning
- Disparate impact analysis and demographic parity testing under ISO/IEC TR 24027
- Feature importance and local explainability verification using SHAP and LIME frameworks

Day 4: Advanced Vulnerabilities, Threat Controls, and Ongoing Monitoring

- Vulnerability mitigation based on the OWASP Top 10 for Large Language Model Applications
- Adversarial attack categorization and evasion testing aligned with MITRE ATLAS
- External model assurance questionnaires and vendor supply chain verification
- Statistical drift tracking using the Population Stability Index and divergence indicators
- Human-in-the-loop oversight workflows and empirical revalidation trigger design

Day 5: Practical Evaluation and Assurance Pack Assembly

- Retail credit scoring evaluation reviewing validation metrics and threshold challenges
- Customer service conversational assistant testing covering prompt security and output bounds
- Risk catalogue structuring and internal control mapping for an active enterprise model
- Compilation of the complete risk assessment and model assurance documentation
- Simulated model risk panel defence presenting independent assurance findings

Skills You Will Gain

- Algorithmic Risk Classification
- Model Validation Testing
- Fairness and Bias Auditing
- Model Explainability Analysis
- Generative System Threat Mitigation
- AI Risk Control Mapping
- Algorithmic Vendor Due Diligence
- Model Assurance Reporting

Why Attend This Course

- Acquire structured testing protocols to verify that machine learning models perform reliably across changing operating conditions.
- Formulate precise technical questions to identify unverified assumptions in data science pipelines and vendor models.
- Supply executive stakeholders and oversight bodies with structured evaluations of high-exposure algorithmic deployments.
- Leave equipped with a structured, repeatable model assessment pack evaluated through rigorous peer challenge.

Conclusion

Sustained value from machine learning systems requires disciplined oversight prior to production release and continuous verification thereafter. This training traces the full trajectory of algorithmic risk, moving from system inventories and governing standards to empirical validation metrics, generative vulnerability controls, and drift detection. By structuring evidence across real-world case scenarios, practitioners acquire a standard-aligned framework to independently assess, challenge, and validate artificial intelligence deployments across their enterprise.