CONCEPT

Agentic AI Governance: Designing and Controlling Autonomous AI Agents



Agentic AI Governance: Designing and Controlling Autonomous AI Agents

Technical depth: Practitioner · **Practical mode:** Case study

Introduction

Organisations are moving from AI assistants that answer questions to AI agents that plan tasks, call tools and act on business systems with delegated authority. Without clear limits, an agent can approve a payment, expose data or chain errors across processes faster than anyone notices. This Core Concept course equips managers to design, govern and control autonomous agents using recognised AI management, risk and security frameworks. Participants apply each method to case material from several sectors and leave with an Agent Governance and Control Plan for a use case in their own organisation.

Course Objectives

- Distinguish agentic systems from conventional generative AI tools and classify proposed use cases by autonomy, reversibility and potential impact
- Select agent architectures, reasoning patterns and interoperability protocols, including the Model Context Protocol, suited to a given business process
- Specify agent goals, tool permissions, approval checkpoints and stop conditions using least-privilege design
- Assess agent threats and risks with the OWASP Top 10 for Agentic Applications, CSA MAESTRO and ISO/IEC 23894
- Establish evaluation, runtime monitoring and intervention mechanisms that keep agents within their delegated authority
- Produce an Agent Governance and Control Plan ready for submission to an AI governance committee

Target Audience

- Managers who own business processes where AI agents are being piloted or scaled
- AI product and solution owners responsible for specifying agent behaviour and acceptance criteria
- Enterprise and solution architecture managers approving agent integration with core systems
- AI governance, risk and compliance managers overseeing autonomous system deployments
- Information security managers responsible for identity, access and threat coverage of AI agents
- Digital transformation managers building the business case for agentic automation

Course Outline

Day 1: Agentic AI Foundations and the Use Case Landscape

- Agent Anatomy: Model, Planner, Memory, Tools and Environment
- Autonomy Spectrum from Assistant to Fully Autonomous Agent
- ISO/IEC 22989 Concepts Applied to Agent Systems
- Use Case Screening Matrix: Value, Reversibility and Blast Radius
- Current-State Agent Inventory and Shadow Agent Discovery

Day 2: Architectures, Protocols and Management Standards

- ReAct and Plan-and-Execute Reasoning Patterns
- Single-Agent, Orchestrator-Worker and Multi-Agent Topologies
- Model Context Protocol and Agent2Agent Protocol for Tool and Agent Interoperability
- ISO/IEC 42001 AI Management System Controls for Agent Deployment
- NIST AI RMF Govern, Map, Measure and Manage Functions for Agents

Day 3: Designing Controlled Agents

- Agent Specification Canvas: Goals, Permissions, Tools and Stop Conditions
- Least-Privilege Tool Access and Scoped Credential Design
- Human-in-the-Loop Approval Checkpoints and Escalation Thresholds
- Guardrail Layering: Input Filters, Output Validation and Policy Engines
- Agent Evaluation Test Suite: Task Success, Trajectory and Safety Metrics

Day 4: Agentic Threats, Impact and Runtime Control

- OWASP Top 10 for Agentic Applications: Goal Hijack, Tool Misuse and Memory Poisoning
- CSA MAESTRO Layered Threat Modelling for Multi-Agent Systems
- Indirect Prompt Injection Through Retrieved Content and Tool Outputs
- ISO/IEC 42005 Impact Assessment for Autonomous Decisions
- Runtime Monitoring, Audit Trails and Kill-Switch Procedures

Day 5: Case Work and the Agent Governance and Control Plan

- Procurement Agent Case Study: Approval Limits and Supplier Data Access
- Customer Service Agent Case Study: Refund Authority and Human Handover
- Agent Risk Register Build Aligned to ISO/IEC 23894
- Agent Governance and Control Plan Drafting
- Design Review Board and Plan Defence

Skills You Will Gain

- Agent Architecture Design
- Autonomy Risk Classification
- Tool Permission Design
- Agentic Threat Modelling
- Human Oversight Design
- Agent Evaluation and Testing
- AI Management System Controls
- Runtime Agent Monitoring

Why Attend This Course

- Take back an Agent Governance and Control Plan for a real use case, already challenged by peers acting as a design review board
- Brief technical teams and vendors precisely on what an agent may do, which tools it may call and when a person must decide
- Recognise the failure patterns that turn a successful pilot into an uncontrolled production agent
- Compare agent governance approaches with managers from finance, government, healthcare and industry

Conclusion

Agentic AI delivers value only when every agent works within authority that has been deliberately designed, tested and monitored. This course moves from the anatomy and autonomy levels of agents, through the architectures, protocols and management standards that shape them, to the threats and runtime controls that keep them within bounds. The final day turns that material into an Agent Governance and Control Plan, giving participants a structured basis for approving, constraining and overseeing the next agent their organisation proposes to deploy.