



TECHNICAL LEARNING AND ASSESSMENT DOSSIER

AI Security & Generative AI Risk – Enterprise Security Manager

Course + certificate

780,00 €

VAT excluded

Course code

AI-SEC-GENAI

Minimum duration

424 min

Final pass score

75%

Dossier issue

15/10/2026

1. Course identity, purpose and scope

Technical-management course on AI and GenAI security risks: prompt injection, data leakage, model abuse, poisoning, supply chain, agentic AI, secure architecture, monitoring and risk governance.

This extended technical dossier describes the pathway as published in CIWZ Academy on the issue date. It consolidates purpose, references, learning architecture, module contents, assessment system, passing criteria and certificate conditions. It does not replace the official texts of the standards, laws, guidelines or schemes referenced.

2. Normative, regulatory and technical references

NIST AI RMF 1.0; NIST AI 600-1 GenAI Profile; ISO/IEC 23894:2023; ISO/IEC 5338:2023; ISO/IEC 27001:2022; ISO/IEC 27090 under publication (evolving reference); OWASP guidance for LLM/GenAI

References are used as the learning and assessment framework within the limits of published content and available sources. For professional application, the official and current versions of the referenced documents always prevail.

3. Learning architecture and assessment system

- 8 structured learning modules
- minimum tracked time: 424 min
- intermediate-check bank: 256 questions
- active final-exam bank: 240 questions
- final exam: 30 questions selected from the bank
- passing threshold: 75%
- Access to the pathway through the personal Academy profile or a corporate assignment.
- Sequential study of the modules, with server-side recording of progress and required minimum time.

- Intermediate checks linked to modules to verify understanding, application capability and quality of reasoning.
- Final exam unlock only after the pathway requirements have been completed.
- Protected final exam assembled by the platform through selection from the active course question bank.
- Recording of the result and, where applicable, the option to request the final certificate.

4. Official catalogue technical information

Course technical sheet

CIWZ Academy executive/professional pathway with original content, server-tracked learning, 8 modules, randomized intermediate checks, Secure Proctor in assessments and a scenario-based final question bank. Free access; any fee applies only to certificate issuance after passing.

5. Detailed course syllabus

The following syllabus is derived from the active modules published on the platform. Each module shows its minimum study time and the size of its available intermediate-check bank.

01 Threat modelling for AI and GenAI

Minimum time: 53 min | Intermediate questions: 32

02 Prompt injection, jailbreaks and instruction hierarchy

Minimum time: 53 min | Intermediate questions: 32

03 Data leakage, privacy and secret protection

Minimum time: 53 min | Intermediate questions: 32

04 Model and data poisoning integrity

Minimum time: 53 min | Intermediate questions: 32

05 RAG, vector databases and knowledge security

Minimum time: 53 min | Intermediate questions: 32

06 Agentic AI, tool use and autonomous actions

Minimum time: 53 min | Intermediate questions: 32

07 AI supply chain, provider and model security

Minimum time: 53 min | Intermediate questions: 32

08 AI security testing, monitoring and incident response

Minimum time: 53 min | Intermediate questions: 32

6. Developed module-by-module contents

The following pages analytically describe the content actually associated with each platform module. The structure is designed so that learners, companies, HR, Quality, HSE or CISO functions can understand precisely what is studied and assessed.

MODULE 01**Threat modelling for AI and GenAI**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Distinguish assets and trust boundaries across data, prompts, embeddings, models, tools, agents, APIs, vector stores, plugins and downstream systems.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Build an AI-specific threat model linking actors, capabilities, attack paths, impact and existing controls.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Build an AI-specific threat model linking actors, capabilities, attack paths, impact and existing controls.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Architecture diagram, data flow, model inventory, trust-boundary map, misuse cases, attack trees and risk decisions.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

An AI copilot can access a knowledge base and ticketing system: assess what happens if untrusted input influences tool calls. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Share of AI components with threat models updated after architectural changes.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: Threat model copied from the web; no trust-boundary map; security limited to the model while ignoring data and integrations.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to "Threat modelling for AI and GenAI".

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an

ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

MODULE 02**Prompt injection, jailbreaks and instruction hierarchy**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Understand direct/indirect injection, jailbreaks, instruction conflicts and risks from external content being treated as trusted instructions.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Separate instructions from data, minimize privileges, validate tool calls, use content boundaries, policy enforcement and adversarial tests.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Separate instructions from data, minimize privileges, validate tool calls, use content boundaries, policy enforcement and adversarial tests.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Red-team cases, policy tests, tool-call allowlists, prompt templates, injection telemetry and blocked-action evidence.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

An indexed document contains malicious instructions executed during RAG: design defenses beyond simple input filtering. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Injection-test success rate and share of tool calls subject to server-side authorization.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: Relying on the system prompt; phrase blacklists; overprivileged tools; no output validation before action.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to "Prompt injection, jailbreaks and instruction hierarchy".

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

MODULE 03**Data leakage, privacy and secret protection**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Manage risk from sensitive prompts, provider retention, training on inputs, embeddings, logs, memorization, secret exposure and output inference.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Classify data, apply DLP, minimization, redaction, tenant controls, encryption, access control and provider-use policies.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Classify data, apply DLP, minimization, redaction, tenant controls, encryption, access control and provider-use policies.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Data classification, DLP rules, provider configuration, retention settings, access logs, privacy assessment and incident records.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

An employee pastes proprietary code into a public chatbot: manage prevention, detection, response and contractual review. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Number of sensitive-data policy violations and share of AI flows covered by DLP/retention controls.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: “Do not enter secrets” policy without enforcement; logs containing full prompts; unknown retention; unclassified embeddings.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to “Data leakage, privacy and secret protection”.

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

MODULE 04**Model and data poisoning integrity**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Protect training/fine-tuning data, feedback loops, model artefacts and pipelines from manipulation that changes behaviour, introduces backdoors or degrades performance.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Apply provenance, integrity checks, trusted sources, quarantine, validation, signed artefacts, controlled retraining and rollback.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Apply provenance, integrity checks, trusted sources, quarantine, validation, signed artefacts, controlled retraining and rollback.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Dataset lineage, approval records, hashes/signatures, retraining logs, validation reports, rollback and anomaly evidence.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

User feedback feeds automatic retraining and is manipulated by a coordinated group: redesign trust and validation gates. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Share of datasets/model versions with provenance and integrity verification before release.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: Automatic retraining without gates; unsigned artefacts; untracked data sources; untested rollback.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to "Model and data poisoning integrity".

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

MODULE 05**RAG, vector databases and knowledge security**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Assess access-control, cross-tenant leakage, malicious-document, stale-knowledge, poisoning and over-retrieval risks in RAG systems.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Enforce authorization before and after retrieval, segment indexes, validate ingestion, filter metadata, apply least privilege and source attribution.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Enforce authorization before and after retrieval, segment indexes, validate ingestion, filter metadata, apply least privilege and source attribution.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Index access rules, ingestion workflow, source provenance, retrieval logs, permission tests and stale-content controls.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

A user without document access obtains a summary through RAG: identify where authorization was lost. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Unauthorized-retrieval rate in tests and share of sources with owner/version/provenance.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: Permissions applied only in UI; shared index; unverified ingestion sources; no stale-content management.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to “RAG, vector databases and knowledge security”.

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

MODULE 06**Agentic AI, tool use and autonomous actions**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Manage risk when AI plans actions, invokes tools, changes systems, spends money or interacts with privileged accounts.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Apply least privilege, scoped credentials, approval gates, transaction limits, sandboxing, idempotency, deterministic validation and kill switches.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Apply least privilege, scoped credentials, approval gates, transaction limits, sandboxing, idempotency, deterministic validation and kill switches.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Tool registry, permission matrix, approval logs, limits, sandbox configuration, transaction evidence and emergency-stop tests.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

An agent can open tickets, run scripts and send email: define which actions require human confirmation and which technical limits apply. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Share of high-impact actions with approval gates and policy-blocked tool invocation rate.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: Shared credentials; generic tools; no transaction limits; audit logs not correlated with model decisions.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to “Agentic AI, tool use and autonomous actions”.

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

MODULE 07**AI supply chain, provider and model security**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Govern API providers, open-source models, fine-tuning services, libraries, model hubs and cloud dependencies for integrity, availability, licensing and change risk.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Perform due diligence, pinning/versioning, artefact verification, contract controls, failover, rate-limit planning and change monitoring.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Perform due diligence, pinning/versioning, artefact verification, contract controls, failover, rate-limit planning and change monitoring.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Supplier assessments, model sources, signatures/hashes, SBOM/ML-BOM where used, SLA, change notices and fallback tests.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

A provider retires a model with short notice: assess business continuity and forced-migration risk. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Share of AI components with tracked version, verified provenance and defined fallback.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: Automatically using latest version in production; unverified model downloads; single dependency without exit plan.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to "AI supply chain, provider and model security".

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

MODULE 08**AI security testing, monitoring and incident response**

- Minimum required time: 53 min
- Available intermediate-check bank: 32 questions

Executive briefing

Integrate red teaming, abuse testing, security regression, telemetry, anomaly detection, incident response and lessons learned for AI systems.

Professional objectives

By the end of the module the participant should turn the topic into decisions, controls and verification. Operational focus: Define test libraries, release gates, runtime monitoring, abuse signals, incident taxonomy, evidence preservation, rollback and communication plans.

Operational method

Apply a structured cycle: define context and purpose, set criteria before decisions, assign accountability, execute controls, collect evidence and review results. For this topic: Define test libraries, release gates, runtime monitoring, abuse signals, incident taxonomy, evidence preservation, rollback and communication plans.

Evidence and auditability

Professional quality is demonstrated by coherent, reconstructable evidence, not merely by procedures. Key evidence: Red-team reports, regression results, telemetry dashboards, incident timelines, evidence packs, rollback and corrective actions.

Decisions, risk and proportionality

Decisions should be proportionate to criticality, exposure, impact, information reliability and applicable requirements. Document assumptions, confidence, exceptions and reassessment triggers.

Applied case

A known jailbreak reappears after a model update: decide whether the issue is test coverage, change management or control effectiveness. Separate facts from hypotheses, identify missing information, choose the initial action and define what evidence will demonstrate that the decision was effective.

Indicators and weak signals

A useful indicator needs a formula, source, frequency, owner, threshold and associated decision. For this module monitor at least: Regression escape rate, time to contain AI security incidents and red-team finding closure rate.

Common failure modes

Avoid merely formal approaches. Typical weaknesses: Red team only before go-live; no security regression; missing security telemetry; AI incidents not correlated to versions.

Decision laboratory

Scenario: a condition changes suddenly (technology, supplier, requirement, volume, threat or responsibility). Build a five-step response: impact, criterion, decision, evidence, review. Apply it to "AI security testing, monitoring and incident response".

Assessment preparation

Before the assessment, make sure you can explain the topic without reading the slides, justify a decision in an ambiguous scenario, identify at least two independent pieces of evidence and distinguish a substantive response from a documentary one.

7. Assessment system, final exam and passing criteria

- Intermediate checks are linked to individual modules and provide progressive learning control before access to the final assessment.
- The final assessment consists of 30 questions selected from the active course bank. The platform applies an anti-repetition logic for new attempts when the bank size allows it.
- Final question-bank size detected at generation time: 240 active questions.
- Published passing threshold for the course: 75%.
- Questions are intended to assess not only recall but also interpretation, application, evidence selection, scenario assessment and quality of decision-making, depending on the subject matter.

8. Tracking, learning continuity and session integrity

- Path status, completed modules and recorded time are managed server-side when the course uses the Academy learning mode.
- The platform allows the pathway to resume from the recorded point and retains the results of required checks.
- Assessment sessions may be subject to the integrity and proctoring controls enabled by the platform and described in the applicable usage rules.
- Technical and integrity events support assessment and do not replace professional judgement where human review is required.

9. Certificate, use and limitations

The CIWZ certificate, where provided by the pathway and requested after passing, records the result achieved in the stated CIWZ pathway or assessment. It does not automatically replace personnel certification, professional registration, statutory licence, legal qualification or credentials issued by third-party bodies where such requirements are imposed by the applicable scheme.

10. Commercial and tax conditions

- Course + certificate: 780,00 € VAT excluded.
- UAE VAT is applied to UAE customers according to billing data and the applicable tax rules.
- Any specific commercial terms, agreements or corporate orders override the public price only when formally accepted by Certificato IWZ FZCO.

11. Dossier document control

- Document automatically generated by CIWZ Academy from the course data published at the time of the request.

- Course code: AI-SEC-GENAI
- Document language: EN
- Generation date: 15/10/2026
- For normative and professional application, the current official sources and any applicable procedures/schemes prevail.