
AI Agent Security Training Course: Threat Modelling, Access Control and Incident Response

31 May – 4 June 2027
Casablanca





AI Agent Security Training Course: Threat Modelling, Access Control and Incident Response

Ref: 597_30761 **Date:** 31 May – 4 June 2027 **Location:** Casablanca **Fees:** 4000 GBP

Programme Summary

AI agent security is the subject of this one-week training course for security and technology leaders at managerial level, ending in a control plan for one AI agent. Agents that call tools, read data and hold credentials widen the attack surface beyond that of a chat assistant, yet reviews often stop at model behaviour while permissions, tool calls and logs go untested. The programme builds a repeatable method for modelling threats, limiting access, containing agents and responding to incidents, applied to an agent the participant selects. Mars Training Centre UK delivers this programme on AI agent security to mixed-organisation cohorts.

The Business Case

An AI agent acts on behalf of the organisation, so a weakness in its permissions or inputs becomes a weakness in the systems it reaches. Where controls are thin, a manipulated instruction can trigger a payment, expose records or change a configuration before anyone notices. Teams that treat the agent as a privileged user, with tested limits and monitored actions, give executives clearer assurance and a faster route to recovery.

- Data Exposure: Falls as output filtering and data scoping stop agents from returning records they should not see.
- Unauthorised Actions: Fall as least-privilege access control limits what each agent may change.
- Recovery Time: Falls as monitoring and an agent incident response playbook shorten detection and containment.

Principles and Evidence Base

- OWASP Top 10 for LLM Applications: Its risk list, headed by prompt injection and including excessive agency, structures the threat work in Sessions 1 and 2.
- Least privilege: This access principle guides the permission design in Session 2.
- Defence in depth: This layered control model shapes the sandboxing and monitoring work in Session 3.
- Threat modelling: Its method of identifying assets, entry points and abuse paths frames Session 1 and the adversarial testing in Session 4.



Intended Participants

- Security leaders who set controls for systems that act on their own.
- Security architects who design the access, isolation and logging around agents.
- AI and engineering heads who build or buy agents and own their release.
- Risk and technology directors who need assurance that agent controls work.
- Incident response leads who must contain failures involving automated actions.

Typical sectors include banking, energy, telecommunications and public bodies.

Learning Outcomes

1. Participants will be able to construct a threat model for an AI agent from its tools, data and credentials.
2. Participants will be able to evaluate exposure to prompt injection, tool misuse and data leakage.
3. Participants will be able to design identity and least-privilege access control for agents.
4. Participants will be able to prioritise sandboxing and monitoring controls by risk.
5. Participants will be able to diagnose weaknesses through red teaming findings.
6. Participants will be able to justify an incident response playbook for agent failures to executives.

Programme Structure

Day 1: AI Agent Security Foundations

What can go wrong when an agent holds tools, data and credentials?

- Agent Architecture Map
- Asset and Entry Point Inventory
- Threat Model Template
- Prompt Injection Case Review
- Data Leakage Pathway Diagram

Day 2: Identity and Access Control

How are agents identified, and how are their permissions limited?

- Agent Identity Register
- Least-Privilege Permission Matrix
- Credential Handling Rules
- Tool Allow-List Design
- Approval Gate Criteria



Day 3: Sandboxing and Monitoring

Which layers contain an agent, and which signals show misuse?

- Execution Sandbox Layout
- Output Filtering Rules
- Action Logging Standard
- Anomaly Alert Catalogue
- Kill Switch Procedure

Day 4: Red Teaming and Incident Response

How are weaknesses found before attackers do, and how is a failure contained?

- Adversarial Test Plan
- Attack Scenario Library
- Finding Severity Scale
- Agent Incident Response Playbook
- Post-Incident Review Form

Day 5: Applying the Tools to One Agent

Do the threat, access and response tools protect one agent end to end?

- Applying the Session 1 Threat Model Template
- Using the Session 2 Least-Privilege Permission Matrix
- Running the Session 3 Anomaly Alert Catalogue
- Checking the Session 4 Agent Incident Response Playbook
- Completing the Agent Control Plan

Methodology

The programme uses structured exercises built on anonymised agent designs, permission lists, log extracts and incident reports from banking, energy and public settings. Participants work in security teams on a shared agent case, challenge each other's controls and receive written facilitator feedback on every exercise set. Each session ends with a short decision on which control to adopt and why.

- Suggested activity: Tracing how an injected instruction could reach a payment tool.
- Suggested activity: Cutting an over-broad permission list to least privilege.
- Suggested activity: Walking a mock incident from first alert to containment.



Capstone Output

Each participant completes a control plan for one AI agent. It holds a threat model, a permission matrix, sandboxing and monitoring rules, a red teaming test plan and an incident response playbook. The plan is used at work to brief the security committee, guide engineers on required controls and set the review cycle for the agent.

Frequently Asked Questions

Is AI agent security suitable for leaders without hands-on engineering skills?

Yes. AI agent security is taught at managerial level using maps, matrices and case material, so coding is not required, though familiarity with basic security concepts is assumed.

Which models does AI agent security rest on?

AI agent security rests on the OWASP Top 10 for LLM Applications, least privilege, defence in depth and threat modelling, each applied in a named session.

How is the plan from AI agent security used at work?

You use the control plan to brief the security committee, to tell engineers which controls an agent needs and to set its review cycle.

How does AI agent security differ from AI governance?

AI agent security concentrates on technical controls, testing and response for agents, while AI governance concerns policy, accountability and oversight of AI use across an organisation.