JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

AI Security Champion Training Course for IT and Risk Staff

jupiterprime.uk

info@jupiterprime.uk

Registered in the United Kingdom and Saudi Arabia

AI Security Champion Training Course for IT and Risk Staff

AI Security Champion training is a 1-week training course for security analysts, IT engineers and risk officers who take on the champion role for generative and predictive systems in their organisation, and who leave with a working AI security playbook. Attackers now target the models, prompts and training sets behind business tools, while staff adopt unapproved chatbots faster than policy can follow. AI Security Champion training is delivered by Jupiter Prime Training.

Why AI Security Champion Skills Matter Now

Three pressures shape the AI security champion role today. Language models connected to email, files and plug-ins open new attack paths such as prompt injection and data exfiltration through generated output. Employees paste confidential material into unsanctioned services, creating shadow AI that security teams cannot see. Boards and auditors now ask who is accountable for model-related incidents. An organisation with trained champions spots poisoned inputs and leaking integrations early; one without them learns of exposure from customers or the press. The course rests on ISO/IEC 42001, the AI management system standard, and the NIST AI Risk Management Framework AI RMF.

Who Gains Most from AI Security Champion Skills

- Security analysts who monitor alerts and must recognise model-specific attack patterns
- IT engineers who integrate language models with mailboxes, document stores and plug-ins
- Risk officers who maintain the register of AI use cases and their exposure ratings
- Information security awareness leads who brief staff on safe use of generative services
- Application owners who approve or reject new AI features in business systems

A security function that sends an analyst, an engineer and a risk officer together gains most, because the three can build one shared threat picture and escalation path.

AI Security Champion Capabilities

Participants leave able to name, test and contain the attack paths that are specific to machine learning and language model deployments.

- Assess AI use cases with a threat model covering inputs, model, outputs and integrations — the application owner approving release
- Test chat interfaces for direct and indirect prompt injection — the engineer hardening the integration
- Specify data leakage controls for prompts, retrieval stores and generated output — the data protection officer
- Weigh poisoning and supply-chain risks in third-party models and datasets — the procurement decision
- Design a shadow AI discovery and sanctioned-service process — the chief information security officer
- Communicate safe-use rules through targeted awareness sessions — staff in every business unit

Key AI Security Champion Terms

- Prompt injection — crafted text that overrides a model's intended instructions, either typed directly or hidden in documents and web pages the model reads
- Data poisoning — deliberate corruption of training or retrieval data so a model learns or repeats attacker-chosen behaviour
- Shadow AI — use of AI services by staff without approval or visibility from IT and security functions
- Model exfiltration — extraction of a model's parameters, system instructions or confidential context through repeated querying
- Guardrail — a technical filter or policy check placed before or after a model to block unsafe inputs or outputs
- Red teaming — structured adversarial testing of an AI system to find exploitable weaknesses before attackers do

AI Security Champion Day by Day

Day 1: The AI Attack Surface and the Champion Role

- AI System Component Map — Inputs, Models, Outputs and Connected Plug-Ins
- Champion Role Charter — Mandate, Escalation Lines and Reporting Rhythm
- Threat Landscape for Language Models — Prioritising Attacks Seen in Practice
- NIST AI RMF Govern and Map Functions — Assigning Accountability for Model Risk
- ISO/IEC 42001 Control Objectives — Linking Security Duties to Management System

Day 2: Prompt and Output Attacks

- Direct Prompt Injection Techniques — Testing Chat Interfaces Before Release
- Indirect Injection Through Documents and Web Content — Retrieval Pipeline Hardening
- Jailbreak Patterns and Guardrail Bypass — Choosing Filter Layers
- Insecure Output Handling — Preventing Generated Code and Links Reaching Systems
- Excessive Agency in Plug-Ins — Limiting What Connected Actions Can Do

Day 3: Data, Model and Supply-Chain Threats

- Training and Retrieval Data Poisoning — Integrity Checks for Ingested Sources
- Sensitive Information Disclosure — Redaction and Access Control for Context Stores
- Model Exfiltration and System Instruction Leakage — Rate Limits and Monitoring
- Third-Party Model and Dataset Vetting — Supplier Security Questionnaires
- AI Threat Modelling Worksheet — Rating Likelihood and Impact per Use Case

Day 4: Shadow AI, Monitoring and Response

- Shadow AI Discovery Through Network and Expense Signals — Sanctioned Service Lists
- Logging and Telemetry for Model Interactions — Detection Rules for Misuse
- AI Incident Response Runbook — Containment, Rollback and Notification Steps
- Red Teaming Exercise Design — Scope, Rules of Engagement and Findings
- Awareness Session Planning — Role-Based Messages for Business Units

Day 5: Practising the AI Security Playbook

- Threat Model Walkthrough — Banking Chatbot and Hospital Records Assistant Cases
- Injection and Leakage Test Rehearsal — Applying Day Two and Three Methods
- Incident Tabletop Simulation — Exercising the Response Runbook
- Control Mapping to ISO/IEC 42001 and NIST AI RMF — Gap Prioritisation
- AI Security Playbook — Threat Models, Controls, Runbooks and Awareness Calendar

Applying AI Security Champion Skills After the Course

- Run a threat modelling session on the organisation's most exposed AI use case and record ratings and controls in the risk register
- Launch a shadow AI discovery sweep and publish a sanctioned service list with a short awareness briefing for staff

AI Security Champion Questions FAQ

What is an AI Security Champion and why does the role matter now?

An AI Security Champion is the person who spots, tests and contains threats specific to AI systems inside an organisation. The role matters now because language models connect to sensitive data and staff adopt unapproved tools faster than policies change.

Which roles gain most from AI Security Champion training?

Security analysts, IT engineers, risk officers and awareness leads gain most from AI Security Champion training. Cases from banking and healthcare show how prompt injection, data leakage and poisoning affect customer service, records and decision support systems.

How does AI Security Champion training differ from a general cyber security course?

AI Security Champion training focuses on attacks that exist only because of models, prompts and training data. A general cyber security course covers networks, endpoints and identity; here you test injection, poisoning, leakage and shadow AI directly.

What do you bring back from AI Security Champion training?

You bring back a playbook tailored to your organisation's AI use cases. It holds threat models, a control map, an incident runbook and an awareness calendar, ready to present to your security leadership.

AI Security Champion: The Way Forward

The AI security playbook gives each champion threat models, mapped controls, a response runbook and an awareness calendar. Its first decision is which AI use case receives controls and testing before the next release, and security leaders use it to assign ownership. After the first incident or red team exercise, revisit the threat ratings, as attackers adapt quickly to new defences.