AI500 AI Security Professional

Kurzbeschreibung:

Teilnehmer lernen, KI-Anwendungen mit offensiven und defensiven Methoden systematisch zu prüfen und abzusichern. Sie führen Red Teaming gegen LLMs, Agenten und RAG-Systeme durch, inklusive Prompt Injection, Jailbreaks und Adversarial Attacks mit Einordnung über MITRE ATLAS. Auf der Defensivseite setzen sie OWASP Top 10 für LLMs um, etablieren Kontrollen über den gesamten KI-Lebenszyklus sowie Logging, Monitoring, Incident Response, Threat Modelling und Trustworthy AI.

Zielgruppe:

- **IT- und Cybersecurity-Fachkräfte**
  - Security Engineers, Analysten und Penetration Tester
  - Verantwortliche für IT- und Applikationssicherheit
- **KI- und ML-Verantwortliche**
  - Data Scientists, ML Engineers, AI Architects
  - Verantwortliche für KI-Entwicklung und Betrieb
- **Führungskräfte im Bereich IT, Security & AI**
  - Chief Information Security Officers (CISO)
  - Chief Data / AI Officers
- **Entwickler und DevOps / MLOps Teams**
  - Entwickler, die KI-Systeme implementieren und betreiben
  - Teams für DevSecOps / MLSecOps

Voraussetzungen:

Vorheriger Besuch des AI300 AI Technology Implementer oder vergleichbare Vorkenntnisse.

Sonstiges:

Dauer: 5 Tage

Preis: 3650 Euro plus Mwst.

Ziele:

- **Grundverständnis für KI und ML entwickeln**
  - Kenntnisse über Funktionsweise, Modelle und typische Einsatzszenarien
  - Verständnis für Chancen und Risiken von KI-Systemen
- **Offensive Security-Kompetenzen aufbauen**
  - Angriffsflächen von KI-Anwendungen, LLMs, Agenten und RAG-Systemen erkennen
  - Praktische Angriffsstrategien verstehen (Red Teaming, Adversarial Attacks, Evasion)
  - MITRE ATLAS zur Klassifikation von Angriffen anwenden

- **Defensive Security-Fähigkeiten entwickeln**
  - ♦ Sicherheitskontrollen über den gesamten KI-Lebenszyklus implementieren
  - ♦ OWASP Top 10 LLM & Agentic AI verstehen und anwenden
  - ♦ Logging, Monitoring und Incident Response für KI-Systeme aufbauen

- **Trustworthy AI und organisatorische Maßnahmen etablieren**
  - ♦ Prinzipien vertrauenswürdiger KI in Organisationen verankern
  - ♦ Governance, Richtlinien und Prozesse zur KI-Sicherheit implementieren

- **Praxisorientiertes Threat Modelling anwenden**
  - ♦ Bedrohungen für KI-Systeme analysieren und priorisieren
  - ♦ Defensive Maßnahmen systematisch ableiten

Inhalte/Agenda:

- **Grundlagen & Rahmenbedingungen**
  - ◊ KI & ML Grundlagen
  - ◊ AI Act und regulatorische Anforderungen
  - ◊ Risk Management für KI-Systeme

- **Offensive Security für KI**
  - ◊ Einführung in AI Red Teaming
  - ◊ Angriffe auf KI-Modelle, Daten & Pipelines
  - ◊ Angriffe auf Vektordatenbanken und RAG-Systeme
  - ◊ LLM-spezifische Angriffstechniken (Prompt Injection, Jailbreaks etc.)
  - ◊ Angriffe auf KI-Agenten
  - ◊ Evasion-/Adversarial-Angriffe
  - ◊ MITRE ATLAS (offensive Perspektive)

- **Defensive KI-Sicherheit (AI Application Security)**
  - ◊ Grundlagen der KI-spezifischen Anwendungssicherheit
  - ◊ OWASP Top 10 für LLM und Agentic AI / MLSecOps
  - ◊ Security Controls über den gesamten KI-Lebenszyklus
  - ◊ Organisatorische & Governance-Maßnahmen
  - ◊ Trustworthy AI

- **Sicherheitsmethoden & Betrieb**
  - ◊ Threat Modelling für KI-Systeme
  - ◊ Logging & Monitoring
  - ◊ Incident Response für KI

- **Zertifikatsprüfung**