Agentic AI Security Training

Duration: **3 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Overview

With the development of Agentic AI systems built on the foundation of**Large Language Models (LLMs)**, the actions generated by agents can evolve in unpredictable ways from a security perspective. In addition to text, audio, video, and images generated by LLMs, LLM Apps have their own unique security vulnerabilities.

Unlike LLMs, **AI Agents** can also introduce security risks at the action level. These risks may include the poisoning of an agent, changing an agent's goal, manipulating the tools used by an agent, or causing an agent to perform malicious actions.

In this training, we focus on LLM and Agentic AI vulnerabilities through hands-on practices. Using OWASP, which has become a standard reference in software and mobile security, as a reference for LLM and Agentic AI security, we examine vulnerabilities covered by the **OWASP Machine Learning Security Top 10, OWASP LLM Top 10, and OWASP Agentic AI Security Top 10** in depth.

Who Should Attend

This training is intended for participants who want to learn about**AI Security, LLM Security, and Agentic AI Security**, examine security vulnerabilities in artificial intelligence systems, or work on securing these systems.

It is particularly suitable for technical participants working in software, artificial intelligence, and cybersecurity who work with AI systems, LLM Apps, and AI Agents; are interested in AI Blue Teaming and AI Red Teaming; or want to examine **OWASP Machine Learning Security, OWASP LLM, and OWASP Agentic AI Security** vulnerabilities.

Agentic AI Security Training

Duration: **3 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

In this training, you will examine what AI Security is, the concepts of artificial intelligence security and the use of AI in security, and the different types of AI systems used in R&D and production environments.

You will explore what it means for an artificial intelligence system to be secure through the concepts of **AI Robustness and AI Security**. You will also examine the familiar cybersecurity concepts of Blue Teaming and Red Teaming from an AI perspective, including the responsibilities and roles of AI Blue Teams and AI Red Teams. In addition, you will examine the use of AI for malicious cyber activities and the overall AI Security landscape.

Within the **OWASP Machine Learning Security Top 10**, you will review the security of applied Machine Learning and Deep Learning projects.

Within the **OWASP LLM Top 10**, you will examine Prompt Injection, Sensitive Information Disclosure, Supply Chain, Data and Model Poisoning, Improper Output Handling, Excessive Agency, System Prompt Leakage, Vector and Embedding Weaknesses, Misinformation, and Unbounded Consumption.

Within the **OWASP Agentic AI Security Top 10**, you will examine Agent Goal Hijack, Tool Misuse and Exploitation, Identity and Privilege Abuse, Agentic Supply Chain Vulnerabilities, Unexpected Code Execution (RCE), Memory & Context Poisoning, Insecure Inter-Agent Communication, Cascading Failures, Human-Agent Trust Exploitation, and Rogue Agents.

Outline

AI Security

What is AI Security?

- What is artificial intelligence security?
- Artificial intelligence security and the use of AI in security.

Different Types of AI Systems

- Different types of AI systems in R&D and production environments.

What Does "Secure AI" Mean?

- What does it mean for artificial intelligence to be secure?
- AI Robustness and AI Security.

Blue Teaming vs. Red Teaming

- AI-focused Blue Teaming and Red Teaming.
- Responsibilities and roles of AI Blue Teams and AI Red Teams.

Agentic AI Security Training

Duration: **3 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Using AI for Malicious Intents

- The use of artificial intelligence for malicious cyber activities.

The AI Security Landscape

- The AI Security landscape and overall architecture.

OWASP Machine Learning Security Top 10

- Project reviews of applied Machine Learning and Deep Learning security.

OWASP LLM Top 10

LLM01: Prompt Injection

- Security risks created by prompts and manipulating an LLM through Prompt Injection.

LLM02: Sensitive Information Disclosure

- Security risks associated with the disclosure of critical or sensitive information learned by LLMs.

LLM03: Supply Chain

- Supply chain vulnerabilities in AI infrastructure and their security.

LLM04: Data and Model Poisoning

- Data and model poisoning vulnerabilities and protection against these risks.

LLM05: Improper Output Handling

- Security problems associated with LLM outputs and secure output handling.

LLM06: Excessive Agency

- Secure development and design processes for LLM applications.

LLM07: System Prompt Leakage

- Disclosure and protection of system prompts.

LLM08: Vector and Embedding Weaknesses

- Security vulnerabilities in Vector DB and Embedding processes.

LLM09: Misinformation

- The impact of incorrect information on LLM results.

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

LLM10: Unbounded Consumption

- Financial losses caused by excessive use of AI infrastructure and measures that can be taken against them.

OWASP Agentic AI Security Top 10

ASI01: Agent Goal Hijack

- Changing the predefined task or goal of an agent.

ASI02: Tool Misuse and Exploitation

- Manipulation of agent actions and the tools used by agents.

ASI03: Identity and Privilege Abuse

- Security of agent identities and authorization mechanisms.

ASI04: Agentic Supply Chain Vulnerabilities

- Examining supply chain vulnerabilities from an Agentic AI perspective.

ASI05: Unexpected Code Execution (RCE)

- Risks resulting from agents executing tools or generating malicious code.

ASI06: Memory & Context Poisoning

- Poisoning of LLM and Agent memory mechanisms.

ASI07: Insecure Inter-Agent Communication

- Manipulation or attacks against communication between agents through protocols such as A2A.

ASI08: Cascading Failures

- The cascading impact of one agent's output or decision on other agents in Multi-Agent systems.

ASI09: Human-Agent Trust Exploitation

- Malicious or poisoned agents attempting to manipulate human approval in Human-in-the-Loop processes.

ASI10: Rogue Agents

- Agents operating outside their intended design and performing malicious operations, requests, or redirections in the background.