LLM Security Training

Duration: **3 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Overview

LLM Security Training is designed for professionals who want to understand security threats affecting Large Language Models (LLMs), generative AI applications, and Retrieval-Augmented Generation (RAG) systems, and develop effective approaches to protecting these technologies.

The course begins with the fundamental operating principles of large language models and explores security risks such as prompt injection, jailbreak attacks, data leakage, and model manipulation. Participants learn to recognize potential vulnerabilities in LLM-powered applications, analyze attack scenarios, and evaluate appropriate defensive measures.

The program also covers RAG security, sensitive data protection, secure model integration, output filtering, and LLM security testing. Theoretical concepts are supported by practical laboratory exercises, helping participants develop the skills required to design, test, and manage more secure LLM-based systems.

Prerequisites

Introductory familiarity with artificial intelligence concepts and basic cybersecurity knowledge are recommended.

Who Should Attend

- Professionals responsible for improving security in LLM and generative AI projects
- AI Security Specialists
- Data Scientists
- IT Auditors
- Risk Management and Compliance Specialists
- AI Product Managers

LLM Security Training

Duration: **3 Day**

BİLGİNÇ
IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

By the end of this training, you will be able to:

- Understand how large language models operate and identify their main security risks.
- Analyze prompt injection, jailbreak, and model manipulation attacks.
- Evaluate data security and data poisoning risks in RAG systems.
- Recognize sensitive data exposure and model misuse threats.
- Implement secure output mechanisms using guardrails and output filtering.
- Apply LLM security testing and red teaming techniques.
- Use logging, monitoring, and incident response approaches.
- Evaluate the fundamental principles of secure LLM integration in enterprise applications.

Outline

1. LLM and Generative AI Fundamentals

- Introduction to Large Language Models and Generative AI
- LLM Architecture and Operating Principles
- Transformer Fundamentals
- Tokenization and Context Window Concepts
- Prompt Engineering Fundamentals

2. LLM Security Threats and Attack Techniques

- Understanding Prompt Injection Attacks
- Jailbreak Attacks and Attempts to Bypass Security Restrictions
- Role Manipulation Attacks
- Context Manipulation Techniques
- Model Manipulation and Security Risks in LLM-Based Systems

3. RAG and Data Security

- Security Risks in RAG-Based Systems
- Data Poisoning Threats

LLM Security Training

Duration: **3 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

- Vector Database Security
- Sensitive Data Exposure and Data Leakage Risks

4. Model Security and Protection Mechanisms

- Model Abuse and Misuse Scenarios
- Evaluating Hallucination-Related Risks
- Output Filtering Techniques
- Secure Output Control Using Guardrails
- Fundamental Principles of Secure LLM Integration

5. LLM Security Testing and Monitoring

- LLM Security Testing Approaches
- Red Teaming Processes
- Logging and Monitoring Techniques
- Incident Response Processes

6. Practical Security Exercises

- Exploring LLM Security Scenarios
- Practical Evaluation of Attack and Defense Approaches
- Analyzing Security Risks Through Hands-On Laboratory Exercises