LLM Security Eğitimi

Süre: **3 Gün**

BİLGİNÇ
IT Academy
info@bilginc.com
bilginc.com

Overview

LLM Güvenliği Eğitimi, büyük dil modelleri (Large Language Models - LLM), üretken yapay zeka uygulamaları ve Retrieval-Augmented Generation (RAG) tabanlı sistemlerde karşılaşılabilecek güvenlik tehditlerini anlamak ve bu tehditlere karşı etkili koruma yöntemleri geliştirmek isteyen profesyoneller için hazırlanmıştır.

Eğitimde, büyük dil modellerinin çalışma prensiplerinden başlayarak prompt injection, jailbreak, veri sızıntısı ve model manipülasyonu gibi güvenlik riskleri ele alınır. Katılımcılar, LLM tabanlı uygulamalarda oluşabilecek güvenlik açıklarını tanımayı, saldırı senaryolarını analiz etmeyi ve uygun savunma mekanizmalarını değerlendirmeyi öğrenir.

Program kapsamında RAG sistemlerinin güvenliği, hassas verilerin korunması, güvenli model entegrasyonu, çıktı filtreleme ve güvenlik testleri üzerinde durulur. Teorik konular uygulamalı laboratuvar çalışmalarıyla desteklenerek katılımcıların LLM tabanlı sistemleri daha güvenli tasarlama, test etme ve yönetme becerilerini geliştirmesi amaçlanır.

Prerequisites

Katılımcıların temel yapay zeka kavramlarına giriş seviyesinde aşina olmaları ve siber güvenlik konusunda temel bilgiye sahip olmaları önerilir.

Who Should Attend

- LLM ve üretken yapay zeka projelerinde güvenlik süreçlerini geliştirmek isteyen profesyoneller
- Yapay zeka güvenliği uzmanları
- Veri bilimciler
- Bilgi teknolojileri denetçileri
- Risk yönetimi ve uyum uzmanları
- Yapay zeka ürün yöneticileri

LLM Security Eğitimi

Süre: **3 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

Eğitim sonunda:

- Büyük dil modellerinin çalışma prensiplerini ve temel güvenlik risklerini anlayabileceksiniz.
- Prompt injection, jailbreak ve model manipülasyonu saldırılarını analiz edebileceksiniz.
- RAG sistemlerinde veri güvenliği ve data poisoning risklerini değerlendirebileceksiniz.
- Hassas veri sızıntılarını ve modelin kötüye kullanımına ilişkin tehditleri tanıyabileceksiniz.
- Guardrails ve output filtering yöntemleriyle güvenli çıktı mekanizmaları oluşturabileceksiniz.
- LLM security testing ve red teaming süreçlerini uygulayabileceksiniz.
- Logging, monitoring ve incident response yaklaşımlarını kullanabileceksiniz.
- Kurumsal uygulamalarda güvenli LLM entegrasyonu için gerekli temel prensipleri değerlendirebileceksiniz.

Outline

1. LLM ve Generative AI Temelleri

- Büyük dil modelleri (LLM) ve üretken yapay zeka kavramları
- LLM mimarisi ve temel çalışma prensipleri
- Transformer mimarisinin temelleri
- Tokenization ve context window kavramları
- Prompt engineering temelleri

2. LLM Güvenlik Tehditleri ve Saldırı Yöntemleri

- Prompt injection saldırılarının incelenmesi
- Jailbreak saldırıları ve güvenlik kısıtlamalarının aşılması
- Role manipulation saldırıları
- Context manipulation yöntemleri
- LLM tabanlı sistemlerde model manipülasyonu ve güvenlik riskleri

3. RAG ve Veri Güvenliği

- RAG tabanlı sistemlerde güvenlik riskleri

LLM Security Eğitimi

Süre: **3 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

- Data poisoning tehditleri
- Vector database güvenliği
- Sensitive data exposure ve veri sızıntısı riskleri

4. Model Güvenliği ve Koruma Mekanizmaları

- Model abuse ve misuse senaryoları
- Hallucination kaynaklı risklerin değerlendirilmesi
- Output filtering yöntemleri
- Guardrails ile güvenli çıktı kontrolü
- Güvenli LLM entegrasyonu için temel prensipler

5. LLM Güvenlik Testleri ve İzleme

- LLM security testing yaklaşımları
- Red teaming süreçleri
- Logging ve monitoring yöntemleri
- Incident response süreçleri

6. Uygulamalı Güvenlik Çalışmaları

- LLM güvenlik senaryolarının incelenmesi
- Saldırı ve savunma yaklaşımlarının uygulamalı değerlendirilmesi
- Laboratuvar çalışmalarıyla güvenlik risklerinin analiz edilmesi