CORSO AI RED TEAMING: TECNICHE DI ATTACCO SU SISTEMI AI

Durata: 1,5 giorni

PROGRAMMA

MODULO 1: FONDAMENTI DI RED TEAMING SU SISTEMI AI

- Superficie d'attacco dei sistemi AI a confronto con quella tradizionale
- Principi di sicurezza offensiva applicati all'AI
- Mappatura degli attacchi AI al ciclo di vita del red team e alle pratiche di cyber defense

MODULO 2: RICOGNIZIONE SU TARGET AI

- Identificazione di applicazioni AI, componenti di machine learning e infrastruttura dei modelli
- Tecniche di discovery di asset, dipendenze e servizi esposti
- Approcci stealth per evitare il rilevamento durante la fase di ricognizione

MODULO 3: ATTACCHI AGLI AGENTI AI

- Abuso delle istruzioni di prompt per manipolare il comportamento degli agenti
- Sfruttamento dei sistemi di memoria e delle integrazioni con strumenti esterni
- Tecniche di evasione per mantenere la persistenza e operare senza attivare i sistemi di difesa

MODULO 4: ATTACCHI A SISTEMI MULTI-AGENTE E PROTOCOLLI A2A

- Architettura e modelli di fiducia nei sistemi multi-agente
- Manipolazione dei messaggi e corruzione dei flussi di lavoro
- Tecniche di impersonazione di agenti e sfruttamento dei protocolli A2A

MODULO 5: ATTACCHI ALLE PIPELINE RAG

- Avvelenamento delle fonti di conoscenza utilizzate per il recupero delle informazioni
- Manipolazione del layer di recupero per alterare il contesto fornito al modello
- Controllo degli output tramite compromissione della pipeline RAG

MODULO 6: ATTACCHI AGLI EMBEDDING

- Ruolo degli embedding nei sistemi di machine learning
- Tecniche di inversione degli embedding per il recupero di informazioni sensibili
- Estrazione di dati dai modelli tramite attacchi mirati

MODULO 7: ATTACCHI AL MODEL CONTEXT PROTOCOL E ALLE SUPERFICI DI INTEGRAZIONE

- Architettura dei layer di orchestrazione e dei framework di integrazione AI
- Abuso del Model Context Protocol per eseguire azioni non autorizzate
- Tecniche di escalation dei privilegi tramite le superfici di integrazione degli strumenti

CONTATTI

info@nexsys.it

www.nexsys.it

CORSO AI RED TEAMING: TECNICHE DI ATTACCO SU SISTEMI AI

Durata: 1,5 giorni

MODULO 8: Supply chain attack su sistemi AI/ML

- Vettori di attacco nella supply chain AI: dataset, pesi dei modelli, adapter e dipendenze
- Tecniche per l'introduzione di artefatti malevoli prima del deployment
- Rilevamento e mitigazione degli attacchi alla supply chain

MODULO 9: VULNERABILITÀ NELL'INFRASTRUTTURA E NEL DEPLOYMENT AI

- Identificazione di vulnerabilità in piattaforme cloud dedicate all'AI
- Attacchi a server di modelli e workload containerizzati
- Configurazioni errate e vettori di accesso comuni negli ambienti di deployment

MODULO 10: THREAT MODELING PER AMBIENTI AI

- Identificazione di asset critici e confini di fiducia in ambienti AI complessi
- Mappatura dei percorsi di attacco e valutazione del rischio
- Integrazione del threat modeling con la gestione del rischio e le capacità di rilevamento

MODULO 11: ESERCITAZIONE FINALE: RED TEAM ENGAGEMENT COMPLETO

- Pianificazione e conduzione di un'esercitazione di red team su un ambiente enterprise AI simulato
- Applicazione integrata delle tecniche affrontate durante il corso
- Simulazione della compromissione di sistemi AI in produzione e analisi dei risultati

CONTATTI

info@nexsys.it

www.nexsys.it

CORSO AI RED TEAMING: TECNICHE DI ATTACCO SU SISTEMI AI

Durata: 1,5 giorni

OVERVIEW DEL CORSO

I sistemi di intelligenza artificiale stanno ridisegnando il perimetro della sicurezza aziendale. Agenti autonomi, pipeline di recupero delle informazioni, protocolli di orchestrazione e supply chain dei modelli introducono vettori d'attacco che le metodologie tradizionali di red team non sono attrezzate a coprire. Gli avversari lo sanno già — e stanno imparando a sfruttarli.

Questo corso trasferisce la mentalità offensiva del red team direttamente nel dominio AI. In tre mezze giornate di formazione tecnica intensiva, i partecipanti acquisiscono le competenze per identificare, analizzare e attaccare sistemi AI reali: dagli agenti singoli alle architetture multi-agente, dai layer di orchestrazione all'infrastruttura di deployment in ambienti cloud e containerizzati.

Il percorso non si ferma alle tecniche singole. L'ultimo modulo è un'esercitazione completa di red team contro un ambiente enterprise simulato, progettata per mettere alla prova la capacità di ragionare come un avversario in uno scenario realistico e complesso.

A CHI È RIVOLTO IL CORSO?

Il corso è rivolto a:

- Penetration tester e red teamer
- Sistemisti e architetti IT
- Security analyst e blue teamer
- IT Manager e CISO

COSA SAPRAI FARE ALLA FINE DEL CORSO?

Al termine del corso i partecipanti saranno in grado di:

- Condurre ricognizione su ambienti AI per identificare asset, dipendenze e servizi esposti, mappando la superficie d'attacco prima di agire
- Manipolare agenti AI e architetture multi-agente sfruttando istruzioni di prompt, sistemi di memoria, relazioni di fiducia tra agenti e integrazioni con strumenti esterni
- Compromettere pipeline RAG attraverso l'avvelenamento delle fonti di conoscenza e la manipolazione del livello di recupero, fino al controllo degli output del modello
- Estrarre dati sensibili dai modelli tramite attacchi sugli embedding e sfruttare il Model Context Protocol per scalare privilegi o eseguire azioni non autorizzate
- Individuare e sfruttare vulnerabilità nella supply chain AI e nell'infrastruttura di deployment, inclusi server di modelli, ambienti cloud e workload containerizzati
- Costruire modelli di minaccia per ambienti AI complessi e condurre un'esercitazione completa di red team contro uno scenario enterprise realistico

CONTATTI

info@nexsys.it

www.nexsys.it