Offensive AI Exploits and Security (LFWS320)

OEM: Linux Foundation • Duration: 1 Day (8 hrs) • Code: LFWS320

COURSE MODULES & TOPICS

Introduction & Setup

- LLM architectures (RAG, agents, multi-agent), OWASP® Top 10, attack surface mapping.

Direct Prompt Injection

- RAG attacks, semantic retrieval, data extraction, dual-LLM defenses.

Guard Bypass

- Encoding bypass, synonym attacks, multi-step extraction, guard hardening.

IP/Header Spoofing

- X-Forwarded-For, LLM auth delegation risks, context injection.

Agent Tool Abuse

- LangChain/LangGraph abuse, command execution, tool scoping.

Indirect Injection & SSRF

- Data vs instruction boundary, webhook SSRF, sanitization.

Multi-Modal Injection

- Vision model attacks, steganography, dual-vision guards.

Memory Poisoning

- Conversational memory abuse, escalation spoofing.

Schema Confusion

- Function-calling abuse, path traversal, tool ambiguity.

Hands-On Capstone Labs

- Customer Support AI
- Multi-Agent Poisoning
- Capstone & Wrap-Up