INFOSECTRAIN
Educate. Excel. Empower.

AI-powered Security Operations Bootcamp

How LLMs Are Exploited by Hackers

BATCH - 02

11th - 12th July 2026

10 AM - 2 PM (IST)

Ollama · python · Streamlit · TikToken · garak · PyRIT · Chroma · SBERT

8 CPEs

Hands-On

OWASP LLM Top 10

Agentic AI

Why Attend?

AI security is reshaping how modern enterprises build, deploy, and defend intelligent systems, creating both powerful capabilities and new attack surfaces. This bootcamp is designed to bridge that gap by combining AI fundamentals with real-world security practices. It focuses on how LLMs and agentic systems work, where they fail, and how they are exploited in production environments. Through structured, hands-on learning, it builds practical skills in AI threat analysis, red teaming, and secure system design for real-world security operations.

What sets this training apart:

AI Security Lifecycle Coverage

Understanding how LLMs work, fail, and are secured across the full AI system lifecycle

LLM & AI Attack Techniques

Exploring prompt injection, jailbreaks, poisoning attacks, and system prompt exploitation

Agentic AI Security

Securing and attacking multi-agent systems, memory poisoning, and tool misuse scenarios

Hands-On Red Team Labs

Executing real-world AI attack and defense scenarios using frameworks, models, and tooling

OWASP & MITRE Mapping

Mapping AI threats to OWASP LLM, Agentic AI, MCP Top 10, and MITRE ATLAS

AI Red Team Automation

Using tools like PyRIT and Garak for structured adversarial testing and evaluation workflows

Key Takeaways

- Earn 8 CPE credits
- Hands-on AI security lifecycle training
- Real-world LLM attack simulations
- OWASP LLM and Agentic AI mapping
- MITRE ATLAS threat scenario coverage
- End-to-end AI red teaming experience

Meet the Expert

Urvesh

6+ Years of Experience

DFIR, Threat Hunting & Intel | CHFI | eTHP | DCPLA | CTIA | ECIH | CND | CCSE

Urvesh brings 6+ years of experience in information security, specializing in SOC operations, SIEM/XDR implementation, detection engineering, threat hunting, and DFIR. He has worked extensively on building and managing Wazuh-based SOC environments, integrating threat intelligence, and automating incident response workflows.

His specializations include:

- SOC deployment and SIEM/XDR implementation (Wazuh)
- Detection engineering, custom rules, and alert tuning
- Threat hunting and intelligence-driven investigations
- SOAR-based incident response and automation
- DFIR and security monitoring across hybrid environments

Urvesh has trained 300+ professionals in security operations, helping them build practical skills in detection, response, and SOC engineering.

Bootcamp Agenda

Day 1

Module 1: The AI Security Landscape

Builds the foundational vocabulary of AI security - how LLMs actually work under the hood, where they break, and how to map AI threats to the frameworks SOC teams already use every day.

- How LLMs actually work: tokens, context windows, RLHF, temperature, sampling
- OWASP Top 10 for LLMs and MITRE ATLAS, mapped alongside MITRE ATT&CK
- The AI attack surface from training data to inference and output handling
- Responsible AI security research and disclosure

Module 2: AI-Powered Threat Intel

Build a production-grade threat intel workflow where AI accelerates analyst work without becoming the source of truth & paste a real threat report URL, get back an analyst-ready evidence pack.

- AI-augmented IOC extraction from public threat reports
- Dynamic MITRE ATT&CK mapping with hallucination validation against the live ATT&CK dataset
- ATT&CK Navigator layer + SOC hunting pack generation (SPL + KQL starters)
- Analyst-in-the-loop validation: final / review / rejected separation

Module 3: AI for GRC

Use AI to draft, challenge, and validate compliance artifacts across ISO 27001, NIST CSF 2.0, EU AI Act, NIST AI RMF, and ISO 42001 without losing audit defensibility.

- ISO 27001:2022 Statement of Applicability drafting at scale
- NIST CSF 2.0 maturity assessment and gap analysis
- AI Governance through EU AI Act + ISO 42001 + NIST AI RMF lens
- GRC validation gate to catch hallucinated framework references and unsupported claims

Module 4: Assisted Detection Engineering Using AI

Bring AI into every stage of detection engineering, from drafting Sigma rules to validating them against telemetry, and then turn detection engineering back on AI systems themselves.

- AI-assisted Sigma rule generation from attacker behaviour
- AI-powered log triage, incident timeline construction, and IR report drafting
- Detection validation against real Windows Security and Sysmon logs
- Sigma detection rules FOR AI agent telemetry - tool-call anomalies, exfiltration patterns, memory-write events

Day 2

Module 5: LLM Architecture

Understand the LLM internals that an attacker actually exploits, such as tokenisation, system prompt boundaries, sampling, and trust failures, because every attack starts here.

- Tokenisation deep-dive: token smuggling, homoglyphs, encoding bypass
- System prompt trust boundary failures
- Attack reliability across temperature and sampling: measuring ASR, not vibes
- Uncensored vs aligned open-weight models: behavioural comparison

Module 6: Prompt Injection & Jailbreaks

The most active LLM attack surface in 2026, direct, indirect, and multi-turn, all covered with a measured Attack Success Rate.

- Manual Crescendo + Auto Crescendo (attacker/target/judge model loop)
- Skeleton Key direct policy override
- Many-shot Jailbreak and Context Compliance Attack (CCA)
- EchoLeak reproduction - CVE-2025-32711 markdown image exfil chain end to-end
- PoisonedRAG - embedding-layer poisoning of a local FAISS / Chroma knowledge base
- Judge-model scoring, telemetry capture, and Attack Success Rate measurement

Module 7: Agentic Attacks & Red Teaming

Where the 2025–2026 attack frontier actually lives - multi-agent systems, MCP servers, memory poisoning, and automated red-team tooling, all mapped to OWASP Agentic Top 10 (ASI01–ASI10).

- Build a 3-agent system (Orchestrator + Email + File), break it with indirect injection, harden it with 3 defences
- Persistent memory poisoning across sessions, with SIEM-style detection
- Garak vulnerability scanning + aligned vs uncensored model comparison
- MCP Tool Poisoning & Rug Pulls - CVE-2025-54136 (MCPoison) and CVE 2025-54135 (CurXecute).
- PyRIT Automated Red Team - Microsoft's Crescendo and Tree-of-Attacks with Pruning orchestrators against your own targets

INFOSECTRAIN
Educate. Excel. Empower.

Contact us

sales@infosectrain.com

www.infosectrain.com