

Practical AI Security Bootcamp

Understand, Attack, Defend, and Monitor
AI Systems

8 CPEs

Trainer-Led Demos

AI Attack Scenarios

ML & LLM Security

(A Brand of Azpirantz Technologies LLP)



Why Attend?

AI is rapidly becoming part of business applications, security workflows, automation systems, and decision-making processes. But as AI adoption grows, so does the attack surface across data, models, prompts, APIs, agents, pipelines, and infrastructure.

This bootcamp helps participants understand how AI systems work, how attackers target ML, LLM, and agentic AI systems, and how defenders can apply practical controls across the AI security lifecycle. Through trainer-led demonstrations, real-world scenarios, and framework-based explanations, participants will see how AI attacks, threat modelling, defenses, monitoring, and incident response work in practice.

What Sets This Training Apart

AI Security Lifecycle

Learn fundamentals, attacks, defense, monitoring, and response

Trainer-Led Demonstrations

Practical walkthroughs across ML, LLM, and agentic systems

ML, LLM & Agent Coverage

Understand risks across modern AI systems

Framework-Based Approach

Map risks using MITRE ATLAS, STRIDE, and OWASP

Practitioner Tool Exposure

Explore tools for attacks, guardrails, monitoring, and defense

AI Security Career Edge

Earn 8 CPEs, build AI governance skills for GRC, audit, risk, and security roles

Key Takeaways

✓ Understand AI, ML, LLM, and agentic AI risks

✓ Identify attack surfaces across models, prompts, APIs, and data

✓ Explore AI attacks through trainer-led practical demonstrations

✓ Map AI threats using OWASP, MITRE ATLAS, and STRIDE

✓ Learn defenses for guardrails, data, models, and applications

✓ Earn 8 CPEs and build practical AI security readiness

Meet the Expert



Avnish

7+ Years of Experience

AI Security | AI Governance | Cloud Security | Data Security | Consultant & Trainer

Avnish is an Information Security and Cloud Security Consultant with 7+ years of experience in AI security, cloud security, data security, AI-assisted threat detection, and securing AI/ML pipelines. For this bootcamp, he brings practical expertise in AI attacks, threat modelling, defensive controls, monitoring, and governance through real-world examples and trainer-led demonstrations.

His specializations include:

- AI security, AI governance, and model risk assessment
- Securing AI/ML pipelines, models, and data flows
- AI-assisted threat detection and SOC automation
- AWS, Azure, ELK, Splunk, and cloud security monitoring
- Threat modelling, incident response, and vulnerability remediation
- NIST AI RMF, AI governance, and model risk validation

Bootcamp Agenda

🔖 Day 1

Module 1: AI Fundamentals & Introduction to AI Security

- Understanding AI: Components of an AI system, and Predictive vs Generative vs Agentic AI
- How Machines Learn: Machine Learning vs Deep Learning, Model Development Lifecycle
- Transformer Architecture and Understanding how LLMs work: tokens, embeddings, attention, context window, system vs user prompts
- Agentic AI Basics: planning, memory, tool use
- Understanding how AI Systems vary from Traditional Software
- The AI Attack Surface: prompt, data, model, API, pipeline, and infrastructure risk layers
- Limitations of AI: hallucination, bias, explainability gaps, overreliance

Lab: Map the architecture of a sample AI application and identify its assets, entry points, and data flows

🔖 Day 2

Module 2: Attacking ML & DL Systems

- OWASP ML Security Top 10 overview
- Poisoning attacks: data and model poisoning concepts
- Evasion attacks: crafting inputs to bypass ML-based detectors
- Privacy attacks: Extraction and inference
- Attack tooling: Adversarial Robustness Toolbox (ART)

Lab: Run a pre-built evasion attack against a sample network attack detector

Module 3: Attacking LLMs and Agentic Systems— 1 Hour

- OWASP LLM Top 10 and OWASP Agentic AI Top 10 overview
- Direct and indirect prompt injection; system prompt leakage
- Jailbreaking techniques
- Context-window overflow and vector store poisoning
- Tool misuse, excessive agency, and MCP-specific attack vectors
- Automated testing tools: Garak, PyRIT, Promptfoo

Lab: Run an automated vulnerability scan against a sample AI application and identify exploitable weaknesses

🔗 Day 3

Module 4: AI Threat Modelling

- Threat modelling fundamentals: threat actors, assets, abuse cases, entry points, data flows, existing controls
- Applying MITRE ATLAS and STRIDE to AI applications
- Trust boundaries in AI systems: data, model, pipeline, infrastructure, and (for LLM/agentic systems) tool and memory boundaries

Lab: Produce a trust-boundary/data-flow diagram and threat model for a sample AI application

Module 5: Defending AI Systems

- Defense-in-Depth for AI: the layered security model across data, model, application, and infrastructure
- Data Security: data classification, leakage patterns (prompt, output, log, vector DB), sanitization and de-identification techniques
- Data protection tooling: Microsoft Presidio (PII detection/masking), Gitleaks/TruffleHog (secrets detection)
- Model Security for ML/DL: adversarial training, artifact integrity, model registry protection
- Model registry tooling: MLflow (versioning, model cards)
- Model Security for LLMs & Agents: input/output guardrails, guardrail approaches (rule-based, classifier-based, LLM-as-judge)
- Guardrail tooling: LLM-Guard, LlamaGuard, GuardrailsAI
- AI Application Security: API authentication/authorization, rate limiting, and cost budgeting
- Gateway tooling: LiteLLM (rate limiting, authentication, request logging)

Lab: Apply input/output guardrails to a vulnerable LLM application and verify they block malicious prompts

🔮 Day 4: AI Monitoring and Incident Response

Module 6: AI Monitoring and Incident Response – 2 Hours

- AI security monitoring overview: prompt, API usage, model query, and tool/agent action monitoring
- Monitoring/observability tooling: Evidently AI (drift reports), Arize Phoenix, Langfuse/Langsmith (LLM tracing)
- Abuse detection: high-volume querying, repeated extraction, and repeated jailbreak attempts
- AI incident scenarios: prompt injection, sensitive data exposure, model extraction, poisoned dataset
- AI Supply Chain Security: open-source model and dataset risk, third-party API risk, model provenance verification
- Supply chain tooling: SBOM/AI-BOM concepts

Lab: Trace LLM calls and tool invocations using Arize Phoenix, and identify repeated jailbreak/extraction attempts from the trace data

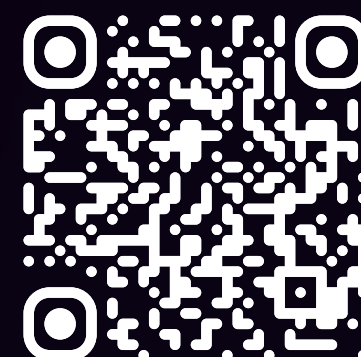


Get In Touch

 sales@infosectrain.com

 www.infosectrain.com

(A Brand of Azpirantz Technologies LLP)



Scan to Know More