

Tradecraft Academy

Adversary Intelligence: LLM Tradecraft

Agentic systems run inside your environments, orchestrating tools, holding memory, and carrying identities of their own. Adversaries reach them through prompt injection, jailbreaks, and weaknesses in the infrastructure agents depend on, and practitioners are asked to secure these systems before anyone explains how they work.

Adversary Intelligence: LLM Tradecraft closes that gap for defenders with self-contained training modules and hosted labs that teach LLM concept and security fundamentals, developed by SpecterOps in partnership with OpenAI.

A Closer Look at the Course

Participants will leave the course with the foundations to evaluate AI systems critically, apply AI capabilities to defensive security workflows, and assess and secure agentic systems in enterprise environments. The course begins with a study of machine learning applied to security problems such as phishing detection and malware classification. Then the course builds to LLMs, agents, and Codex workflows for security tasks including incident investigation, threat modeling, and reverse engineering.

One throughline runs across every module: the model is one part of an agent, and the harness around it determines what information the model receives, which tools it can reach, how its state persists, and how execution proceeds. That framing carries into security assessment, where an LLM application still has identities, APIs, dependencies, and data-access decisions, now shaped by what enters model context.

Knowing where these systems are vulnerable is what makes the defensive coverage meaningful. Participants learn the basics of how attackers approach these systems through prompt injection, jailbreaks, insecure MCP configurations, and overprivileged agent identities, building the understanding needed to then enforce defenses at the input, output, and infrastructure layers. Hosted labs run through every module, from building classic ML models to evaluating agent runs with MLflow and using Codex to reverse malware.

Course Modules

Foundations

- **AI Foundations**
Feature engineering, model families, training and evaluation
- **Large Language Models**
Tokenization, embeddings, attention, retrieval-augmented generation
- **Prompting**
Message roles, prompting techniques, task decomposition, structured outputs
- **AI Agents**
Harness architecture, tool calls, planning, memory, multi-agent design
- **Coding Agents and Codex**
Skills, subagents, sandboxing, approvals
- **Observability, Evaluation, and Optimization**
MLflow, deterministic verifiers, LLM-as-judge

(Continued on next page)

Why SpecterOps

SpecterOps assesses AI systems the way adversaries approach them, through AI red team engagements, AI security assessments, and cyber ranges built to test model performance against known completion paths. That work informs what this course teaches. Developed in partnership with OpenAI, the course carries the same commitment to fundamentals that shape all of our training courses.

Before You Register

Who Should Attend

Security practitioners, researchers, engineers, defenders, and technical leaders who need to understand how LLMs and agentic systems work, how adversaries target them, and how to secure AI-enabled workflows.

Prerequisites

Participants should be comfortable with technical security concepts and basic command-line workflows. Prior LLM security experience is not required; the course covers the underlying mechanics as it goes and keeps the math light. Familiarity with programming, security testing, threat modeling, or reverse engineering helps in the labs.

Hardware Requirements

Participants provide their own computer with a modern web browser. All course materials and labs are hosted, with no local virtual machines required. Labs use browser-based environments, command-line tooling, and guided AI-assisted workflows.

What's Included

Modules stand on their own, so participants can work the full progression or focus on the areas relevant to their work. New cohorts open monthly, each with 30 days of access to course materials and hosted labs, plus a ChatGPT Pro subscription from OpenAI. On completion, participants receive the course slides, a certificate of completion, and a digital badge confirming SpecterOps Training alumni status.

Enroll at academy.specterops.io



Course Modules

Security Applications

- **Model Context Protocol**
Tool-description injection, tool hijacking, SSRF
- **The LLM Threat Model**
Data flows, identities, trust boundaries
- **Jailbreaks and Prompt Injection**
Direct and indirect injection, prompt extraction, LLM safety and guardrails
- **Reverse Engineering with LLMs**
Ghidra workflows, function identification, type information, LLM integrations

Participants Will Learn How To

- ✓ **Build and evaluate LLMs and agentic systems,**
including how to assess their architecture, limitations, and security implications in real enterprise environments.
- ✓ **Identify and understand attacker techniques**
targeting LLMs and agentic systems, including prompt injection, jailbreaks, and exploitable weaknesses in AI infrastructure and agent design.
- ✓ **Apply AI to defensive security workflows,**
including threat modeling, security testing, and AI-assisted reverse engineering using leading AI technologies.