SEC537: **AI for Offensive Security™**

GOAA Offensive AI Analyst giac.org/goaa

GAIPT AI Penetration Tester giac.org/gaipt

6 Day Program | 36 CPEs | Laptop Required

You Will Be Able To

- Apply AI across the full attack lifecycle, from reconnaissance to post-exploitation
- Engineer AI tooling for reconnaissance, exploit development, and malware creation
- Automate vulnerability research through AI patch diffing, OSINT, and exploit generation
- Assess LLMs, RAG pipelines, Agents, and MCP servers using real-world adversarial TTPs
- Exploit AI-specific attack paths including prompt injection, model attacks, and agent abuse
- Map offensive and adversarial AI TTPs to defensive controls and architectural improvements

Business Takeaways

- Build complete AI capabilities within your offensive security program
- Reduce risk by evaluating AI systems with the same TTPs adversaries use in production
- Strengthen security by integrating AI offensive operations with adversarial AI testing
- Identify architectural weaknesses before AI-enabled applications become business risks
- Provide engineering teams with findings that improve AI security controls and system design
- Prepare security teams to assess both traditional attack surfaces and the AI systems

Artificial intelligence (AI) has changed offensive security in two important ways. Security practitioners now use AI throughout offensive operations to automate and enhance traditional attack workflows. At the same time, organizations are deploying AI-enabled applications that introduce entirely new attack surfaces requiring specialized testing.

Attack With AI and Attack AI Itself

This course pairs *SEC535: Offensive AI – Attack Tools and Techniques* with *SEC536: Adversarial AI – Penetration Testing AI Systems*, teaching you how to leverage AI in your operations and how to find and exploit vulnerabilities in AI systems others deploy.

If you'd rather focus on just one angle, SEC535 and SEC536 are also available individually.

Prepare for Both Sides of Modern Offensive AI

SEC537 brings these two different AI usages together into a single place. During the first three days, you will learn how to apply AI across offensive operations, from reconnaissance and social engineering to exploit development and malware creation — Offensive AI. The last three days of the course shift focus to Adversarial AI, where you will pentest LLM-integrated applications, RAG pipelines, AI agents, model infrastructure, and related AI components using the same attack techniques criminals use against live production systems.

Through realistic labs and hands-on exercises, you will build practical experience using AI as an offensive capability while learning how to identify and exploit weaknesses in modern AI deployments. Each attack is paired with defensive context, helping you understand not only how AI systems fail, but also the architectural decisions and controls that reduce real-world risk.

Whether your organization is adopting AI or your offensive team is already using it during engagements, SEC537 prepares you for both sides of modern Offensive AI and Adversarial AI in one complementary discipline.

GOAA Offensive AI Analyst giac.org/goaa

GIAC Offensive AI Analyst

The GIAC Offensive AI Analyst (GOAA) certification validates ability to apply practical, real-world offensive artificial intelligence techniques to modern cybersecurity challenges.

- SOC and incident response analysts
- Red team and penetration testers
- National security and law enforcement analysts
- Security consultants and risk managers
- Security engineers and architects
- Security managers
- AI and machine-learning enthusiasts

GAIPT AI Penetration Tester giac.org/gaipt

GIAC AI Penetration Tester

The GIAC AI Penetration Tester (GAIPT) certification validates a practitioner's ability to penetration test enterprise AI deployments, including LLM-based applications, RAG pipelines, and agentic systems.

- Penetration testers and red teamers
- Security practitioners
- Data scientists and AI engineers
- Technical leadership

Section Descriptions

SECTION 1: Introduction to Offensive AI and Vulnerability Exploitation

Discover how threat actors leverage AI to gather intelligence and exploit vulnerabilities. Students explore advanced OSINT tools enhanced by machine learning, examine RAG-powered penetration testing workflows, and exploit vulnerabilities using AI.

TOPICS:

- Introduction to Artificial Intelligence
- Introduction to AI Models and Capabilities
- OSINT AI For Penetration Testers
- Network
- Reconnaissance and Vulnerability Exploitation
- Web Exploitation Using AI

SECTION 2: Social Engineering Attacks

In this section, we'll explore the art of developing convincing phishing content, as well as how to supplement it with both audio and video deepfake content. We will then finish the day by performing patch diffing with AI.

TOPICS:

- Introduction to Social Engineering
- Creating AI-Powered Phishing Emails
- Creating Audio Deepfakes
- Creating Image and Video Deepfakes
- Patch Diffing

SECTION 3: Malware Development and Security Control Evasion

Explore cutting-edge AI applications in malware engineering and defense evasion. Students learn how neural networks transform malware development, grasp how attackers circumvent AI safety guardrails, demonstrate sophisticated evasion techniques powered by machine learning, and examine living-off-the-land tactics enhanced through computational intelligence.

TOPICS:

- Fundamentals of Malware
- Writing Malware with AI
- Hiding, Obscuring, and Trojanizing Persistent Malware
- Agentic Malware
- Bypassing Security Controls

SECTION 4: Foundations of Attack Techniques

In this section, we establish how AI models process input, then use that understanding to locate live deployments and subvert their instructions. Coverage moves from model internals through AI-specific reconnaissance into direct and indirect injection, memory and RAG poisoning, and multimodal attacks.

TOPICS:

- How LLMs process input: tokenization, embeddings, attention, and sampling
- Frameworks and case studies: OWASP LLM Top 10, MITRE ATLAS, EchoLeak
- AI-specific reconnaissance, stack mapping, and model fingerprinting
- Direct and indirect injection: prefill, role confusion, lethal trifecta, Memory and RAG poisoning, canary tokens, and multimodal evasion attacks

SECTION 5: Infrastructure, Integrations, and Advanced Attacks

In this section, we explore how the components surrounding AI models create their own attack surface. From safety training and guardrail evasion to infrastructure exposure, timing side channels, and API misconfiguration, this section focuses on exploiting the implementation layer: where and how AI meets the real world.

TOPICS:

- How safety training works and why jailbreaking succeeds
- Jailbreak techniques: personas, logic traps, framing, and policy puppetry
- Defense stack fingerprinting and classifier evasion techniques
- Exposed inference hosts, ML platforms, and observability dashboards, Timing side channels, API authorization flaws, and weight extraction

SECTION 6: Alignment, Interfaces, Agents, and MCP

Alignment is not a control; it is a probability adjustment, and the interfaces around the model are their own attack surface. This section covers the alignment failures you can test in a production deployment, sycophancy and constitutional AI abuse, the streaming and rendering layers that carry stored XSS and SSRF, and agentic systems where a single injection inherits every capability the agent holds.

TOPICS:

- Emergent behavior, bias, and misalignment as a testable attack surface
- Alignment attacks: reward hacking, goal misgeneralization, and sandbagging
- Sycophancy and constitutional AI: principle conflict, scope abuse, leakage
- Streaming response forgery, stored XSS in AI chat, SSRF, and sandbox escape
- Agentic attacks, multi-agent topologies, and MCP exploitation

Who Should Attend

SEC537 is designed for security practitioners who want to both use AI throughout offensive operations and assess the security of modern AI systems. This course is recommended for:

- **Penetration testers** looking to integrate Offensive AI into their methodology while learning to assess AI-enabled applications
- **Red team operators** who want to combine AI-assisted offensive tradecraft with adversarial testing of LLMs, RAG pipelines, AI agents, and related AI systems
- **Application security engineers and security architects** responsible for evaluating the security of AI features as they move into production
- **Security consultants** conducting offensive security assessments for organizations deploying AI-enabled applications
- **Security researchers** exploring Offensive AI techniques, adversarial AI testing, and emerging attack methodologies
- **Incident responders, SOC personnel, and cyber defenders** who need to understand how AI is changing attacker capabilities and expanding the enterprise attack surface
- **AI and machine learning engineers** who want to understand how deployed models, agents, and AI infrastructure appear from an attacker's perspective
- **Security managers, directors, and CISOs** building offensive security programs that must address both AI-assisted attacks and the security of enterprise AI deployments