Training: OffSec
# OffSec AI-300 Advanced AI Red Teaming

OffSec | Learning Partner

## TRAINING GOALS:



Advanced AI Red Teaming (AI-300) is OffSec's advanced AI cybersecurity training course designed to help security professionals assess and exploit vulnerabilities in modern AI systems. As organizations increasingly adopt generative AI, machine learning models, and autonomous AI applications, the attack surface for cyber threats is rapidly expanding, increasing the need for stronger threat intelligence, risk management, and modern cyber defense strategies. Traditional penetration testing approaches were not designed for AI-enabled environments, where models, data pipelines, agents, and orchestration frameworks introduce entirely new security risks. As organizations deploy generative AI across production environments, the AI attack surface continues to expand, requiring new approaches to AI security testing and offensive assessment.

AI-300 teaches learners how to apply an adversary mindset to modern artificial intelligence technology, combining proven cybersecurity methodology with offensive techniques tailored for AI applications, deep learning systems, and emerging AI technologies. The course focuses on how real attackers identify weaknesses in AI-enabled environments, manipulate model behavior, and compromise the infrastructure supporting modern AI deployments.

The training emphasizes hands-on AI cybersecurity labs that simulate real-world environments where AI systems operate alongside traditional infrastructure. Learners interact with enterprise-style AI architectures that include LLMs, vector databases, multi-agent systems, model orchestration frameworks, and cloud security environments supportingAI infrastructure, reflecting how modern AI technology is deployed in production environments.



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

Throughout the course, learners develop practical skills to identify vulnerabilities, simulate adversarial attacks, and analyze the impact of weaknesses across complex AI ecosystems. By applying offensive security techniques to generative AI platforms, machine learning pipelines, and AI deployment environments, learners gain the expertise required to evaluate AI-enabled systems from an attacker's perspective while supporting real-world security operations and incident response readiness.

The training culminates in the OffSec AI Red Teamer (OSAI) certification exam, a rigorous 24-hour practical red team engagement where learners must compromise a realistic AI-enabled enterprise environment. Successful candidates earn the OffSec AI Red Teamer (OSAI and OSAI+) certification, demonstrating practical expertise in assessing and exploiting modern AI systems.

***Each participant in an authorized OffSec AI-300 training held at Compendium CE receives a Learn One" license, which includes, among other benefits, a free OSAI/OSAI+ exam voucher.***

After completing this course, learners will be able to:

- Identify and map attack surfaces across modern AI systems, including generative AI, LLM applications, and machine learning environments
- Perform reconnaissance and threat detection and modeling for AI-enabled systems, identifying trust boundaries and high-value targets
- Exploit vulnerabilities in AI agents and multi-agent systems, including prompt injection and memory manipulation attacks
- Compromise RAG pipelines and vector databases through data poisoning and retrieval-layer manipulation
- Conduct embedding attacks and extract sensitive information from AI models and machine learning systems
- Exploit weaknesses in AI orchestration layers and tool integration frameworks used by modern AI applications
- Identify and exploit vulnerabilities across the AI supply chain, including datasets, models, and adapters
- Attack AI infrastructure and deployment environments, including model servers, cloud security platforms, and containerized workloads
- Perform model extraction, adversarial machine learning attacks, and AI system manipulation techniques
- Apply offensive methodology to assess AI cybersecurity risks and improve risk management strategies across AI environments



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

Who is this course for?

AI-300 is designed for experienced cybersecurity professionals looking to expand their expertise into AI security and machine learning security, including penetration testers, red teamers, security engineers, and professionals pursuing roles such as AI security specialist or certified AI security professional.

The course is also suitable for AI engineers and developers who want to better understand how adversaries target AI-enabled systems and learn practical techniques for identifying and mitigating AI cybersecurity risks.

This also includes professionals seeking an AI security certification or advancing their skills through an AI cybersecurity course focused on real-world adversarial techniques.

## CONSPECT:

- Introduction to Red Teaming AI Systems
  - Understand how artificial intelligence systems change the traditional attack surface. This module introduces the core concepts of AI cybersecurity, explains how adversaries target AI-enabled environments, and maps AI attacks to the red team lifecycle.
- Reconnaissance for AI Targets
  - Learn how to identify and map AI applications, machine learning components, and model infrastructure within a target environment. Students practice reconnaissance techniques used to discover AI assets, dependencies, and exposed services without alerting defenders.
- Attacking AI Agents
  - Explore offensive techniques for manipulating AI agents by abusing prompt instructions, memory systems, and tool integrations. This module demonstrates how attackers influence autonomous AI applications while maintaining stealth.
- Attacking Multi-Agent Systems and A2A Protocols
  - Analyze the architecture of multi-agent AI systems and learn how adversaries exploit trust relationships between agents. Students practice attacks such as message manipulation, agent impersonation, and workflow corruption.
- Exploiting RAG Pipelines
  - Examine how attackers compromise retrieval-augmented generation (RAG) systems by poisoning knowledge sources and manipulating retrieval layers to control model outputs.
- Attacking Embeddings



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70

BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

- Understand the role of embeddings in machine learning systems and perform attacks such as embedding inversion and information extraction to recover sensitive data from AI models.

- Attacking Model Context Protocol and Tool Surfaces
  - Explore how orchestration layers and AI tool integration frameworks can be abused to escalate privileges or execute unintended actions within AI systems.
- Supply Chain Attacks on AI/ML Systems
  - Learn how adversaries target the AI supply chain, including datasets, model weights, adapters, and dependencies. Students practice techniques used to introduce malicious artifacts into AI environments before deployment.
- AI Infrastructure and Deployment Exploits
  - Identify vulnerabilities in AI infrastructure, including cloud AI platforms, model servers, and containerized machine learning workloads.
- Threat Modeling for AI-Enabled Targets
  - Develop strategies for identifying high-value AI assets, trust boundaries, and potential attack paths in complex AI environments.
- Assembling The Pieces - Capstone Red Team Engagement
  - Apply the techniques learned throughout the course during a full-spectrum red team engagement against a realistic enterprise AI environment, simulating how adversaries compromise production AI systems.

## REQUIREMENTS:

AI-300 is an advanced AI cybersecurity course designed for learners with a strong foundation in cybersecurity. Students should have experience with penetration testing concepts, networking, Linux and Windows systems, and basic scripting.

A basic familiarity with AI systems or machine learning concepts, such as LLMs or generative AI applications, is helpful but not required. The course focuses on offensive security techniques for assessing AI-enabled systems, and learners without prior AI experience can still succeed.

OSCP or equivalent hands-on experience is recommended.

## Difficulty level



## CERTIFICATE:

The participants will obtain certificates signed by Compendium CE (course completion).

The AI-300 course and online lab prepares you for the OSAI/OSAI+ AI Red Teamer certification. Learn more about the OSAI/OSAI+ exam



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

https://help.offsec.com/hc/en-us/articles/46593096734612-OSAI-Exam-Guide

***Each participant in an authorized OffSec AI-300 training held at Compendium CE receives a Learn One" license, which includes, among other benefits, a free OSAI/OSAI+ exam voucher.***

## TRAINER:

Authorized OffSec Trainer

## ADDITIONAL INFORMATION:

The course includes a license "Learn One"

The license includes:

- One 200 or 300-level course
- 365 days of access
- 2 exam attempts
- Bonus KLCP and OSWP courses + exams
- 200+ Proving Grounds Practice labs



Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70

BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007