AI Hacking 101

Category: Networking & Security | **Vendor:** TCM Security

Duration: 8.00 hours (1 days) **6.5 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course Overview

The AI Hacking 101 ILT teaches students the fundamentals of penetration testing AI/LLM based applications such as customer facing chatbots. The course focuses on demonstrating how to detect and exploit common AI vulnerabilities such as: Prompt Injection Sensitive Information Disclosure Improper Output Handling System Prompt Leakage Misinformation Excessive Agency Not only will students learn about these exploits, but they will also spend hands-on time in a custom-built environment exploiting and uncovering these vulnerabilities. The online lab features the TCM Vulnerable Chatbot, a customer service chatbot that can interact with customers' tickets and improve its responses via Retrieval Augmented Generation (RAG) using the company's knowledge base.

About This Course

The AI Hacking 101 ILT teaches students the fundamentals of penetration testing AI/LLM based applications such as customer facing chatbots. The course focuses on demonstrating how to detect and exploit common AI vulnerabilities such as: Prompt Injection Sensitive Information Disclosure Improper Output Handling System Prompt Leakage Misinformation Excessive Agency Not only will students learn about these exploits, but they will also spend hands-on time in a custom-built environment exploiting and uncovering these vulnerabilities. The online lab features the TCM Vulnerable Chatbot, a customer service chatbot that can interact with customers' tickets and improve its responses via Retrieval Augmented Generation (RAG) using the company's knowledge base.

Who Should Attend

- Penetration testers looking to add AI/LLM pen testing to their tool kit
- Developers working with AI and LLM applications
- Defenders looking to understand AI risks and how they can impact their organizations
- Anyone interested in AI and its risks and dangers

Prerequisites & Entry Requirements

General Prerequisites:

A preliminary understanding of penetration testing methodology is suggested.

Detailed Course Outline

1 - AI Fundamentals Review

- A quick review of some of the fundamentals of AI such as how they operate and standard terms such as model parameters, temperature, top-p, inference, training, LLMs.

2 – AI Threat Model

- Discuss the threat actors, assets, adversary goals and attack surfaces for modern AI applications and the specific AI application used in the course

3 - Reconnaissance, Model Mapping and Baseline Behavior and Fingerprinting

- Demonstrate techniques for performing reconnaissance of AI applications with a specific focus on fingerprinting underlying AI models and their settings.

4 - Prompt Injection and Jailbreaking

- Demonstrate common techniques for prompt injection and jail breaking

5 - Prompt Injection Tools and Resources

- Show common tools and repositories of prompts used for prompt injection and jailbreaking

6 - Bypassing Common Protections

- Showcase how to bypass common protections for prompt injection such as input/output filtering

7 - Testing for harmful output/hate speech/misinformation/off-topic content and resource drainage

- Demonstrate tests for verifying the model responds correctly to requests for generating harmful or Off-topic content or attempts to waste resources.

8 - Data Exfiltration

- Demonstrate how retrieval augmented generation works and vulnerabilities associated with it such as leakage of confidential material and PII.

9 - RAG and Vector DB Attacks

- Demonstrate attacks the focus on the retrieval of documents and the ticket base, showcase vector poisoning attacks.

10 – Excessive Agency

- Demonstrate how excessive agency in applications can be exploited and tested for.

Frequently Asked Questions

Q: What delivery options are available for AI Hacking 101?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 1-day AI Hacking 101 course provides up to 6.5 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the AI Hacking 101 training?

The training takes place over 1 day(s), with each day lasting approximately 8.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for AI Hacking 101?

Yes, we provide corporate training, dedicated training, and closed classes for AI Hacking 101. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: Why choose Nexus Human for AI Hacking 101?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your AI Hacking 101 training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 889 8700 (Ireland) | +44 20 8142 8961 (UK)

© 2026 Nexus Human. All rights reserved. This brochure was generated on 19/10/2026.