Lab – Defending AI

Please join the following rooms on TryHackMe platform and complete all the tasks.

Learn how to detect and defend against adversarial attacks and use AI to supercharge investigations and enhance blue team operations.

This lab explores how adversaries exploit machine learning models using adversarial inputs, data poisoning, and evasion techniques that bypass traditional defenses. You will start by identifying these attacks and learning how they impact model integrity. Then, you will implement defensive strategies like adversarial training and input validation to harden your systems. Finally, you will leverage AI itself to assist in blue team operations by automating triage, detecting anomalies, and accelerating forensic investigations. By the end, you will know how to secure ML pipelines and weaponise AI for defensive advantage.

AI/ML Security Threats

Learn AI basics, key terms, and how it's used by both attackers and defenders.

https://tryhackme.com/room/aimlsecuritythreats

Detecting Adversarial Attacks

Learn how to identify and analyse adversarial attacks.

https://tryhackme.com/room/idadversarialattacks

Defending Adversarial Attacks

Learn defence mechanisms to harden machine learning models.

https://tryhackme.com/room/defadversarialattacks

AI Forensics

Explore AI DFIR and learn how it boosts your investigation capabilities.

https://tryhackme.com/room/aiforensics