Techonquer

CERTIFICATION PROGRAM

AI Security Expert

Become a job ready AI red teamer. Break and defend LLMs, RAG systems, and AI agents through real hands on attacks, then earn a certificate that proves you can do it.

100% HANDS ON | BEGINNER TO ADVANCED | VERIFIABLE CERTIFICATE | CAREER FOCUSED

LEVEL	FORMAT	FOCUS	MODULES
Basic to Advanced	Module Based	AI Red Teaming	09 Modules

+91 6367098233 | INFO@TECHONQUER.ORG

TECHONQUER.ORG/AI-SECURITY-TRAINING

01 / OVERVIEW

Attack it. Understand it. Defend it.

AI is everywhere, and so are its attackers. The Techonquer Certified AI Security Expert program turns you into a hands on AI red teamer who can break, test, and secure modern AI systems. Start from the fundamentals, master offensive attacks across ML, LLMs, RAG, agents, and multimodal AI, then learn to defend them like a pro. Every skill maps to industry frameworks such as MITRE ATLAS and the OWASP Top 10 for LLM Applications, so you finish job ready with proof to show for it.

09	140+	10+	2
MODULES	TOPICS COVERED	TOOLS & FRAMEWORKS	RED & BLUE TEAM

Who should enroll

- Security engineers and penetration testers moving into AI security
- ML and AI engineers who need to secure the systems they build
- Red teamers and bug bounty hunters expanding into AI targets
- Students and professionals building an AI security career
- Blue teamers and SOC analysts defending AI powered products

What you will be able to do

- Map AI attack surfaces and threat models with MITRE ATLAS and OWASP LLM Top 10
- Execute prompt injection, jailbreaks, and data exfiltration on LLM apps
- Attack ML models with poisoning, evasion, extraction, and inversion
- Red team RAG systems, AI agents, and multimodal models
- Automate testing with Garak, Promptfoo, and custom tooling
- Build guardrails, monitoring, and AI incident response
- Write professional AI red team reports for stakeholders

02 / CURRICULUM

Course modules

M01 AI Foundations

Core concepts of artificial intelligence and modern generative AI systems.

- Introduction to AI, machine learning, and deep learning
- Supervised, unsupervised, and reinforcement learning
- Neural networks, weights, biases, and training
- Training vs inference
- Model inputs, outputs, and prediction pipelines
- Introduction to generative AI
- Large language model fundamentals
- Transformers, tokens, embeddings, context windows
- Prompting basics and model behavior
- LLM APIs, inference endpoints, and parameters

M02 AI Red Teaming Fundamentals

The foundations of AI red teaming and how attackers analyze AI systems.

- Introduction to AI security and AI red teaming
- AI safety vs AI security vs AI ethics
- AI threat landscape and real world attacks
- Understanding AI attack surfaces
- MITRE ATLAS framework overview
- OWASP Top 10 for LLM Applications
- Identifying AI components in web apps
- Mapping AI trust boundaries
- AI red team lab setup
- Overview of AI red teaming tools

M03 Reconnaissance & Attack Surface Discovery

Gathering information about AI systems before launching attacks.

- Identifying AI powered features in applications
- Fingerprinting LLMs and ML models
- Discovering hidden AI endpoints and APIs
- Extracting system prompts and hidden instructions
- Analyzing model behavior through prompt testing
- Mapping data sources used by AI systems
- Identifying inputs, tools, plugins, integrations
- Understanding AI permissions and access boundaries
- Preparing attack paths for AI exploitation

M04 Traditional Machine Learning Red Teaming

Offensive techniques used against classical machine learning models.

- Introduction to adversarial machine learning
- Data poisoning attacks
- Label flipping and training data manipulation
- Backdoor attacks in ML models
- Dataset contamination and integrity attacks
- Evasion attacks and adversarial examples
- White box vs black box attacks
- FGSM and PGD attack techniques
- Fooling image, text, and tabular models
- Model extraction and model stealing
- Query based model replication
- Surrogate model creation
- Membership inference attacks
- Model inversion attacks
- Training data and PII extraction
- ML supply chain and malicious pretrained models
- Pickle deserialization and model file exploitation

M05 Large Language Model Red Teaming

Attacking LLMs, chatbots, and prompt based AI applications. This is the core offensive module of the program.

- Introduction to LLM vulnerabilities
- Direct prompt injection
- Instruction override attacks
- System prompt manipulation
- Context hijacking
- Bypassing basic AI restrictions
- Advanced jailbreaking techniques
- Persona based jailbreaks
- Role playing exploits
- Encoding bypasses with Base64, ROT13, hex
- Payload splitting and multi step jailbreaks
- Multilingual prompt attacks
- Indirect prompt injection
- Hidden instructions in documents
- Poisoned web content attacks
- Email and file based prompt injection
- System prompt leakage
- Training data memorization attacks
- Sensitive data extraction from LLMs
- Exfiltration via markdown, links, tool output
- LLM red teaming exercises and challenges

KEY FOCUS AREAS

PROMPT INJECTION | JAILBREAKS | INDIRECT INJECTION | SYSTEM PROMPT LEAKAGE | DATA EXFILTRATION

M06 Advanced Generative AI System Red Teaming

Attacks against modern AI architectures such as RAG systems, AI agents, and multimodal AI.

- Retrieval augmented generation architecture
- Red teaming RAG applications
- Document poisoning
- Vector database manipulation
- Retrieval hijacking
- Access control bypass in RAG systems
- Cross tenant data leakage in AI apps
- Red teaming AI agents
- Tool use exploitation
- Function calling abuse
- Command injection through AI agents
- SSRF attacks via AI tools
- Remote code execution through agent capabilities
- Resource exhaustion and infinite loop attacks
- Multimodal AI red teaming
- Vision language model attacks
- Hidden text inside images
- Adversarial image prompts
- Audio based injection attacks
- Fine tuning and alignment exploitation
- Safety alignment bypass and reward hacking
- LoRA adapter security risks
- Session and memory abuse in AI apps

KEY FOCUS AREAS

RAG SYSTEMS | AI AGENTS | TOOL USE ABUSE | MULTIMODAL | ALIGNMENT BYPASS

M07 AI Red Team Automation & Operations

Scaling AI red teaming and running structured offensive security assessments.

- Automated AI vulnerability scanning
- Prompt fuzzing techniques
- Automated jailbreak testing
- Using Garak for LLM security testing
- Using Promptfoo for prompt evaluation
- Custom Python scripts for AI attacks
- Chaining AI with web vulnerabilities
- XSS through AI outputs
- SQL injection via AI generated queries
- SSRF through AI connected tools
- MLOps and AI infrastructure attacks
- ML pipeline exploitation
- Model registry attacks
- CI/CD poisoning
- Kubernetes and container risks for AI
- Offensive use of AI in cyberattacks
- AI powered phishing
- Deepfake based social engineering
- Voice cloning attacks
- AI assisted malware generation
- Scoping AI red team engagements
- Rules of engagement
- Legal and ethical considerations
- Threat actor emulation
- Multi stage AI attack simulation
- Exploit chaining and risk prioritization
- Evidence collection and proof of concept

KEY FOCUS AREAS

AUTOMATED TESTING | GARAK | PROMPTFOO | MLOPS ATTACKS | EXPLOIT CHAINING | ENGAGEMENT SCOPING

M08 AI Red Team Reporting & Risk Communication

Turning technical attacks into professional red team reports.

- AI red team reporting structure
- Executive summary creation
- Technical vulnerability documentation
- Attack scenario explanation
- Proof of concept presentation
- Severity scoring for AI vulnerabilities
- Business impact analysis
- Mapping findings to OWASP and MITRE ATLAS
- Remediation recommendations
- Presenting AI security risks to stakeholders

M09 Blue Team AI Defense & Monitoring

Defensive security measures to protect AI systems.

- Introduction to AI blue teaming
- Input validation for AI applications
- Output filtering and content controls
- LLM firewalls and guardrails
- Secure prompt design
- System prompt protection techniques
- Access control for AI systems
- Secure RAG architecture
- Secure AI agent design
- Adversarial training basics
- Safety alignment and model hardening
- AI logging and audit trails
- Detecting prompt injection attempts
- Monitoring abnormal model behavior
- Drift detection and anomaly detection
- AI incident response
- Building an AI security monitoring strategy

03 / TOOLING

Tools & frameworks covered

01 Python for AI security testing

02 Hugging Face

03 Local LLM deployment tools

04 Garak

05 Promptfoo

06 Adversarial Robustness Toolbox

07 MITRE ATLAS

08 OWASP Top 10 for LLM Apps

09 AI red team lab environments

10 LLM firewall & guardrail concepts

Techonquer Certified AI Security Expert

On completion, participants earn the Techonquer Certified AI Security Expert credential, validating hands on capability across the full AI attack and defense lifecycle.

- ✓ Full AI red teaming methodology
- ✓ Findings mapped to MITRE ATLAS and OWASP
- ✓ Professional reporting workflow
- ✓ Offensive skills across ML, LLM, RAG, agents
- ✓ Blue team defense and monitoring
- ✓ Verifiable certificate of completion

Launch your AI security career

Join the Techonquer Certified AI Security Expert program and gain the offensive and defensive skills companies are hiring for right now. Beginner friendly and 100% hands on.

WEB techonquer.org/ai-security-training

CALL +91 6367098233

MAIL info@techonquer.org

Enroll Now