OFFENSIVE SECURITY

AI Penetration Testing & AI Red Teaming

Test the Attack Surface Traditional Security Programs Were Never Built to Find

AI systems introduce risks across the model, data, pipeline, integrations, and human decision points. Traditional penetration testing was not designed to evaluate them.

01 — THE CHALLENGE

Organizations are deploying AI faster than many security and risk programs can adapt.

Prompt injection, training data poisoning, model extraction, jailbreaks, RAG leakage, and agentic system abuse create attack paths that are often absent from traditional testing scopes. Teams developing AI may lack offensive security expertise, while security teams may lack experience evaluating AI architectures.

02 — WHY ARCTIQ

Arctiq brings adversarial testing to the AI estate.

We test direct and indirect prompt injection, jailbreaking, RAG pipelines, agentic systems, training and fine-tuning workflows, and the integration layers surrounding commercial and self-hosted models.

> Testing extends beyond the model to system prompts, guardrails, tool definitions, authentication, authorization, data egress controls, and downstream trust relationships. Findings are aligned to OWASP Top 10 for LLM Applications, MITRE ATLAS, and the NIST AI Risk Management Framework.

03 — ADVISORY CAPABILITIES

- AI Direct and indirect prompt injection testing
- AI Jailbreak and system prompt leakage testing
- AI RAG pipeline security
- AI Agentic workflow and tool abuse testing
- Training data poisoning
- AI Model extraction and membership inference
- ✓ Insecure output handling
- Authentication and authorization testing
- Data egress and downstream trust analysis
- AI OpenAI, Anthropic, Azure OpenAI, Bedrock, and Vertex AI integrations
- AI Fine-tuned and self-hosted model testing

FRAMEWORKS & PLATFORMS

OWASP LLM Top 10 · MITRE ATLAS · NIST AI RMF · OpenAI · Anthropic · Bedrock · Vertex AI

04 — BUSINESS OUTCOMES

Organizations gain:

- ✓ Visibility into AI-specific attack paths
- ✓ Validation of model and application guardrails
- ✓ Greater understanding of AI data risk
- ✓ Prioritized remediation tied to operational impact
- ✓ Stronger alignment between AI security and governance
- ✓ A defensible view of the security posture of AI-enabled systems

05 — WHY IT MATTERS

AI systems create new attack paths that traditional testing may not identify. Arctiq helps organizations understand how AI-enabled applications and agentic systems perform against adversarial techniques designed specifically for AI environments.

Test your AI estate →

www.arctiq.com