CANARYTRAP

Catching Vulnerabilities and Trapping Exploits

Canary Trap's elite team of security experts come armed with the tools, experience and credentials to help improve your organization's security resiliency and cyber risk posture.

ARTIFICIAL INTELLIGENCE (AI) PENETRATION TESTING

Specialized testing to ensure AI trustworthiness

SERVICE OVERVIEW

Artificial Intelligence (AI) and Large Language Models (LLMs) are transforming how enterprises operate, innovate, and interact with data. As organizations build, train, and deploy AI driven systems, security teams face a rapidly evolving challenge: ensuring that these models and the systems they connect to remain secure, trustworthy, and resilient against adversarial manipulation.

AI and LLMs represent a fundamentally new class of software with unique attack surfaces. They can be influenced through crafted prompts, manipulated into leaking sensitive information, or exploited through insecure integrations and model driven logic flaws. These risks fall outside the scope of traditional penetration testing and require specialized adversarial techniques to properly evaluate and mitigate.

Canary Trap provides structured, model-aware penetration testing that's designed to uncover vulnerabilities across the full AI stack. We identify weaknesses in prompt handling, data retrieval, model behavior, system integrations, and operational safeguards - delivering the assurance needed to deploy AI responsibly and securely.

At the end of the engagement, Canary Trap will deliver a comprehensive Report of Findings. We will highlight identified vulnerabilities along with any associated opportunities to improve security and mitigate cyber risk.

Canary Trap combines human expertise with sophisticated tools, proven methodologies and, where appropriate, threat intelligence to ensure a thorough, in-depth approach to security testing and assessments.

Engage Canary Trap

Complete our Scoping Questionnaire at www.canarytrap.com or Contact Us directly by telephone or email.

Report of Findings

Canary Trap will deliver a Report of Findings highlighting any identified vulnerabilities for remediation.

The Canary Trap Approach

- ✓ **Step 1:** Define
- ✓ **Step 2:** Uncover
- ✓ **Step 3:** Report
- ✓ **Step 4:** Remediate
- ✓ **Step 5:** Retest

inquiries@canarytrap.com (844) 750-2018 www.canarytrap.com