Service Definition

AI Red Team Assessment

Introduction

Intertek's AI Red Teaming service simulates real-world adversarial attacks to identify vulnerabilities before malicious actors can exploit them. Using proprietary AI-powered testing tools combined with decades of cybersecurity expertise, Intertek helps organisations secure their AI implementations while aligning with ISO/IEC 42001 and the EU AI Act.

Scoping

AI systems are constantly evolving, and so are the threats targeting them. Intertek will work closely with the organisation to agree the most effective approach to simulate adversarial attacks to identify how AI systems can be misled, manipulated, or exploited to extract sensitive data or elicit malicious, harmful, criminal or unethical behaviour.

The collaborative scoping exercise will help to determine whether a full spectrum of Generative AI and Agentic AI security testing is required or if a modular or continuous AI assurance provides the best solution for resilient AI deployments. These conversations will typically include scope components consisting of:

- LLM Adversarial Payload Testing
- Web Application & API Testing
- Agentic AI Integration Testing
- Cloud Configuration Review
- Infrastructure Testing
- Threat Scenario Simulation

Service Overview

Adversarial AI Testing

Intertek conducts controlled Red Team exercises to uncover how AI can be manipulated.

LLM Adversarial Payload Testing

- Test if models withhold sensitive information
- Prevent malicious, harmful, or unethical responses
- Ensure compliance with safety and governance requirements

Agentic AI Integration Testing

- Evaluate autonomous or semi-autonomous agents in system interactions
- Identify potential exploitation paths or unsafe behaviours
- Verify reliability, security, and adherence to intended boundaries

Continuous AI Assurance

AI evolves constantly, and so do the threats. Intertek's subscription-based Continuous AI Assurance provides ongoing monitoring and testing to minimise risk and maintain resilience.

Options include:

- Regular Red Teaming to detect vulnerabilities as AI evolves
- Change-triggered assessments wx`hen models, data, or configurations are updated
- Remediation verification and retesting
- Threat evolution tracking for emerging attack methods

Service Definition

AI Red Team Assessment

This proactive approach ensures your AI remains secure, compliant, and resilient over time.

Deliverables
Throughout the duration of the exercise, Intertek will provide regular updates to appropriate staff with relevant findings, summary of activities to date, progress/completion status of phase tasks/activities together with further details of any significant vulnerabilities identified.

The importance of reporting quality should not be overlooked or underestimated, especially given the natural language and non-deterministic nature of AI/LLM interactions.

Our detailed reporting style clearly maps vulnerabilities identified to governance frameworks, such as the OWASP Top 10 for LLM Applications, the OWASP Top 10 (APP/API) and OWASP Application Security Verification Standard, accompanied by detailed remediation recommendations and ample visual evidence of vulnerability exploitation. This facilitates efficient replication and identification of causation for each reported vulnerability, significantly aiding effective client remediation activity.

The engagement will conclude with a final presentation of findings to senior management with anecdotal evidence of findings easily interpreted by a non-technical audience and detailing business and operational impact with clear suggestions for addressing problems found.

Service Management
You will be appointed a dedicated account manager and provided with points of contact for commercial and technical questions, to ensure that the service is properly executed.

Intertek NTA has a Service Level Agreement that is monitored by a R.A.G system to ensure that delivery standards are met on all projects. The SLA includes timeframes for project delivery with technical staff working to targets to ensure that deadlines are met.

Additionally, Intertek NTA has a customer services department that is responsible for the key administrative aspects of delivery such as liaising with you to confirm site details, scheduling the test, ensuring reports are sent out in a timely manner, scheduling follow up tests and debriefs etc. This is documented and monitored through the use of an internal database system, and links into other key processes regarding the correct method of sending the report according to the protective marking of the contents.

Information Assurance
Intertek NTA is security cleared to perform information security testing services on infrastructure, platforms or software containing data protectively marked up to OFFICIAL-SENSITIVE

Ordering and Invoicing Process
Once the scope of work and number of units/days required has been agreed between the customer and Intertek NTA, a valid purchase order is required to enable scheduling and delivery. An invoice will then be raised upon delivery of the service.

Service Definition

AI Red Team Assessment

Customer Responsibilities

The customer is responsible for engaging with Intertek NTA to discuss perceived threat vectors, actors/agents and agreeing targets of interest for the engagement.