CASE STUDY - Transforming AI-related risks into organisational resilience

Securing government innovation in the age of AI

Company
A UK Government department

Region
Europe

Intertek Solutions
Full-scale, multi-component AI Red Teaming engagement

As a leading cybersecurity consultancy, Intertek specialises in providing AI Red Teaming and Agentic Assurance for enterprise and government organisations.

We ensure that the transition from risk to resilience is seamless, aligning with global standards such as the EU AI Act, ISO 42001 and the OWASP Top 10 for LLM Applications.

Intertek's technical expertise transforms a high-risk innovation into a secure, resilient tool for public service.

"While AI integration benefits were clear, the stakes were exceptionally high. The application handles vast amounts of sensitive data, making it a prime target for malicious actors. Intertek's approach went beyond traditional penetration testing by focusing on the unique vulnerabilities of the AI/LLM ecosystem."

The challenge

The customer's application handles vast amounts of sensitive data, a primary target for malicious actors. Among the main concerns:

- Emerging Attack Vectors: The integration of AI introduced novel threats such as Prompt Injection, Jailbreaking, and PII Disclosure.
- Rapidly Evolving Landscape: As AI techniques advance, so do the methods used by cyber-criminals. Static security measures are no longer enough to ensure long-term safety.
- Operational Integrity: it was vital to ensure the AI did not exhibit "Excessive Agency" – acting beyond its intended boundaries in a way that could lead to unauthorised system interactions.

The solution

Key elements of Intertek's solution included:

- Advanced Adversarial Testing: Using our proprietary AI-powered tooling, we simulated real-world attacks to identify weaknesses in the model's guardrails.
- Environment Comprehensive Evaluation: From user-facing interfaces to backend API endpoints and cloud configurations.
- "LLM-as-a-Judge" Grading: To manage the vast amount of data, we utilised air-gapped, local infrastructure to grade AI outputs against strict security rubrics, ensuring 100% data sovereignty and zero exposure.
- Framework Alignment: Our testing was mapped to the OWASP Top 10 for LLM Applications, covering critical risks like System Prompt Leakage and Vector/Embedding Weaknesses.

The result

The AI Red Teaming exercise provided the client with independent technical assurance, allowing them to advance from development to production with confidence.

- Vulnerability Remediation: Our detailed reporting allowed the client to visualise exact failure points, and action fixes for vulnerabilities like prompt injection and sensitive data exfiltration.
- Strategic Resilience: By aligning the project with the EU AI Act, ISO 42001 and NIST frameworks, the department could assess posture against global regulatory compliance.
- Continuous Assurance: Through our Continuous AI Subscription, the department now benefits from ongoing monitoring and assessments as their models evolve over time.

For more information

nta-sales-dept@intertek.com

Contact form

intertek.com/ai