G DATA advanced analytics

AI Application Penetration Test

Finden Sie heraus, wie sicher Ihre KI-Anwendung ist

Gibt der Chatbot auf Ihrer Webseite auch bei manipulierten Eingaben wirklich keine Interna heraus? Wird über Ihre KI-API nur das ans CRM übertragen, was auch übertragen werden soll? Ein spezialisierter Penetration Test deckt Schwachstellen in Ihrer KI-Anwendung und deren Integration in Backend-Systeme, Schnittstellen und Datenflüsse auf.

Penetration Tests nach der OWASP LLM Application & Generative AI Security Top 10

Ideale Grundlage, um compliant mit Standards wie ISO/IEC 42001 oder dem EU AI Act zu werden

Erfahrene Spezialisten – mit der Expertise aus 40 Jahren IT-Sicherheit made in Germany

Welche Schwachstellen wir aufdecken

Prompt-Injection-Schwachstellen, einschließlich:

- Direct Prompt Injection
- Indirect Prompt Injection
- Jailbreaking
- Universal Jailbreaking

Supply-Chain-Risiken, wie z. B.:

- Verwendung von Drittanbieterpaketen mit bekannten Schwachstellen
- Einsatz veralteter oder unsicherer KI-Modelle

Die unsachgemäße Verarbeitung von Ausgaben, z. B. durch:

- Cross-Site Scripting (XSS)
- SQL Injection
- Remote Code Execution (RCE)
- XML External Entity (XXE)

Schwachstellen bei Schnittstellen, auf die die KI Zugriff hat, z. B. bei:

- APIs
- Datenquellen wie Fileserver oder SQL Server
- MCP Server

Die unbeabsichtigte Preisgabe sensibler Infos, etwa durch:

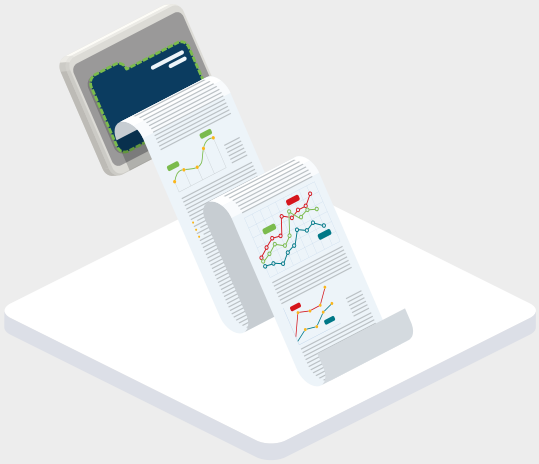
- Offenlegung des System-Prompts
- Preisgabe von Trainingsdaten, die für das Finetuning verwendet wurden
- Weitergabe von Nutzerdaten, die durch andere Benutzer eingebracht wurden
- Einsicht in sensible Informationen innerhalb der Anwendung

Welche Systeme wir für Sie analysieren

Mit einem AI Application Penetration Test kann nahezu jede produktiv eingesetzte LLM-gestützte oder generative KI-Anwendung überprüft werden – unabhängig davon, ob sie „off the shelf“ integriert, via API angebunden oder individuell entwickelt wurde. Entscheidend ist die Analyse des Zusammenspiels aus Modell, Logik, Schnittstellen und Benutzerinteraktion.

- Chatbots für Kundenservice oder HR
- KI-Assistenten für interne Prozesse (z. B. Ticketbearbeitung)
- Copiloten für Entwickler oder Analysten
- Textklassifikations- oder Zusammenfassungsdienste
- Text-to-Image- & Image-to-Text-Tools
- KI-gestützte Präsentations- oder Content-Generatoren
- MCP (Model Context Protocol) Server
- und viele mehr

Ihr Ergebnisbericht



Nach dem Penetration Test erhalten Sie einen umfassenden Abschlussbericht. Darin aufgelistet:

- eine Übersicht über alle gefundenen Schwachstellen, die von Angreifern ausgenutzt werden könnten
- die Einschätzung der Ausnutzbarkeit & der möglichen Auswirkungen
- Empfehlungen zur Behebung
- Management Summary

Zertifizierte Expertise

OSCP · eMAPT · eWPTX · CRTP · PNPT · OSEP

Erkennen Sie Schwachstellen in Ihrer KI-Anwendung, bevor andere sie ausnutzen.

sales@gdata-adan.de +49 (0) 234 / 9762-820

G DATA advanced analytics