solidlab

OFFENSIVE SECURITY

SECURITY ASSESSMENT OF AI-POWERED APPLICATIONS

KEY ASSESSMENT AREAS

- Discovering of attack surface – UI, APIs, and every agent in a multi-agent system, with LLMs, integrations, and security controls
- Exploitability assessment of AI-specific threats – prompt injection, agent manipulation, and more (covering the OWASP Top 10 for LLM Applications and beyond)
- Integration and access review – down to exactly what your AI agents are allowed to do across internal and third-party services
- Identifying abuse paths in GenAI features that break business logic or bypass built-in restrictions

THREATS TO AI APPLICATIONS

- **Prompt Injection**
  Crafted prompts let attackers hijack an AI agent – and access data or take actions on the system's behalf
- **Sensitive Information Disclosure**
  LLM responses can leak sensitive information: user data, internal documents, API keys, infrastructure details, and contents of RAG sources
- **Unbounded Consumption**
  Attack-driven token burn that inflates inference costs – the "denial-of-wallet" scenario
- **Excessive Agency**
  AI agents with excessive access to tools and services enable large-scale attacks – from pivoting into internal infrastructure through integrated tooling to running spam and phishing campaigns
- **Business Logic Bypass**
  The agent's context may miss subtle properties of the underlying business logic – and an attacker can exploit that to bypass restrictions without any specialized payloads at all

WHAT WE COVER

Architecture
Architecture and attack surface review

Integrations
Integrations and access permissions

AI-Specific Threats
AI-specific vulnerabilities and attack scenarios

Business Logic
Misuse of GenAI features

Security Controls
Review of implemented security features

Production Environment
Application hosting and delivery infrastructure

WHY AI SYSTEMS NEED A SPECIALIZED ASSESSMENT

Non-Deterministic Behavior

The same prompt can get a different answer every time: LLM output is non-deterministic

AI-Specific Threat Classes

From Prompt Injection and jailbreaks to data disclosure, integration flaws and over-privileged agents, AI-specific threats call for specialized testing methods

An Expanded Attack Surface

An AI agent becomes an autonomous execution unit with privileges of its own. Multi-agent systems compound this further, expanding the attack surface dramatically

WHO NEEDS AI SECURITY ASSESSMENT

- E-commerce
- Banking & Fintech
- Telecom & Industrial
- SaaS & Tech Companies
- Healthcare
- Government Agencies

WHAT YOU GET

- Technical report
- Executive summary
- Remediation guidance
- A retest to confirm your fixes are sound
- Expert advice on fixing vulnerabilities and defending against AI-specific threats

ASSESSMENT STAGES

Scoping

We agree on the scope, identify AI components and review your documentation and architecture

Architecture Analysis

Mapping the attack surface: LLMs, agents, integrations, APIs, user interfaces and data flows

Report & Recommendations

A detailed, actionable report

Hands-On Security Testing

Active vulnerability research guided by OWASP recommendations for AI systems, MITRE ATLAS, and Google SAIF; integration testing and validation of security control effectiveness

Exploitation & Risk Analysis

Demonstrating the real-world impact of discovered vulnerabilities, with severity ratings and business impact assessment

47A Vavilova St., Moscow 117312, Russia

+7 499 705 76 57

info@solidlab.ru

For more than 15 years, SolidLab Group has been protecting organizations from cyber threats, preventing attacks and minimizing the risks of financial and reputational damage. SolidLab provides a full range of services: from pentesting to application protection, incident response and secure SDLC implementation.