

Your AI can take action. Keep it within bounds.

AI penetration testing and red teaming for applications, RAG workflows, and agents.

TEST THE WORKFLOW. VERIFY THE BOUNDARIES.

An AI agent can follow the wrong instructions with legitimate credentials. Examine the path from untrusted content to data access and tool execution, including the application controls around the model.

01 / REVIEW

Architecture review

Examine permissions, data flows, trust boundaries, and approval controls before deeper testing.

02 / VALIDATE

AI penetration testing

Test an agreed application or workflow for exploitable weaknesses in AI behavior and supporting controls.

03 / CHALLENGE

Adversarial scenarios

Evaluate misuse across multiple steps, including tools, retrieved content, and persistent memory.

Specialist perspective.

Our team includes an OffSec AI Red Teamer (OSAI) certified professional. The testing scope is tailored to your workflow, permissions, and business risks.

Discuss your AI security testing scope

info@ironfistlabs.com | +1 (313) 306-2048

SCOPE AND OUTCOMES

From observed behavior to practical remediation.

Agree the boundaries, evaluate realistic scenarios, and review evidence with your team.

WHERE TESTING CAN FOCUS

AI APPLICATIONS	Prompt injection, sensitive information exposure, and behavior relevant to the agreed use case.
RAG & KNOWLEDGE	Document access, data boundaries, and isolation between users or customers.
AGENTS & TOOLS	Tool permissions, destinations, approval checks, and unauthorized action attempts.
MEMORY & STATE	Persistent influence, provenance, and separation between sessions or tenants.
SUPPORTING SYSTEMS	Authentication, API authorization, and conventional application weaknesses within scope.

What your team receives

A leadership summary, evidence-backed findings, practical remediation guidance, and a results walkthrough. Any retesting scope and window are defined in your proposal.

Results reflect the scenarios and configuration tested. Testing does not guarantee complete safety, and changes to models or integrations may require further evaluation.

Start with your workflow diagram, tool inventory, user roles, and key concerns.