ARTIFICIAL INTELLIGENCE SECURITY

AI Security Assessment

A hands-on security review of your LLM applications and AI pipelines — from prompt-layer risks to the permissions behind agentic systems — so you can ship AI features without shipping a new attack surface.

OWASP LLM
TOP-10 ALIGNED

Agentic
RED-TEAM READY

CycloneDX
AI-BOM OUTPUT

Senior
PRACTITIONER-LED

What we assess

LLM applications
System prompts, retrieval pipelines, and every path untrusted input takes to reach the model.

Agents & tool use
Tool integrations, autonomous workflows, and the permissions behind every action an agent can take.

Model supply chain
Third-party models, fine-tuning data, and the dependencies your AI stack quietly inherits.

What you get

- ✓ AI Bill of Materials (AI-BOM) in CycloneDX ML-BOM format
- ✓ Testing for prompt injection, jailbreaks, and unsafe output handling
- ✓ Data-leakage testing across retrieval pipelines and trust boundaries
- ✓ Red-teaming of agentic workflows, tools, and their permissions
- ✓ Review of model supply chain and third-party integration risk
- ✓ Deployment and guardrail guidance for your architecture
- ✓ Prioritized findings with buildable remediation steps

How it works

1. Scope your AI stack, data flows, and trust boundaries
2. Test the prompt layer, pipeline, and third-party integrations
3. Red-team agentic workflows and the permissions behind them
4. Deliver findings and guardrail guidance, then retest the fixes

A good fit if

- ✓ You have shipped an LLM feature and nobody has tested it adversarially.
- ✓ Your agents can call tools, move money, or reach customer data.
- ✓ Customers or auditors are asking how you secure your AI — or for an inventory of it.
- ✓ You want findings tied to real business impact, not a list of clever jailbreak prompts.

Probably not, if

- × You want certification that a model is “safe” — that is not something anyone can honestly issue.
- × You need accuracy, bias, or alignment evaluation — that is AI safety work, not security testing.
- × The feature is a prototype with no real data, users, or permissions behind it yet.

How we engage

Fixed scope
A fixed-scope proposal at a fixed price. No hourly meter running.

Senior-led
The practitioner who scopes the work does the work and writes the report.

Plain language
One report your engineers and your board can both act on.

Retest included
We verify the fixes landed — one round of retesting is in scope.