ADVISORY WORKSHOP

AI Agent Security Threat Modeling

Map the risks before they become incidents.

A structured engagement that identifies the specific security risks created when autonomous AI agents interact with your APIs, SaaS platforms, and internal systems. Your team leaves with a concrete threat model and a prioritized plan to address the risks that matter most.

DURATION	INVESTMENT	AUDIENCE
1–2 days	$15,000–$25,000	CISOs, security architects, identity teams, AI platform leads

WHY THIS MATTERS NOW

AI agents are being deployed faster than security teams can assess them. Traditional threat modeling was built for applications, not for autonomous actors that choose their own tools, delegate to other agents, and act at machine speed.

Focus Areas

- Privilege escalation through agent delegation chains
- Credential exposure and token misuse by autonomous agents
- Agent-to-agent manipulation and prompt injection vectors
- API misuse, hallucinated actions, and uncontrolled tool invocation
- Data exfiltration paths through ungoverned tool access
- Uncontrolled delegation and authority propagation

Deliverables

- ✓ Comprehensive threat model tailored to your agent architecture
- ✓ Risk severity summary with likelihood and impact assessment
- ✓ Recommended controls mapped to identified threats
- ✓ Prioritized remediation guidance your team can act on

IDEAL FOR

Organizations deploying AI agents that access customer data, execute transactions, call sensitive APIs, or operate in regulated environments. Best for teams that need to demonstrate due diligence to executives, auditors, or regulators before expanding agent usage.

1. **Pre-Workshop Discovery**
   We review your environment and current controls. A brief intake questionnaire ensures we arrive prepared.
2. **Live Working Sessions**
   Hands-on sessions with your security and technical teams. Real architecture, real decisions.
3. **Architecture Review**
   We synthesize findings into actionable artifacts: blueprints, gap analyses, and roadmaps.
4. **Final Delivery**
   A structured readout with your leadership team. Clear findings and a path forward.

Book a Workshop

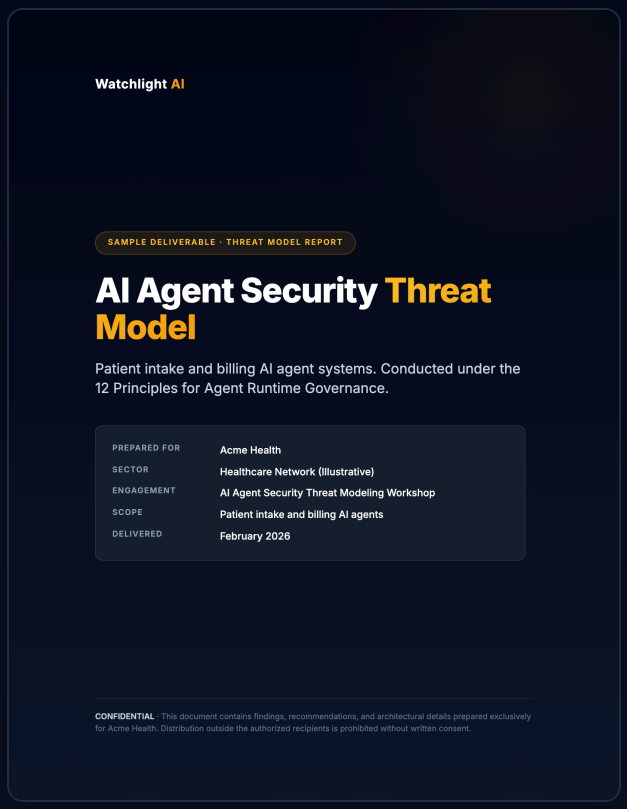
We typically respond within 2 business days.

watchlight.ai/workshops/book

SAMPLE DELIVERABLE

What You Actually Receive

Every workshop produces a structured, actionable deliverable. Below is a preview of the sample threat model report produced for a sample client, showing the depth and format of what your team will receive.



DOCUMENT	Sample Threat Model Report
SAMPLE CLIENT	Acme Health (Illustrative)
LENGTH	6 pages
FORMAT	PDF (with cover, sections, and appendices)

What the Deliverable Contains

Executive Summary
Overall risk posture, top priorities, engagement overview with risk counts.

Agent Inventory
Complete table of agents assessed, owners, systems accessed, and risk rating per agent.

Threat Findings
23 identified threats by severity with affected agents, attack scenarios, and impact analysis.

Recommended Controls
Controls mapped to threats and governance principles, sequenced by priority and effort.

Prioritized Roadmap
Four-phase remediation plan from immediate actions through runtime governance rollout.

Next Steps
Clear, time-bounded recommendations your team can act on immediately.

See the full sample deliverable
A complete 6 pages PDF with sample data, tailored to show exactly what you will receive.

Download Sample (PDF)

Sample data only · Client name and details are illustrative