

Agent Evaluation Harness & Red-Team Retainer

Monthly Ongoing · Foundry-Native · 12-Month Minimum

Weekly CI/CD regression · Monthly OWASP LLM Top 10 v2 · MITRE ATLAS v5.4 · Quarterly deep-scenario
AI Red Teaming Agent (PyRIT v0.11) · Defender for AI · Sentinel data lake · Copilot for Security SOAR

 M12 = SETUP (Implementation)	M16 = OPERATE (Retainer)	73% of enterprises run agents in production WITHOUT continuous eval (Gartner 2025)	MITRE ATLAS v5.4 added	11 agent-specific techniques	March 2026
---	---------------------------------	--	------------------------	------------------------------	------------

THE 90-DAY EVAL DECAY PROBLEM

A one-time eval harness setup (M12) **decays in 90 days** — OWASP LLM Top 10 v2 evolves quarterly, jailbreak corpora publish bypasses weekly, and MITRE ATLAS v5.4 added 11 agent-specific techniques in March 2026 (Publish Poisoned AI Agent Tool + Escape to Host + memory manipulation) that static test suites miss. SaaS platform vendors charge **\$30-\$400K/year recurring** (Lakera + HiddenLayer + Protect AI + CalypsoAI) with platform lock-in, not service. M16 delivers **native-Foundry ongoing operating retainer** at **\$180K-\$660K annual ACV** — weekly automated regression + monthly red-team + quarterly deep-scenario + 24x7 named engineers (Enterprise) + Anthropic MCP Security Working Group intelligence flow.

WHAT YOU RECEIVE MONTHLY

7 deliverables · Recurring · Native-Foundry

- Weekly CI/CD-gated regression via Foundry Evaluations SDK + Azure Monitor alerts on **≥5%** score deltas
- Monthly OWASP LLM Top 10 v2 attack battery (LLM01 + LLM02 + LLM05 + LLM08 + LLM09)
- Monthly MITRE ATLAS v5.4 agent-specific TTP simulation (Publish Poisoned + Escape to Host + memory)
- Monthly jailbreak corpus refresh + Anthropic MCP Security Working Group intelligence flow
- Quarterly deep-scenario red-team (business-logic abuse + chained tool + cross-agent A2A adversarial)
- Continuous Sentinel correlation + Defender for AI alerts + Purview DSPM data-lineage verification
- Monthly exec report (STANDARD+) · Quarterly Board pack (ENTERPRISE) (Caremark + SEC Item 106/1.05)

M12 → M16 SEQUENCING

PART 1

SETUP

M12

2–4 weeks · \$25K–\$75K one-time

- Implementation
- Foundry Evaluations + Tracing
- Defender for AI free-tier
- Sentinel data lake setup
- App Insights “Agents (Preview)”

↓

PART 2

OPERATE

M16

12-mo min · \$180K–\$660K annual

- Customer support retainer
- Weekly regression + drift
- Monthly red-team batteries
- Quarterly deep-scenario
- 24×7 named (Enterprise)

VS PLATFORM VENDORS

LAKERA+HL+PAI+CAI

\$30–\$400K/yr

platform lock-in

M16

Native-Foundry service · No lock-in

OWNERSHIP

Customer owns all evaluators

★ FIRST RECURRING SKU

M10–M15 = one-time engagements · M16 = FIRST RECURRING in M10–M18 Foundry chain · Parallel to M5 in Healthcare flagship.

INVESTMENT

FOUNDATION

\$15K/mo

\$180K annual · 12-mo min

- Single entity · 3 agents
- 8×5 P1 4hr response
- Monthly red-team + quarterly deep
- Weekly regression + drift detection

RECOMMENDED

STANDARD

\$32K/mo

\$384K annual · 3 entities · 5 agents

- 24×7 P1 2hr response
- Monthly exec briefing 30-min
- Custom evaluator authoring 2/mo
- Cross-entity scenarios

ENTERPRISE

\$55K/mo

\$660K annual · 10 entities · 10 agents

- 24×7 named engineers · P1 1hr
- Weekly exec readout 30-min
- Quarterly Board pack (Caremark + SEC 106/1.05)
- Multi-region + unlimited evaluators
- Dedicated Teams channel

Add-on: +\$5K/mo per additional agent · MACC drawdown ideal for 12-month retainer burn-down

- Microsoft AI Pursuit + ECIF \$25K–\$500K co-funding
- Tier-upgrade pro-rated mid-term
- Tier-downgrade at annual renewal only

WHY KRIV AI

★

Anthropic MCP Security Working Group

CPN-approved participant · Direct intelligence feed

▲

Named US-based engineers (Enterprise)

HIPAA BAA + SR 26–2 cleared · Not offshore SOC

✓

SAM.gov Registered

UEI: DATARJAUUXQ2