🔐 AI Security Market Research Report: February 2026

Market Snapshot: Major consolidation via acquisitions (Cisco→Robust Intelligence, Palo Alto→Protect AI, Check Point→Lakera, SentinelOne→Prompt Security). Landscape split: AI-native startups vs legacy security vendors adding LLM features. OWASP LLM Top 10 = de facto standard.

📊 Section A — Market Categories

1. 🛡️ LLM Firewalls/Gateways

- Protects: Prompt injection, jailbreaks, data exfil
- Buyers: CISO, AppSec
- Deploy: API proxy, SDK wrapper

2. 📋 AI Governance & GRC

- Protects: Compliance gaps, regulatory violations
- Buyers: GRC, CDO, Responsible AI
- Deploy: SaaS dashboard, policy engine

3. 📈 AI Runtime Monitoring

- Protects: Drift, hallucinations, cost overruns
- Buyers: ML Engineers, Platform Engineers
- Deploy: SDK instrumentation, OpenTelemetry

4. 🔒 GenAI DLP & PII Redaction

- Protects: PII/PHI in prompts, secrets exposure

- Buyers: CISO, DLP teams
- Deploy: API scanner, browser plugin, proxy

5. ⚔️ AI Red Teaming & Evals

- Protects: Jailbreaks, safety violations, adversarial attacks
- Buyers: Security teams, AI Safety
- Deploy: CLI tools, eval harness, services

6. 📦 Model Supply Chain Security

- Protects: Malicious models, backdoors, tampering
- Buyers: MLOps, Security
- Deploy: CI/CD scanner, model registry

7. 🔍 AI-SPM (Security Posture Management)

- Protects: Shadow AI, misconfigurations, attack paths
- Buyers: Cloud Security, CISO
- Deploy: Agentless cloud scanner

8. 🔑 Agent Identity & Access

- Protects: Unauthorized tool use, privilege escalation
- Buyers: IAM teams, Platform Engineers
- Deploy: OAuth/OIDC, policy-as-code

🏢 Section B — Vendor Details

🛡️ LLM Firewalls/Gateways

Lakera (acquired by Check Point)

- 🛡️ Protects: Prompt injection, jailbreaks, data leakage, toxic content
- ⚙️ How: Classifier <50ms latency, 1M+ Gandalf game threat intel
- 🎯 Target: Fortune 500, AI-native companies
- 🧬 Origin: AI-native (aerospace safety founders)
- 📦 OSS: Gandalf game only
- 🌐 Deploy: SaaS (platform.lakera.ai)
- 💰 Price: Enterprise; free tier available
- 🔗 Links: lakera.ai, docs.lakera.ai
- ⭐ Confidence: High

Protect AI (acquired by Palo Alto Networks)

- 🛡️ Protects: Prompt injection, jailbreaks, PII leakage, model deserialization
- ⚙️ How: I/O scanners, CPU-optimized inference, regex+LLM detection
- 🎯 Target: Enterprise DevSecOps teams
- 🧬 Origin: AI-native (Amazon-Oracle founders)
- 📦 OSS: **YES** - LLM Guard, ModelScan (github.com/protectai/llm-guard)
- 🌐 Deploy: API/library, self-hosted
- 💰 Price: OSS free; enterprise pricing TBD
- 🔗 Links: protectai.com
- ⭐ Confidence: High

Robust Intelligence (acquired by Cisco)

- 🛡️ Protects: Prompt injection, data poisoning, jailbreaking, model theft
- ⚙️ How: 3 modules - Model scanner + AI Validation + AI Protection (runtime)
- 🎯 Target: Fortune 500 (JPMorgan, IBM, BMW)
- 🧬 Origin: AI-native (Harvard AI research)
- 📦 OSS: No
- 🌐 Deploy: Cisco Security Cloud integration
- 💰 Price: Enterprise (Cisco pricing)
- 🔗 Links: cisco.com/robust-intelligence
- ⭐ Confidence: High

Lasso Security

- 🛡️ Protects: Prompt injection, guardrail bypasses, agent misuse, shadow AI
- ⚙️ How: MCP Secure Gateway, plugin architecture, token masking, tool reputation
- 🎯 Target: Enterprise security teams
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: SaaS
- 💰 Price: Enterprise only
- 🔗 Links: lasso.security
- ⭐ Confidence: High

CalypsoAI

- 🛡️ Protects: Prompt injection, PII leakage, toxic content, jailbreaks
- ⚙️ How: 3 pillars - Red-Team (pre-test) + Defend (real-time) + Observe (audit)

- 🎯 Target: Enterprise, finance, government
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: SaaS + API
- 💰 Price: Enterprise; beta free tier
- 🔗 Links: calypsoai.com
- ⭐ Confidence: High

Arthur AI

- 🛡️ Protects: PII leakage, toxic language, hallucinations, prompt injection
- ⚙️ How: Firewall between app and model; validates prompts/responses
- 🎯 Target: Enterprise LLM deployments
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: API endpoints
- 💰 Price: Enterprise
- 🔗 Links: arthur.ai/firewall
- ⭐ Confidence: High

HiddenLayer

- 🛡️ Protects: Prompt injection, model extraction, poisoning, malware, backdoors
- ⚙️ How: 4 modules - Discovery + Supply Chain + Attack Simulation + Runtime (AIDR)
- 🎯 Target: Enterprise, Federal, Finance
- 🧬 Origin: AI-native (Cylance veterans)
- 📦 OSS: No

- 🌐 Deploy: SaaS, on-prem, air-gapped
- 💰 Price: Enterprise (AWS/Azure Marketplace)
- 🔗 Links: hiddenlayer.com
- ⭐ Confidence: High

Pangea

- 🛡️ Protects: Prompt injection, PII/PHI, malicious URLs/IPs, toxic content, secrets
- ⚙️ How: API-based "recipes" (detector collections); block/report/redact/encrypt
- 🎯 Target: Developers, enterprise
- 🧬 Origin: AI-native
- 📦 OSS: MCP Proxy on GitHub
- 🌐 Deploy: SaaS, self-hosted, gateway
- 💰 Price: Free tier; usage-based
- 🔗 Links: pangea.cloud, pangea.cloud/docs/ai-guard
- ⭐ Confidence: High

Prompt Security (acquired by SentinelOne)

- 🛡️ Protects: Shadow AI, prompt injection, data privacy, MCP security
- ⚙️ How: OWASP-aligned controls, real-time enforcement, browser extension
- 🎯 Target: Enterprise
- 🧬 Origin: AI-native (OWASP LLM Top 10 core team)
- 📦 OSS: Prompt Fuzzer
- 🌐 Deploy: SaaS, browser
- 💰 Price: Enterprise
- 🔗 Links: prompt.security
- ⭐ Confidence: High

NVIDIA NeMo Guardrails

- 🛡️ Protects: Jailbreaks, topic drift, PII, unsafe content, hallucinations
- ⚙️ How: 5 rail types via Colang DSL (Input, Dialog, Retrieval, Execution, Output)
- 🎯 Target: Enterprise AI developers
- 🧬 Origin: AI-native (NVIDIA)
- 📦 OSS: **YES** - Apache 2.0 (github.com/NVIDIA-NeMo/Guardrails)
- 🌐 Deploy: Library, Docker, NIM
- 💰 Price: Free; enterprise support available
- 🔗 Links: developer.nvidia.com/nemo-guardrails
- ⭐ Confidence: High

Guardrails AI

- 🛡️ Protects: Structured validation, PII, toxicity, prompt injection, hallucinations
- ⚙️ How: Guard wrappers around LLM calls; validators from Guardrails Hub
- 🎯 Target: ML engineers, developers
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - GitHub (github.com/guardrails-ai/guardrails)
- 🌐 Deploy: Self-hosted, Flask service
- 💰 Price: Free (OSS); enterprise plans
- 🔗 Links: guardrailsai.com
- ⭐ Confidence: High

📋 AI Governance & GRC

Credo AI

- 🛡️ Protects: Bias, compliance gaps, regulatory violations, third-party AI risk
- ⚙️ How: Policy Packs encoding regulations (EU AI Act, NIST, ISO 42001); AI Registry; Policy Intelligence Engine
- 🎯 Target: CISO, GRC, CDO, Responsible AI teams
- 🧬 Origin: AI-native (2020)
- 📦 OSS: Credo Lens
- 🌐 Deploy: Cloud SaaS
- 💰 Price: Enterprise
- 🔗 Links: credo.ai, credo.ai/product
- ⭐ Confidence: High

Fiddler AI

- 🛡️ Protects: Model drift, integrity issues, bias, fairness, compliance (OCC)
- ⚙️ How: Explainable AI (Shapley, Integrated Gradients); continuous monitoring; real-time alerting
- 🎯 Target: Data scientists, MRM teams, Fortune 500
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: Cloud SaaS (integrates Datadog)
- 💰 Price: Events/models based
- 🔗 Links: fiddler.ai
- ⭐ Confidence: High

ModelOp

- 🛡️ Protects: AI lifecycle fragmentation, compliance, policy enforcement
- ⚙️ How: AI system of record; workflow automation; templates for NIST/EU AI Act/ISO 42001

- 🎯 Target: Complex enterprises
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: Enterprise deployment
- 💰 Price: Enterprise
- 🔗 Links: modelop.com
- ⭐ Confidence: Medium

Fairly AI (now Asenion)

- 🛡️ Protects: AI risk, security, governance gaps
- ⚙️ How: End-to-end AI risk management; patent-pending technology
- 🎯 Target: Enterprise
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: SaaS
- 💰 Price: Enterprise
- 🔗 Links: fairly.ai
- ⭐ Confidence: Medium

📈 AI Runtime Monitoring/Observability

Arize AI

- 🛡️ Protects: Hallucinations, RAG quality issues, agent failures, drift, cost
- ⚙️ How: Phoenix (OSS) - OpenTelemetry tracing, LLM-as-judge evals, UMAP visualization
- 🎯 Target: AI engineers (PepsiCo, Siemens)
- 🧬 Origin: AI-native

- 📦 OSS: **YES** - Phoenix (Elastic License 2.0) - github.com/Arize-ai/phoenix
- 🌐 Deploy: Self-hosted, Phoenix Cloud
- 💰 Price: Phoenix free; AX from $50/mo
- 🔗 Links: arize.com
- ⭐ Confidence: High

WhyLabs

- 🛡️ Protects: Data drift, quality, toxicity, prompt injection, PII, hallucinations
- ⚙️ How: whylogs (privacy-preserving profiling); LangKit; containerized guardrails; MITRE ATLAS aligned
- 🎯 Target: Data science, MLOps teams
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - Platform open-sourced
- 🌐 Deploy: SaaS, self-hosted, VPC
- 💰 Price: Free tier; enterprise plans
- 🔗 Links: whylabs.ai, docs.whylabs.ai
- ⭐ Confidence: High

Langfuse

- 🛡️ Protects: Debugging, prompt quality, latency, cost, regressions
- ⚙️ How: OpenTelemetry tracing; @observe() decorator; prompt management; LLM-as-judge evals
- 🎯 Target: LLM developers
- 🧬 Origin: AI-native (YC W23)
- 📦 OSS: **YES** - Self-hostable - github.com/langfuse/langfuse
- 🌐 Deploy: Self-hosted, Langfuse Cloud

- 💰 Price: Free self-host; Cloud from $59/mo
- 🔗 Links: langfuse.com
- ⭐ Confidence: High

Portkey

- 🛡️ Protects: Provider outages, cost, latency, security, prompt injection
- ⚙️ How: Gateway routes to 200+ LLMs; 50+ guardrails; fallbacks; retries; MCP Gateway
- 🎯 Target: GenAI builders, platform engineers
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - Gateway OSS - github.com/Portkey-AI/gateway
- 🌐 Deploy: Cloud SaaS, private cloud
- 💰 Price: Tiered; enterprise plans
- 🔗 Links: portkey.ai
- ⭐ Confidence: High

Helicone

- 🛡️ Protects: Cost, latency, prompt injection, quality
- ⚙️ How: One-line proxy integration; session tracking; built-in caching; Meta Prompt Guard/Llama Guard
- 🎯 Target: Developers
- 🧬 Origin: AI-native (YC W23)
- 📦 OSS: **YES** - github.com/Helicone/helicone
- 🌐 Deploy: SaaS, self-hosted
- 💰 Price: Free 10k req/mo; paid plans
- 🔗 Links: helicone.ai

- ⭐ Confidence: High

Patronus AI

- 🛡️ Protects: Hallucinations, unsafe content, PII/PHI, toxicity, prompt injection
- ⚙️ How: SimpleSafetyTests; GLIDER (3B eval model); Lynx hallucination model; Percival debugger
- 🎯 Target: AI developers, ML engineers
- 🧬 Origin: AI-native (2023)
- 📦 OSS: SimpleSafetyTests public
- 🌐 Deploy: Cloud API, SDK
- 💰 Price: API-based
- 🔗 Links: patronus.ai, docs.patronus.ai
- ⭐ Confidence: High

🔒 GenAI DLP & PII Redaction

Nightfall AI

- 🛡️ Protects: PII, PHI, PCI, secrets, API keys, NHIs, IP
- ⚙️ How: ML-based detectors (95%+ precision); real-time scanning; redact/delete/block; browser plugin
- 🎯 Target: Tech, healthcare, finance
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: SaaS, browser extensions
- 💰 Price: Tiered SaaS
- 🔗 Links: nightfall.ai

- ⭐ Confidence: High

Private AI

- 🛡️ Protects: PII across 50+ types in 47 languages; text, PDF, images, audio
- ⚙️ How: Transformer NLP; redaction, masking, synthetic replacement; deterministic tokenization
- 🎯 Target: Enterprises fine-tuning LLMs
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: On-prem, private cloud, SaaS
- 💰 Price: Free tier (75 calls/day); enterprise plans
- 🔗 Links: private-ai.com
- ⭐ Confidence: High

Microsoft Presidio

- 🛡️ Protects: 180+ PII entity types
- ⚙️ How: Analyzer (NER, regex, checksum) + Anonymizer (replace, hash, encrypt); pluggable NLP
- 🎯 Target: Developers, Azure/Fabric users
- 🧬 Origin: Open-source toolkit
- 📦 OSS: **YES** - Apache 2.0 - github.com/microsoft/presidio
- 🌐 Deploy: Python, Docker, K8s, REST
- 💰 Price: Free
- 🔗 Links: github.com/microsoft/presidio
- ⭐ Confidence: High

Strac

- 🛡️ Protects: PII, PHI, PCI, secrets, source code
- ⚙️ How: ML/OCR detection; redaction, tokenization, vault; GenAI DLP via browser extension
- 🎯 Target: Mid-market to enterprise
- 🧬 Origin: AI-native (YC)
- 📦 OSS: No
- 🌐 Deploy: SaaS (agentless)
- 💰 Price: SaaS subscription
- 🔗 Links: strac.io
- ⭐ Confidence: High

Skyflow

- 🛡️ Protects: PII, PCI, PHI, IP for LLM training/inference
- ⚙️ How: Data privacy vault; tokenization preserving LLM quality; polymorphic encryption
- 🎯 Target: Fintech, healthcare
- 🧬 Origin: Privacy-native → AI
- 📦 OSS: No
- 🌐 Deploy: SaaS, multi-cloud
- 💰 Price: Usage-based
- 🔗 Links: skyflow.com/llm-vault
- ⭐ Confidence: High

LLM Guard (Protect AI)

- 🛡️ Protects: PII anonymization, prompt injection, jailbreaks, toxic content
- ⚙️ How: Input scanners (Anonymize, PromptInjection, Toxicity); Output scanners (Deanonymize, Sensitive)

- 🎯 Target: Developers, security teams
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - Permissive license - github.com/protectai/llm-guard
- 🌐 Deploy: Python package, API server
- 💰 Price: Free (OSS)
- 🔗 Links: llm-guard.com
- ⭐ Confidence: High

GitGuardian

- 🛡️ Protects: Secrets/credentials, API keys (500+ types), tokens in code and prompts
- ⚙️ How: Multi-layer - regex, entropy, contextual validation; MCP server for LLM gateway
- 🎯 Target: Developers, DevSecOps
- 🧬 Origin: Security-native → AI
- 📦 OSS: **YES** - ggshield CLI - github.com/GitGuardian/ggshield
- 🌐 Deploy: SaaS, CLI, CI/CD
- 💰 Price: Free tier; Business plans
- 🔗 Links: gitguardian.com
- ⭐ Confidence: High

⚔️ AI Red Teaming & Evaluations

Garak (NVIDIA)

- 🛡️ Protects: Prompt injection, jailbreaks, toxicity, data leakage, bias
- ⚙️ How: Probes → Generators → Detectors → Harness; static/dynamic/adaptive probing
- 🎯 Target: Security teams, researchers
- 🧬 Origin: AI-native

- 📦 OSS: **YES** - Apache 2.0 - github.com/NVIDIA/garak
- 🌐 Deploy: CLI tool
- 💰 Price: Free
- 🔗 Links: garak.ai
- ⭐ Confidence: High

Promptfoo

- 🛡️ Protects: Security vulns, jailbreaks, regressions, OWASP Top 10
- ⚙️ How: Declarative YAML config; eval harness; LLM-as-judge; CI/CD native
- 🎯 Target: Developers, DevOps
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - github.com/promptfoo/promptfoo
- 🌐 Deploy: CLI, npm, CI/CD
- 💰 Price: Free
- 🔗 Links: promptfoo.dev
- ⭐ Confidence: High

Adversa AI

- 🛡️ Protects: Prompt injection, jailbreaks, goal hijacking, tool misuse, memory poisoning
- ⚙️ How: Automated continuous red teaming; patented attack generation; agentic AI scenarios
- 🎯 Target: Fortune 500, finance
- 🧬 Origin: AI-native
- 📦 OSS: Research published
- 🌐 Deploy: SaaS
- 💰 Price: Enterprise
- 🔗 Links: adversa.ai

- ⭐ Confidence: High

Mindgard

- 🛡️ Protects: Jailbreaks, prompt injection, model extraction, evasion, poisoning
- ⚙️ How: Automated red teaming; PhD-led attack scenarios; runtime detection; SIEM integration
- 🎯 Target: Finance, healthcare, manufacturing
- 🧬 Origin: AI-native (2016 research, 2022 commercial)
- 📦 OSS: No
- 🌐 Deploy: SaaS
- 💰 Price: Enterprise
- 🔗 Links: mindgard.ai
- ⭐ Confidence: High

Haize Labs

- 🛡️ Protects: Jailbreaks, safety violations, multimodal vulnerabilities
- ⚙️ How: ACG attacks (38x faster than GCG); DSPy-based optimization; model-based evaluators
- 🎯 Target: Model providers (OpenAI, Anthropic)
- 🧬 Origin: AI-native / research-driven
- 📦 OSS: dspy-redteam, j1-micro reward models - github.com/haizelabs
- 🌐 Deploy: Platform + services
- 💰 Price: Enterprise
- 🔗 Links: haizelabs.com
- ⭐ Confidence: High

Inspect (UK AI Safety Institute)

- 🛡️ Protects: Reasoning, autonomous capabilities, safety behaviors
- ⚙️ How: Python framework - Datasets → Solvers → Scorers; 100+ pre-built evals; MCP tool calling
- 🎯 Target: AI Safety Institutes, researchers
- 🧬 Origin: AI-native (government)
- 📦 OSS: **YES** - github.com/UKGovernmentBEIS/inspect_ai
- 🌐 Deploy: Python package
- 💰 Price: Free
- 🔗 Links: inspect.aisi.org.uk
- ⭐ Confidence: High

AI Verify (Singapore IMDA)

- 🛡️ Protects: 11 AI ethics principles, fairness, explainability, robustness
- ⚙️ How: Technical tests + process checks; GenAI via Project Moonshot
- 🎯 Target: Enterprises, auditors
- 🧬 Origin: Government framework
- 📦 OSS: **YES** - GitHub
- 🌐 Deploy: Local (Docker)
- 💰 Price: Free
- 🔗 Links: aiverifyfoundation.sg
- ⭐ Confidence: High

Deepchecks

- 🛡️ Protects: Hallucinations, groundedness, bias, toxicity, PII, drift
- ⚙️ How: Swarm of SLMs + NLP; no-code LLM-as-judge; Golden Set management

- 🎯 Target: ML engineers, product owners
- 🧬 Origin: AI-native
- 📦 OSS: Testing package OSS
- 🌐 Deploy: SaaS
- 💰 Price: Commercial tiered
- 🔗 Links: deepchecks.com
- ⭐ Confidence: High

Scale AI

- 🛡️ Protects: Safety policy violations, jailbreaks, over-refusal, agentic risks
- ⚙️ How: Expert human + model-assisted approaches; SEAL research lab; product-focused threat modeling
- 🎯 Target: Model developers, government
- 🧬 Origin: Hybrid (human + AI)
- 📦 OSS: Research publications
- 🌐 Deploy: Managed service
- 💰 Price: Enterprise
- 🔗 Links: scale.com/evaluation
- ⭐ Confidence: High

HuggingFace LightEval

- 🛡️ Protects: Standard benchmarks (MMLU, TruthfulQA, GSM8K), safety (XSTest)
- ⚙️ How: Multiple backends; 1000+ eval tasks; Open LLM Leaderboard integration
- 🎯 Target: Researchers, ML practitioners
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - github.com/huggingface/lighteval

- 🌐 Deploy: Python library
- 💰 Price: Free
- 🔗 Links: github.com/huggingface/lighteval
- ⭐ Confidence: High

📦 Model Supply Chain Security

HiddenLayer

- 🛡️ Protects: Malware in models, serialization attacks, backdoors, CVE exploits
- ⚙️ How: Model scanning (35+ formats); AIBOM generation; model genealogy; CI/CD integration
- 🎯 Target: CISO, MLOps, Federal
- 🧬 Origin: AI-native
- 📦 OSS: No
- 🌐 Deploy: SaaS, on-prem, air-gapped
- 💰 Price: Enterprise
- 🔗 Links: hiddenlayer.com/aidr
- ⭐ Confidence: High

Protect AI Guardian

- 🛡️ Protects: Deserialization attacks, architectural backdoors, runtime threats
- ⚙️ How: Scans 35+ model formats; native HuggingFace integration; AI model gateway
- 🎯 Target: Security, ML engineers
- 🧬 Origin: AI-native
- 📦 OSS: ModelScan - github.com/protectai/modelscan
- 🌐 Deploy: SaaS, on-prem, CI/CD
- 💰 Price: Enterprise (Palo Alto)

- 🔗 Links: protectai.com/guardian
- ⭐ Confidence: High

JFrog ML

- 🛡️ Protects: Serialization attacks, malicious code, CVEs, backdoors, license issues
- ⚙️ How: Deep decompilation; HuggingFace partnership; 96% FP reduction; model versioning
- 🎯 Target: DevOps, MLOps, Data Scientists
- 🧬 Origin: Legacy DevSecOps → ML
- 📦 OSS: No
- 🌐 Deploy: SaaS, self-hosted, hybrid
- 💰 Price: JFrog Platform pricing
- 🔗 Links: jfrog.com/mlsecops
- ⭐ Confidence: High

Sigstore Model Signing

- 🛡️ Protects: Tampering, integrity, supply chain attacks
- ⚙️ How: Cryptographic signing via Sigstore; Rekor transparency log; K8s validation operator
- 🎯 Target: Model hubs, MLOps
- 🧬 Origin: OSS standard (Google, OpenSSF)
- 📦 OSS: **YES** - Apache 2.0 - github.com/sigstore/model-transparency
- 🌐 Deploy: CLI, library, K8s
- 💰 Price: Free
- 🔗 Links: github.com/sigstore/model-transparency
- ⭐ Confidence: High

Bosch AIShield

- 🛡️ Protects: Model extraction, evasion, poisoning, inference attacks, notebook vulns
- ⚙️ How: AISpectra (discovery + scanning); Guardian (runtime); Watchtower (OSS scanner)
- 🎯 Target: Manufacturing, automotive, IoT
- 🧬 Origin: AI-native (Bosch incubated)
- 📦 OSS: Watchtower - github.com/bosch-aisecurity-aishield/watchtower
- 🌐 Deploy: SaaS, edge, consulting
- 💰 Price: SaaS; pay-per-assessment
- 🔗 Links: boschaishield.com
- ⭐ Confidence: High

Mithril Security

- 🛡️ Protects: Data privacy during inference, model IP, data leakage to providers
- ⚙️ How: BlindBox - TEEs (Intel SGX/AMD SEV); AICert - cryptographic provenance via TPMs
- 🎯 Target: Healthcare, legal, regulated industries
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - BlindBox, BlindAI - github.com/mithril-security/blindbox
- 🌐 Deploy: Azure Confidential, self-hosted
- 💰 Price: Startup pricing
- 🔗 Links: mithrilsecurity.io
- ⭐ Confidence: Medium

🔑 Agent Identity & Access

LangChain/LangGraph

- 🛡️ Protects: Unauthorized tool use, resource access
- ⚙️ How: @auth.authenticate, @auth.on handlers; owner-based scoping; Permit.io integration

- 🎯 Target: Developers
- 🧬 Origin: AI-native
- 📦 OSS: **YES** - github.com/langchain-ai/langchain
- 🌐 Deploy: Self-hosted
- 💰 Price: Free
- 🔗 Links: docs.langchain.com
- ⭐ Confidence: High

Okta Auth0

- 🛡️ Protects: Agent authentication, unauthorized tool access, delegation abuse
- ⚙️ How: OAuth 2.0 flows; full auditability; MCP protocol support
- 🎯 Target: Enterprise
- 🧬 Origin: Legacy IAM → Agents
- 📦 OSS: No
- 🌐 Deploy: SaaS
- 💰 Price: Enterprise
- 🔗 Links: auth0.com
- ⭐ Confidence: High

Ping Identity

- 🛡️ Protects: Agent identity, lifecycle governance, privilege escalation
- ⚙️ How: 5 pillars - Visibility, Onboarding, AuthN/AuthZ, Human Oversight, Threat Protection
- 🎯 Target: Enterprise
- 🧬 Origin: Legacy IAM → Agents
- 📦 OSS: No

- 🌐 Deploy: SaaS, on-prem
- 💰 Price: Enterprise
- 🔗 Links: pingidentity.com/agentic-ai
- ⭐ Confidence: High

CyberArk

- 🛡️ Protects: Machine identity abuse, secrets exposure, privilege escalation
- ⚙️ How: Zero Trust for agents; automated secret rotation; agent lifecycle management
- 🎯 Target: Enterprise with PAM
- 🧬 Origin: Legacy PAM → Agents
- 📦 OSS: No
- 🌐 Deploy: SaaS, on-prem
- 💰 Price: Enterprise
- 🔗 Links: cyberark.com/agentic-ai
- ⭐ Confidence: High

Strata Identity

- 🛡️ Protects: Delegation chains, unauthorized tool use, credential abuse
- ⚙️ How: JIT identity provisioning; OAuth OBO; policy-as-code; HITL for high-risk actions
- 🎯 Target: Multi-cloud enterprises
- 🧬 Origin: IAM orchestration → Agents
- 📦 OSS: No
- 🌐 Deploy: SaaS
- 💰 Price: Enterprise
- 🔗 Links: strata.io/ai-agent-identity
- ⭐ Confidence: High

Permit.io

- 🛡️ Protects: Unauthorized actions, document access violations
- ⚙️ How: RBAC/ABAC/ReBAC policies; JWT validation; LangChain integration
- 🎯 Target: Developers
- 🧬 Origin: Auth-native → AI
- 📦 OSS: LangChain-Permit integration
- 🌐 Deploy: SaaS
- 💰 Price: Tiered
- 🔗 Links: permit.io/langchain
- ⭐ Confidence: High

🏛️ Section C — Legacy Security Vendors Adding LLM

Palo Alto Networks

- 🔧 Traditional: NGFW, Prisma SASE, CASB
- ✨ AI Features: AI Access Security (500+ GenAI apps, DLP for prompts), AI Runtime Security (prompt injection via NVIDIA NeMo), AI-SPM, Prisma AIRS
- 📅 Launch: May 2024 (Precision AI); GA Q4 FY24
- 🔄 Integration: **Deep** - native to Strata platform, unified DLP/CASB, single management
- ⚙️ Mechanism: Proxy + policy engine + NVIDIA integration
- 🔗 Link: paloaltonetworks.com/sase/ai-access-security

CrowdStrike

- 🔧 Traditional: Falcon EDR/XDR

- ✨ AI Features: AI-SPM (shadow AI discovery), Falcon Data Protection for GenAI (browser + local apps), AI Red Team Services
- 📅 Launch: Fal.Con Sept 2025; GenAI DLP GA Jan 2026
- 🔄 Integration: **Deep** - single sensor architecture extends to AI
- ⚙️ Mechanism: Lightweight agent, browser-based
- 🔗 Link: crowdstrike.com/solutions/secure-your-ai

Microsoft Purview

- 🔧 Traditional: DLP, Compliance, eDiscovery
- ✨ AI Features: DSPM for AI, AI-specific DLP policies (Copilot + 100+ third-party), Communication Compliance for AI, Agent 365 integration
- 📅 Launch: Dec 2023 preview; Agent governance Nov 2025
- 🔄 Integration: **Deep** - same sensitivity labels and policies
- ⚙️ Mechanism: Policy engine, browser controls
- 🔗 Link: learn.microsoft.com/purview/ai-microsoft-purview

Zscaler

- 🔧 Traditional: ZIA/ZPA, SWG, CASB
- ✨ AI Features: GenAI Security (shadow AI, prompt DLP), Browser Isolation for GenAI, SSPM for Copilot
- 📅 Launch: 2023 ChatGPT controls; expanded 2024-25
- 🔄 Integration: **Integrated** via platform - extends Zero Trust Exchange
- ⚙️ Mechanism: Proxy-based inline inspection
- 🔗 Link: zscaler.com/products-and-solutions/securing-generative-ai

Cloudflare

- 🔧 Traditional: CDN, WAF, DDoS
- ✨ AI Features: Firewall for AI (prompt injection, PII, toxic via Llama Guard), AI Gateway (caching, rate limiting, cost)
- 📅 Launch: AI Gateway 2023; Firewall for AI beta 2025
- 🔄 Integration: **Moderately bolt-on** - new products leveraging edge
- ⚙️ Mechanism: Reverse proxy, edge deployment
- 🔗 Link: cloudflare.com/application-services/products/firewall-for-ai

Cisco

- 🔧 Traditional: Network, ASA/Firepower, Splunk SIEM
- ✨ AI Features: Cisco AI Defense (via Robust Intelligence acquisition) - AI Firewall, model scanning, AI Validation
- 📅 Launch: Acquisition Oct 2024
- 🔄 Integration: **Acquisition bolt-on** - integrating into Security Cloud
- ⚙️ Mechanism: Robust Intelligence tech being integrated
- 🔗 Link: cisco.com/robust-intelligence

Fortinet

- 🔧 Traditional: FortiGate, FortiSIEM, FortiEDR
- ✨ AI Features: FortiAI-Protect (GenAI visibility, shadow AI, DLP, promptware), FortiGuard AI-Powered Services
- 📅 Launch: 2023 FortiAI; expanded 2024-25
- 🔄 Integration: **Deep** - embedded across Security Fabric
- ⚙️ Mechanism: Firewall + policy engine
- 🔗 Link: fortinet.com/solutions/enterprise-midsize-business/fortiai

Splunk (Cisco)

- 🔧 Traditional: SIEM, Log Management
- ✨ AI Features: AI Agent Monitoring (Alpha), AI Infrastructure Monitoring (GA), LLM Observability, APM LLM Services
- 📅 Launch: 2024-25; AI Agent at .conf25
- 🔄 Integration: **Integrated** via platform - extends Observability Cloud
- ⚙️ Mechanism: OpenTelemetry instrumentation
- 🔗 Link: splunk.com/genai-observability

Datadog

- 🔧 Traditional: APM, Infrastructure Monitoring
- ✨ AI Features: LLM Observability (GA Aug 2024), AI Agent Monitoring (GA), AI Agents Console (Preview), Sensitive Data Scanner
- 📅 Launch: Aug 2024 GA; expanded 2025
- 🔄 Integration: **Integrated** via platform - correlates with APM/RUM/Security
- ⚙️ Mechanism: SDK instrumentation
- 🔗 Link: datadoghq.com/product/llm-observability

Wiz

- 🔧 Traditional: CSPM, CNAPP
- ✨ AI Features: AI-SPM (first CNAPP with AI-SPM Nov 2023), AI-BOM, AI Attack Path Analysis, DSPM for AI
- 📅 Launch: Nov 2023
- 🔄 Integration: **Deep** - native Security Graph extension
- ⚙️ Mechanism: Agentless cloud scanning

- 🔗 Link: wiz.io/solutions/ai-spm

Snyk

- 🔧 Traditional: SAST, SCA, Container Security
- ✨ AI Features: Snyk Code for AI-generated code, MCP Server integration, OWASP LLM education
- 📅 Launch: 2024-25; OWASP sponsorship Nov 2024
- 🔄 Integration: **Integrated** via platform - extends existing SAST
- ⚙️ Mechanism: IDE/CI scanner
- 🔗 Link: snyk.io/solutions/secure-ai-generated-code

Tenable

- 🔧 Traditional: Nessus, Tenable One
- ✨ AI Features: Tenable One AI Exposure (GA Jan 2026), AI-SPM, AI Aware, ExposureAI
- 📅 Launch: AI Aware 2023; AI Exposure GA Jan 2026
- 🔄 Integration: **Deep** - part of Tenable One platform
- ⚙️ Mechanism: Agentless scanning
- 🔗 Link: tenable.com/products/ai-exposure

Check Point

- 🔧 Traditional: Quantum Firewall, CloudGuard, Harmony
- ✨ AI Features: GenAI Protect (shadow AI, DLP), AI Cloud Protect (NVIDIA partnership), Harmony DLP with GenAI
- 📅 Launch: 2024 preview; Lakera acquisition pending Q4 2025
- 🔄 Integration: **Moderate** - SaaS portal + firewall; Lakera will bolster
- ⚙️ Mechanism: SaaS + firewall policy

- 🔗 Link: checkpoint.com/solutions/genai-security

Akamai

- 🔧 Traditional: CDN, WAF, Bot Management
- ✨ AI Features: Firewall for AI (prompt injection, PII, toxic output), API LLM Discovery
- 📅 Launch: March 2025
- 🔄 Integration: **Moderately bolt-on** - new purpose-built product
- ⚙️ Mechanism: Edge/reverse proxy
- 🔗 Link: akamai.com/products/firewall-for-ai

SentinelOne

- 🔧 Traditional: Falcon EDR/XDR, CNAPP
- ✨ AI Features: Prompt Security acquisition - GenAI DLP, prompt injection, shadow AI, jailbreak prevention
- 📅 Launch: 2025 acquisition
- 🔄 Integration: **Acquisition integrating** - being merged into Singularity
- ⚙️ Mechanism: Agent + browser extension
- 🔗 Link: sentinelone.com/prompt-security

Rapid7

- 🔧 Traditional: InsightIDR, InsightAppSec, InsightCloudSec
- ✨ AI Features: AI Attack Coverage (6 OWASP LLM modules), AI/ML compliance pack (Bedrock, Azure OpenAI, Vertex), LLM scanning
- 📅 Launch: Dec 2023 compliance; June 2025 attack coverage
- 🔄 Integration: **Integrated** via platform - extends AppSec/CloudSec
- ⚙️ Mechanism: DAST + compliance checks

- 🔗 Link: rapid7.com/ai-attack-coverage

💡 Section D — Quick Takeaways

1. 🏟️ **LLM Firewalls = Crowded Space**
   - Lakera, Protect AI, CalypsoAI, Arthur, HiddenLayer, Pangea, Prompt Security all compete for "proxy between app and model"
   - Differentiation: latency claims, threat intel sources, compliance certifications
2. 📈 **AI-SPM = Fastest-Growing Category**
   - Wiz pioneered Nov 2023 → Now CrowdStrike, Palo Alto, Tenable all offer
   - Prediction: Every CNAPP vendor will add AI-SPM within 12 months
3. 📦 **Model Supply Chain = Under-Adopted**
   - Despite high-profile risks (malicious HuggingFace models), only HiddenLayer, Protect AI/JFrog, Bosch AIShield offer mature scanning
   - Sigstore model signing emerging as open standard
4. 🔑 **Agent Identity = Nascent & Fragmented**
   - OpenID Foundation & CSA published standards, but most solutions (Okta, Ping, CyberArk) adapt existing IAM
   - No clear leader for agentic-native identity yet
   - LangChain and Permit.io lead in developer tooling
5. ⌛ **Watch Acquisition Integration Delays**
   - Cisco/Robust Intelligence, Palo Alto/Protect AI, Check Point/Lakera, SentinelOne/Prompt Security all mid-integration
   - ⚠️ Verify current feature availability, not acquisition press releases
6. 🌐 **Open Source = Table Stakes**

   - LLM Guard (2.5M+ downloads), NeMo Guardrails, Garak, Promptfoo, Presidio, Guardrails AI, Langfuse dominate developer mindshare
   - Commercial vendors increasingly open-source components to gain adoption

7. 🎯 **Common Blind Spot: Indirect Prompt Injection**
   - Most vendors focus on direct prompt attacks
   - Indirect injection (via documents, tool outputs, RAG retrieval) requires different detection approaches
   - PromptArmor & Adversa AI specialize in this area

8. 💰 **Pricing Opacity = Universal**
   - Nearly all enterprise vendors require "contact sales"
   - OSS alternatives or free tiers (Pangea, Helicone, LLM Guard) may be sufficient for early-stage deployments

9. 🎯 **Buyer Guide by Use Case: Building RAG Application:**
   - ✅ Guardrails AI + Langfuse + LLM Guard (all OSS-first)

   Agentic Workflows:
   - ✅ Lasso Security + Adversa AI + LangChain auth + Portkey

   Internal Copilot:
   - ✅ Nightfall AI + Microsoft Purview + Credo AI

   Model Fine-Tuning:
   - ✅ Private AI + HiddenLayer + Sigstore signing

   Shadow AI Governance:
   - ✅ Wiz AI-SPM + CrowdStrike + Zscaler

10. ⚖️ **Regulatory Deadlines = Driving Urgency**
    - 🇪🇺 EU AI Act full compliance required August 2026

- Vendors aligning to NIST AI RMF, ISO 42001, OWASP LLM Top 10 will win regulated buyers
- Credo AI & AI Verify (Singapore) lead in compliance frameworks

---

📊 **Report Compiled:** February 5, 2026
🔍 **Methodology:** Primary source verification (vendor docs, product pages, GitHub, pricing pages) + reputable third-party validation
⭐ **Quality Standard:** Evidence-based only; unknowns explicitly labeled