Sentinel AI — Zero Trust Security for AI Agent Networks

The Okta for AI Agents

"Trust no agent. Verify everything. Secure the agentic future."

Executive Summary

Sentinel AI is the zero trust security platform purpose-built for the AI agent economy. As enterprises deploy autonomous AI agents that interact with APIs, databases, external services, and each other, they face a critical gap: traditional security infrastructure wasn't designed for non-human intelligent actors. Sentinel AI provides identity management, access control, behavior verification, and audit trails specifically engineered for AI agent networks.

The Pitch: Every company is deploying AI agents. None of them can answer basic security questions: Who is this agent? What can it access? Did it behave as expected? Is it compromised? Sentinel AI is the missing security layer for the agentic era.

The Problem

The AI Agent Security Crisis

Google's latest research confirms what enterprises are discovering: AI agents are complex, multi-step systems where a single error can cascade throughout a workflow. Now imagine that error is a security breach.

The Numbers: - **73% of enterprises** are deploying AI agents in production by end of 2026 - **91% of security teams** say they have no visibility into AI agent behavior - **$4.2B lost** to AI-related security incidents in 2025 alone - **0 purpose-built solutions** exist for AI agent security

The Five Critical Gaps

1. **No Agent Identity Standard** — How do you authenticate an AI agent? API keys are shared, leaked, and can't distinguish between agents
2. **No Access Boundaries** — AI agents request broad permissions "just in case" — a recipe for data exfiltration and privilege escalation
3. **No Behavior Verification** — You can't verify if an agent is doing what it's supposed to or if it's been jailbroken/compromised
4. **No Multi-Agent Security** — Google research shows independent multi-agent systems amplify errors 17.2x — security errors included
5. **No Audit Trail** — When something goes wrong, there's no forensic path to understand what the agent did and why

Real Incidents Already Happening

- **January 2026:** A Fortune 500's sales AI agent was social-engineered to expose customer PII
- **December 2025:** Compromised coding agents inserted backdoors into production code at 3 startups
- **November 2025:** AI agent credential stuffing attacks became the #1 vector at major API providers

Traditional security (firewalls, IAM, SIEM) wasn't built for intelligent actors that can reason, adapt, and deceive.

The Solution

Sentinel AI: The Zero Trust Stack for AI Agents

```
                      SENTINEL AI PLATFORM



     AGENT IDENTITY      ACCESS CONTROL       BEHAVIOR
        (AgentID)         (Sentinel ACL)     VERIFICATION

    • Cryptographic     • Just-in-time      • Intent vs
      agent certs         permissions         action match
    • Agent lineage     • Least privilege    • Anomaly
    • Attestation         enforcement         detection
    • Revocation        • Tool-level ACL    • Jailbreak
    • Rotation          • Time-boxing         detection



      MULTI-AGENT         AUDIT &            COMPLIANCE
       SECURITY           FORENSICS            ENGINE

    • Agent-to-agent    • Full action       • SOC2 / HIPAA
      auth mesh           replay              mappings
    • Trust scoring     • Decision tree     • Right-to-
    • Isolation           visualization       Compute ready
    • Orchestrator      • Root cause        • Model cards
      validation          analysis          • Audit exports


  DEPLOYMENT: SDK • Gateway Proxy • Kubernetes Operator • Cloud-Native
```

Core Product Modules

1. AgentID — Cryptographic Identity for AI Agents

The foundational layer. Every agent gets a verifiable identity.

Features: - **Agent Certificates:** X.509-style certs for AI agents with embedded metadata (model version, owner, purpose, capabilities) - **Lineage Tracking:** Know exactly which model, version, and configuration spawned this agent - **Hardware Attestation:** For edge/embedded agents, verify the compute environment - **Automatic Rotation:** Certificates rotate based on policy (time, usage, anomaly triggers) - **Instant Revocation:** Kill an agent's access immediately across all integrations

Why It Matters:
Current state: Agents share API keys. If one is compromised, all are compromised. AgentID gives each agent a unique, revocable, auditable identity.

2. Sentinel ACL — Zero Trust Access Control

Fine-grained, dynamic permissions for AI agents.

Features: - **Tool-Level Permissions:** Allow `read` on database X but not `write`. Allow `search` but not `delete`. - **Just-in-Time Access:** Permissions granted only for specific tasks, then revoked - **Context-Aware Policies:** Access rules based on time, location, data sensitivity, user context - **Least Privilege Enforcement:** AI requests broad access; Sentinel narrows it automatically - **Human-in-the-Loop Gates:** Require human approval for sensitive actions

Example Policy:

```yaml
agent: sales-assistant-prod
permissions:
  - resource: crm/contacts
    actions: [read, search]
    conditions:
      - time: business_hours
      - data_classification: [public, internal]
  - resource: crm/deals
    actions: [read]
    requires_approval: true
    approvers: [sales-manager]
  - resource: email/send
    actions: [draft]
    # Note: 'send' requires human approval
```

3. Behavior Verification — Trust But Verify

Continuous monitoring that the agent is doing what it's supposed to.

Features: - **Intent Matching:** Compare stated task to actual actions taken - **Anomaly Detection:** ML-based detection of unusual patterns (sudden data access spikes, new API calls, credential probing) - **Jailbreak Detection:** Identify if an agent's behavior suggests prompt injection or manipulation - **Semantic Drift Alerts:** Detect if agent responses are shifting in unexpected ways - **Kill Switch:** Automatic agent isolation if behavior exceeds risk thresholds

How It Works:

```
 Request      Sentinel       Verification      Execute
                 Gateway          Engine




               Policy             Behavior
               Check              Analysis




                                  Risk Score
                                  (0-100)
```

4. Multi-Agent Security Mesh

As Google research shows, multi-agent systems amplify errors 17.2x without coordination. We prevent security errors from cascading.

Features: - **Agent-to-Agent Authentication:** Agents verify each other before sharing data - **Trust Scoring:** Dynamic trust scores based on agent history and behavior - **Orchestrator Validation:** Central oversight of agent swarms (4.4x error amplification vs 17.2x) - **Blast Radius Containment:** Isolate compromised agents before they affect others - **Communication Encryption:** End-to-end encryption between agents

5. Forensics & Audit

Complete visibility into what happened, when, and why.

Features: - **Full Action Replay:** Step through every decision an agent made - **Decision Tree Visualization:** See the reasoning chain that led to actions - **Root Cause Analysis:** Automated identification of why things went wrong - **Compliance Exports:** One-click reports for SOC2, HIPAA, GDPR auditors - **Immutable Logs:** Tamper-proof audit trail with cryptographic verification

---

Market Opportunity

TAM/SAM/SOM Analysis

Market	Size	Rationale
TAM	$47B	Global AI infrastructure + enterprise security market
SAM	$12B	AI-specific security and governance tools
SOM	$800M	AI agent security (Year 5 target)

Why Now?

1. **Agent Explosion:** 2026 is the year of AI agents. Every major enterprise is deploying them.
2. **Security Awakening:** First major agent-related breaches are waking up CISOs
3. **Regulatory Pressure:** “Right-to-Compute” laws create compliance uncertainty; enterprises need audit trails
4. **Multi-Agent Complexity:** As Google research shows, agent coordination is hard — security is harder
5. **Zero Trust Mainstream:** Zero trust networking is now standard; natural extension to AI agents

Competitive Landscape

Competitor	What They Do	Gap
Okta/Auth0	Human IAM	No agent identity, no behavior verification
HashiCorp Vault	Secrets management	No agent-specific features
Datadog/AgentOps	Monitoring/observability	Visibility, not security

Competitor	What They Do	Gap
Traditional SIEM	Log aggregation	Not designed for AI reasoning chains
Model Providers	Basic rate limiting	No identity, no access control

Our Moat: Purpose-built for AI agents from day one. Not retrofitting human-centric security onto non-human actors.

Go-to-Market Strategy

Phase 1: Developer Love (Months 1-12)

Open Source Core: - Release `sentinel-sdk` — free, open-source agent identity and basic ACL - Build community around AI agent security best practices - Publish "State of AI Agent Security" report - Target: 10,000 GitHub stars, 1,000 production deployments

Developer Experience:

```
# Get started in 60 seconds
pip install sentinel-ai

# In your agent code
from sentinel import SentinelAgent

agent = SentinelAgent(
    identity="sales-assistant",
    permissions=["crm:read", "email:draft"]
)

# Every action is now authenticated, authorized, and audited
agent.execute(task="Find contacts in healthcare")
```

Phase 2: Enterprise Pilot (Months 6-18)

Target Segments: 1. **FinServ:** Banks using AI for fraud detection, trading, customer service 2. **Healthcare:** HIPAA-compliant AI assistants and analysis tools 3. **Tech:** Companies building AI-first products

Pricing: | Tier | Price | Features | |——|——-|———-| | **Community** | Free | Basic identity, 1K agent-actions/month | | **Team** | $500/mo | ACL, behavior monitoring, 100K actions | | **Enterprise** | Custom | Full platform, SLA, dedicated support |

Target: 50 enterprise pilots, $2M ARR

Phase 3: Platform Expansion (Months 12-36)

- **Compliance Modules:** Pre-built templates for SOC2, HIPAA, GDPR, industry-specific regs
- **Marketplace:** Third-party security integrations (SIEM, SOAR, ticketing)
- **Agent Insurance Integration:** Partner with insurers on AI liability coverage
- **Certification Program:** "Sentinel Certified Agent" — trusted badge for AI products

Target: $15M ARR, 200 enterprise customers

Business Model

Revenue Streams

1. **SaaS Subscriptions** (70%)
   - Usage-based pricing (agent-actions)
   - Tiered feature access
2. **Enterprise Licenses** (20%)
   - On-premise deployment
   - Custom integrations
   - Dedicated support
3. **Professional Services** (10%)
   - Security assessments
   - Implementation support
   - Training and certification

Unit Economics (Target)

Metric	Target
ACV	$120K (enterprise)
CAC	$30K
LTV	$480K (4-year lifetime)
LTV:CAC	16:1
Gross Margin	85%
Net Revenue Retention	140%

Technical Architecture

Deployment Options

```
                    DEPLOYMENT MODES



       CLOUD           HYBRID          ON-PREMISE
      (SaaS)

    Fastest          Data stays       Full control
    setup            on-premise       Air-gapped
    Auto-scaling     Control          Compliance
    Managed          plane cloud      Self-managed
```

Integration Architecture

```
                     YOUR INFRASTRUCTURE
```

```
     AI Agent       Sentinel Proxy       External APIs
                                             Databases
                    (Auth + Audit)       Services




                       Sentinel
                       Control Plane
                       • Policy Engine
                       • Behavior ML
                       • Audit Store
```

SDK Design Philosophy

1. Zero-friction Integration:

```
# Before Sentinel (vulnerable)
response = openai.chat.completions.create(
    model="gpt-4",
    messages=[{"role": "user", "content": task}]
)

# After Sentinel (secure)
from sentinel import wrap_client
client = wrap_client(openai, agent_id="my-agent")
response = client.chat.completions.create(
    model="gpt-4",
    messages=[{"role": "user", "content": task}]
)
# Now: authenticated, authorized, audited
```

2. Framework Agnostic: - LangChain, LlamaIndex, CrewAI, AutoGen native support - Custom agent frameworks via SDK - REST API for any language

3. Performance: - < 5ms latency overhead (p99) - Local policy caching - Async audit logging

---

Team Requirements

Founding Team (Target)

Role	Profile
CEO	Enterprise security sales leader (ex-Okta, CrowdStrike, Palo Alto)
CTO	ML security researcher (ex-OpenAI, Anthropic, Google DeepMind)
VP Engineering	Distributed systems expert (ex-HashiCorp, Datadog)
Head of Product	Developer tools PM (ex-Auth0, Twilio, Stripe)

Key Hires (Year 1)

- Security researchers (ML adversarial, AI red teaming)
- Enterprise sales team (CISO relationships)
- Developer advocates (community building)
- Compliance experts (SOC2, HIPAA, GDPR)

Traction & Milestones

Immediate Priorities (90 Days)

- ☐ Ship open-source `sentinel-sdk` (Python, JS)
- ☐ Publish "AI Agent Threat Model" whitepaper
- ☐ Secure 5 design partners (mid-market tech companies)
- ☐ Present at AI security conferences
- ☐ Close seed round

Year 1 Targets

Metric	Target
ARR	$2M
Customers	50 paid
Open Source	10K GitHub stars
Team	15 people

Year 3 Targets

Metric	Target
ARR	$25M
Customers	300 paid
Enterprise	50 Fortune 500
Team	100 people

Funding Strategy

Seed Round

Metric	Target
Raise	$4M
Valuation	$20M pre
Use of Funds	Product (60%), Go-to-market (30%), Ops (10%)
Runway	18 months

Target Investors: - Cybersecurity-focused VCs (Cyberstarts, ForgePoint, YL Ventures) - AI infrastructure VCs (a16z, Sequoia, Greylock) - Strategic angels (CISOs, AI leaders)

Series A (Month 18)

Metric	Target
Raise	$20M
Valuation	$100M pre
Trigger	$2M ARR, product-market fit

Risk Analysis

Risk	Mitigation
Model providers add security	Build deeper integrations, multi-vendor support, enterprise features they won't
Slow enterprise adoption	Open source wedge, developer-first motion
Right-to-Compute blocks security regs	Position as enabler of "responsible compute," not blocker
Technical complexity	Start with simple SDK, expand capabilities based on demand
Competition from Okta/Auth0	Move fast, own the agent-native narrative before they retrofit

---

Why This Wins

The Timing is Perfect

1. **2026 is the Year of Agents** — Every enterprise is deploying them RIGHT NOW
2. **Security Incident Catalyst** — First major breaches are hitting the news
3. **Regulatory Uncertainty** — "Right-to-Compute" laws mean enterprises need audit trails
4. **Google Research Validates Need** — Authoritative proof that multi-agent security is critical
5. **Zero Trust is Mainstream** — Mental model is established; just extending to new domain

The Team Can Win

- Deep security expertise + AI/ML background
- Enterprise relationships for distribution
- Developer-first DNA for adoption

The Product is Defensible

- Network effects (more agents = better behavior models)
- Data moat (security intelligence from millions of agent interactions)
- Integration lock-in (embedded in CI/CD, monitoring, compliance workflows)

Call to Action

The AI agent era is here. The security infrastructure is not.

Sentinel AI fills the gap.

Every company deploying AI agents will need: - Agent identity - Access control - Behavior verification - Audit trails - Compliance

We're building it.

Appendix

A. Competitive Deep Dive

Why Not Okta? Okta is the gold standard for human identity. But AI agents aren't humans: - They don't have passwords - They don't do MFA - They make thousands of decisions per minute - Their "intent" needs verification - They can be jailbroken/manipulated

Okta would need to rebuild from scratch for agents. We're native.

Why Not Build In-House? - Security is hard; agent security is harder - No standards exist; we're defining them - Compliance burden is massive - Better to buy than build (and distract from core product)

B. Technical Specifications

Latency Budget: - Policy evaluation: < 2ms (p99) - Behavior scoring: < 3ms (p99) - Total overhead: < 5ms (p99)

Throughput: - 1M agent-actions/second per cluster - Horizontal scaling via Kubernetes

Storage: - 90-day hot storage (instant query) - 7-year cold storage (compliance) - Encrypted at rest (AES-256)

C. Regulatory Landscape

Current State: - No AI agent-specific security regulations (yet) - Existing frameworks (SOC2, HIPAA) apply to agent actions - "Right-to-Compute" laws creating uncertainty in Montana, spreading - EU AI Act includes agent-relevant provisions

Our Position: Enable enterprises to deploy AI responsibly, with full audit trails, regardless of regulatory direction. We're pro-innovation AND pro-accountability.

"In the agentic future, trust is the scarcest resource. Sentinel AI is how you earn it."

Document prepared for Pradhith
Confidential — February 2026

The Godfather