PRODUCT BRIEF

ARKOSE AGENT TRUST MANAGER

See every agent. Know their intent. Enforce the right response.

For a decade, defenders asked: "Is this a bot or a human?" Every WAF, every device fingerprinting vendor, every bot management platform was built to answer that binary. AI agents broke the binary. Agents act at machine scale with human authorization. The old question no longer produces a useful answer.

GenAI-enabled fraud losses are projected to reach $40B by 2027 — a 30% CAGR. A new class of adversary tooling, what we call Generation 3, runs locally on legitimate hardware. It uses real residential IP addresses, executes with human-plausible timing, handles MFA, retries on failure, and sustains operations indefinitely at no ongoing cost.

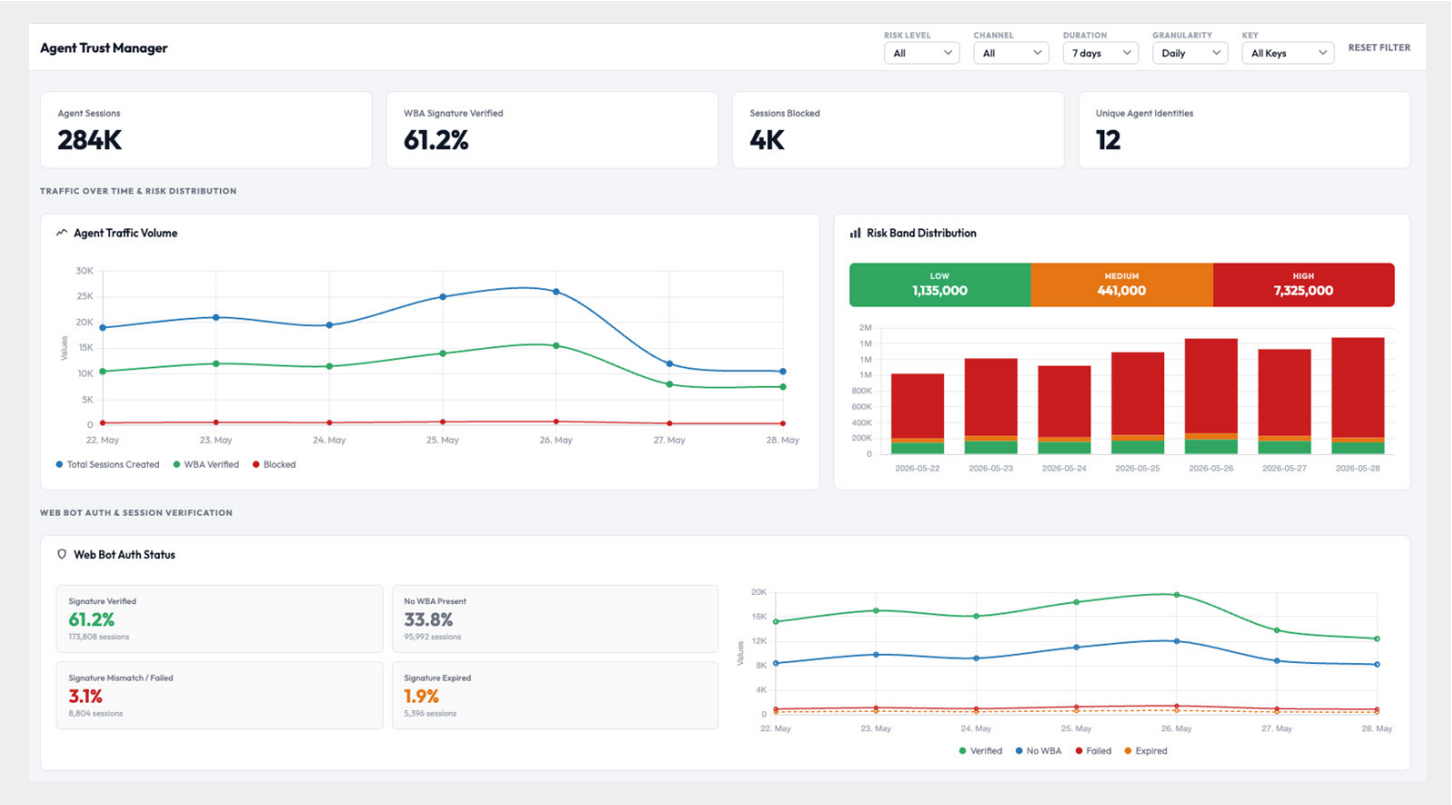
Organizations face a new class of threat that existing defenses were never designed to address. And the new question, “What is the intent of this agent, and should it be trusted?” requires a fundamentally different architecture. Arkose Agent Trust Manager, part of the Arkose Titan platform, is built to answer it.

What Is Arkose Agent Trust Manager?

Arkose Agent Trust Manager is purpose-built for detection, classification, and control of AI agent traffic. Delivered within Arkose Titan, it operates on the same Titan infrastructure.

The product operates through three integrated functions: visibility, classification, and enforcement. Every consumer-facing flow and API endpoint now categorizes traffic into four distinct populations: good users, self-disclosing good agents, non-disclosing good agents, and adversaries. Arkose Agent Trust Manager identifies each and enforces proportionally.

By classifying every agent by intent and applying per-endpoint policy, Arkose Agent Trust Manager gives organizations the confidence to open their platforms to agentic commerce to protect revenue while stopping adversarial agents before they compromise accounts, payments, and platform integrity.



Key Benefits

Achieve full visibility into every agent hitting your flows

Real-time visibility into agent composition across sign-in, sign-up, checkout, and API endpoints, before enforcement decisions are made. This means named attribution: GPTBot, ClaudeBot, Perplexity, Comet, Fellou — not just "bot or not."

Distinguish good agents from bad ones, at machine speed

Three-population classification resolves every agent session as self-disclosing good agent, non-disclosing good agent, or adversary. Intent determines the verdict.

Enforce proportional responses without engineering delays

A five-action enforcement spectrum - allow, monitor, block, throttle and challenge - matches every agent to the right response automatically. Set policy once; enforce it at scale.

Control agent access continuously, without a human in the loop

Customer-defined policy by endpoint, population, and use case. Continuous intent integrity monitoring re-classifies sessions when behavior drifts, blocking threats automatically at the speed of the attack.

Govern every action post-login, not just the gate

A human logs in; an agent executes the money movement. Arkose Agent Trust Manager re-classifies every action across the session lifecycle, so policy applies where the real exposure is.

Protect APIs and MCP servers where browser detection can't reach

As agents call backends directly, JS-based detection becomes structurally absent. Arkose Edge, a component of Arkose Titan, provides server-side detection. No JS tag, no client SDK, with the same agent classification as your browser flows.

Enable agentic actions without fraud exposure

A legitimate AI assistant acting on behalf of a user and an adversarial agent running credential stuffing share the same behavioral signature. Arkose Agent Trust Manager resolves the difference. You no longer have to choose between enabling agentic commerce and increasing revenue or stopping attacks.

How It Works

Arkose Agent Trust Manager operates through three integrated functions: See every agent. Know their intent. Enforce the right response.

VISIBILITY

Real-time view of agent vs. human composition across all critical flows. The agentic AI dashboard surfaces population breakdown and enforcement actions across endpoints. You cannot control what you cannot see.

CLASSIFICATION

Multi-signal evaluation resolves every session into one of four classifications: good user, self-disclosing good agent, non-disclosing good agent, or adversary. Intent detection runs on top; the same session profile receives a different verdict depending on what the agent is doing.

ENFORCEMENT

A multi-action enforcement spectrum - allow, monitor, block, throttle and challenge - matches every session to the proportional response. Customer-defined per-endpoint policy determines which action applies by agent type, intent, and risk level. A price-checking agent on a product page is allowed; the same pattern probing account recovery flows is flagged and blocked. Continuous intent integrity monitoring re-classifies sessions when behavior drifts. Control is not a one-time gate; it is continuous and adaptive.

What Sets Arkose Agent Trust Manager Apart

Delivers Contextualized Intent Intelligence
Goes beyond detecting automation to determine intent. The same behavioral profile receives a different verdict depending on what the agent is doing, enabling legitimate agents while stopping adversarial ones.

Purpose-Built for Generation 3 Agents
Behavioral detection specifically designed for locally hosted, residential-IP agents that bypass IP reputation, cloud ASN detection, and legacy fingerprinting tools.

Runs on the Arkose Titan Platform. No New Integration
Arkose Agent Trust Manager activates on the Arkose Titan platform's shared session infrastructure. Existing Arkose Labs customers are up and running with no new placement, no new integration project.

Recognize Repeat Offenders and Inform Your CIAM Tools
Flag adversarial agents and feed that data into your CIAM system, making it easier to block threats across your full identity stack and stay ahead of coordinated attack campaigns.

Cryptographic Agent Identity via Web Bot Auth
For agents that support it, WBA provides Ed25519 signature verification, cryptographic proof of identity that no spoofing tool can replicate. The scalable mechanism for partner API allowlisting, and the highest-confidence trust signal in the category.

Agent detected and WBA verified

```
"agentic_ai": {
  "detected": true,
  "agent": {
    "name": "ChatGPT",
    "operator": "OpenAI"
  },
  "web_bot_auth": {
    "provided": true,
    "signature_verified": true,
    "signature_fail_reason": null
  }
}
```

Agent detected, no WBA

```
"agentic_ai": {
  "detected": true,
  "agent": { "name": "GPTBot" },
  "web_bot_auth": { "provided": false, "signature_verified": null }
}
```

About the Arkose Titan Platform

Arkose Titan is Arkose Labs' comprehensive platform that delivers end-to-end protection across every touchpoint of the user journey. The platform makes attacks unprofitable while keeping legitimate users moving seamlessly through:

- **Unified Intelligence:** Shared threat data across all touchpoints creates compounding protection where each interaction strengthens the entire system
- **Attack Economics Disruption:** Increases attacker costs exponentially while defender costs remain flat
- **Adaptive Enforcement:** Real-time response that evolves with sophisticated threats including AI-powered attacks
- **Zero-Friction for Legitimate Users:** 98%+ customer satisfaction with invisible protection for real customers

Arkose Titan secures every stage—from first account sign-up through ongoing platform activities—protecting registration, authentication, payments and in-platform interactions with one unified solution.

See Arkose Agent Trust Manager in Action

See how Arkose Agent Trust Manager classifies every agent hitting your flows and enforces by intent, stopping adversarial AI without blocking legitimate activity. Contact us today to schedule your personalized consultation.

BOOK A DEMO

Arkose Labs is the global leader in Agent Trust and Control and helps companies identify whether a site visitor is a customer, AI agent, or adversary. Arkose Titan, its unified platform, provides controls over how they are trusted, challenged, or stopped accordingly. In an era when AI agents are changing the economics of online fraud, Arkose Labs protects against persistent, massive-scale AI-fueled abuse rather than just isolated, bot-driven attacks. Trusted for over 10 years by global financial services, technology, gaming, and travel companies including Anthropic, Meta, Adobe, Roblox, and Expedia, the company has been named to the Deloitte Technology Fast 500 for five consecutive years. Arkose Labs backs Arkose Titan with the industry's only $1 million warranty program. Arkose Labs is headquartered in San Mateo, California, and has research teams around the globe.