Secure every AI interaction.

Defiance AI™ is the **intent engine** for AI. It reads the *why* behind every prompt, response, and agent action in real time, so you can audit, question, and govern AI across your enterprise.

SOLVING AI ADOPTION RISKS

Defend at AI runtime

AI Interaction Security gateway with a built-in intent analysis engine that operates in real time.

Visibility into your AI control plane

Discover and monitor every endpoint, agent, and data path. Eliminate Shadow AI blind spots.

Map agents → users → risk

Mapping risk to every AI asset and to every exchange between a human or autonomous agent and an AI system.

Enforce governance as live controls

Turn NIST, the EU AI Act, ISO 42001, and OWASP into enforcement that runs inline.

What We Stop

- **Prompt Injection**
  Malicious instructions hidden in inputs.
- **Jailbreaks**
  Attempts to override model guardrails.
- **Data & Knowledge Poisoning**
  Tampered training, grounding, retrieval.
- **Sensitive Data Exfiltration**
  Regulated data leaving through AI.
- **Multi-turn Drift**
  Attacks that build across a conversation.
- **Rogue & Over-Permissioned Agents**
  Autonomous actions beyond their scope.
- **Shadow AI**
  Unsanctioned tools and blind spots.
- **Insider Misuse**
  Sanctioned users, unsafe intent.
- **AI Security Risks →**
  See all the AI risks we cover.

> *"This is not classic IT observability into whether something happened but rather contextual visibility into* **why something happened**.*"*
>
> FORRESTER · THE AI CISO, 2026

Current vendors use signatures.
Intent analysis closes the gaps.

Every missing capability is a CISO's blind spot.

CAPABILITY	Defiance AI™	Microsoft Defender, Purview	SASE Zscaler, Netskope, Palo Alto, Cato	Pure-plays Lakera, Prompt Sec, Aim, WitnessAI	OSS NeMo, Guardrails, LLM Guard
'WHAT' happened observability	●	◐	◐	◐	◐
'WHY' happened observability (intent)	●			○	○
Intent graph + denotation resolution	●				
Multi-turn path-risk reasoning	●	○	○	○	
Per-agent identity (IAM for Agents)	●	◐	○	○	
Permission envelope deviation	●				
Forward consequence reasoning	●				
Court-quality signed audit	●	◐	○	○	
Cross-system agent action visibility	●	○	◐		
Tier-3 air-gapped operation	●				○
Bidirectional Brain (boost + RL labels)	●				
Layered architecture (sigs + intent + conseq)	●	○	○	○	

● FULL · ◐ PARTIAL · ○ MINIMAL · BLANK = NOT ADDRESSED

EVERY OTHER TOOL SEES

Surface signals. Post-hoc logs. The what, never the why.

Microsoft Defender / Purview · the request

SASE (Zscaler / Netskope / Palo) · the host

Pure-plays (Lakera / Prompt Sec) · the classifier output

OSS (NeMo / Guardrails) · the SDK boundary

DEFIANCE AI™ SEES

Contextual intent. Every prompt, response, and agent action.

- ✓ The structural intent behind the prompt
- ✓ The evolution of risk over time
- ✓ Intent against policy and alignment
- ✓ The forward consequences of actions

One control plane for every AI interaction.

Defiance AI™ sits inline between everyone and everything using AI. Every prompt, response, and agent action is inspected, scored, and governed in real time.

Chaos

- Untracked agents
- Shadow AI
- Black-box reasoning
- Pinned-cert apps
- Cross-system actions
- Multi-turn drift
- Compliance gaps
- Identity blind spots
- No WHY-visibility

AI INTERACTION GATEWAY

Inspect every prompt & response in real time

01	02	03
Visibility	**Detection**	**Response**
Discover every endpoint & data path	Classify intent & risk	Enforce policy inline

Our self-improving **intent engine** updates detection across every deployment

Control

- ✓ Per-prompt WHY
- ✓ Per-agent identity
- ✓ Path-risk score
- ✓ Permission envelope
- ✓ Court-quality audit
- ✓ Operator dashboards
- ✓ Regulator-ready reports
- ✓ Action-deviation alerts
- ✓ Tier-3 air-gapped ops

AN AI INTERACTION SECURITY STACK

Risk Mapping
Maps agents, users, and AI assets to real business risk.

Agentic Identity
Attaches identity to every agent action, human and machine.

Attack Surface Mgmt
Discovers every AI endpoint, plugin, and data path.

Policy & Governance
Enforces NIST, EU AI Act, ISO 42001 controls inline.

Threat Mitigation
Blocks injection, jailbreaks, exfil, and drift in real time.

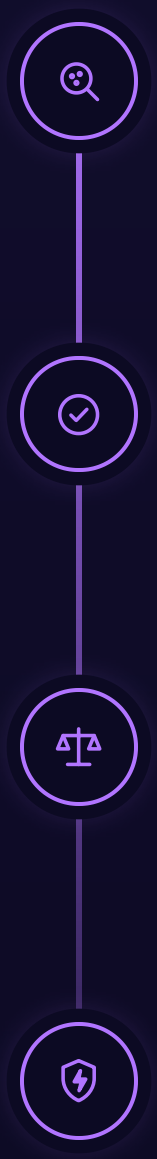
AI Audit Trail
Immutable logging of every prompt and response.

Dashboards & Reporting
Live verdict, stage, and severity distribution.

Expert Support
PurpleSec practitioners staffing deployment and tuning.

One platform. Four modules.

Defiance AI™ is the product suite. Four modules run on the same gateway, taking you from first visibility into AI use through to enforcement across the agentic plane.



Discover

Gain visibility to AI use and independent agentic behaviors. Understand the 'what', 'who' and 'where' of token use to uncover unseen risks.

See what agents are up to! connecting agentic activity to how user prompts are enacted downstream into system actions.

Verify

Proactively listen to AI and receive distillations on usage: both good and bad. Verify how agentic usage is happening to evolve policies and guide business efficiencies.

Understand the real risks the agentic plane introduces to your organization. See "under the hood" to how malicious AI activity enters and infiltrates.

Govern

Turn governance policy into clear enforceable controls. Govern which AI tools are used and how they are permitted to run, throughout the agentic landscape.

A collaborative and living approach to development of AI policy means your organization gets ongoing support and real time informed governance that keeps up with the blazing pace of AI change.

Enforce

Detect and stop threats in real time. Inspect what happened, map full impact and reach, and deploy containment actions on malicious activity before it spreads.

Secure the entire agentic plane, across all AI interactions throughout your organization. Ensures secure operations for models, agents, users, endpoints, identities and assets, with technology that only gets smarter and keeps your policy enforced.

Start with total transparency.
Evolve to full enforcement.

One engine. Three deployment tiers. Deploy where your data and risk live, from endpoint to air-gap.

TIER 1
AI Security Edge (AISE)

Endpoint EXE or SASE plug-in

BEST FOR

Deploying standardized models (Claude, OpenAI, Gemini) for chatbot/co-work, or augmenting business processes with agentic workflows.

TIER 2
Enterprise

Gateway on your network stack

BEST FOR

Mature security environments using AI to process sensitive data, building or fine-tuning their own models, or with customer-facing AI in regulated industries.

TIER 3
NonPublic

Fully air-gapped, one-way CTI

BEST FOR

Running AI in air-gapped environments where cloud LLMs are legally off-limits (IC, DoD, CMMC, FedRAMP-High, sovereign-cloud, OT, and classified research).

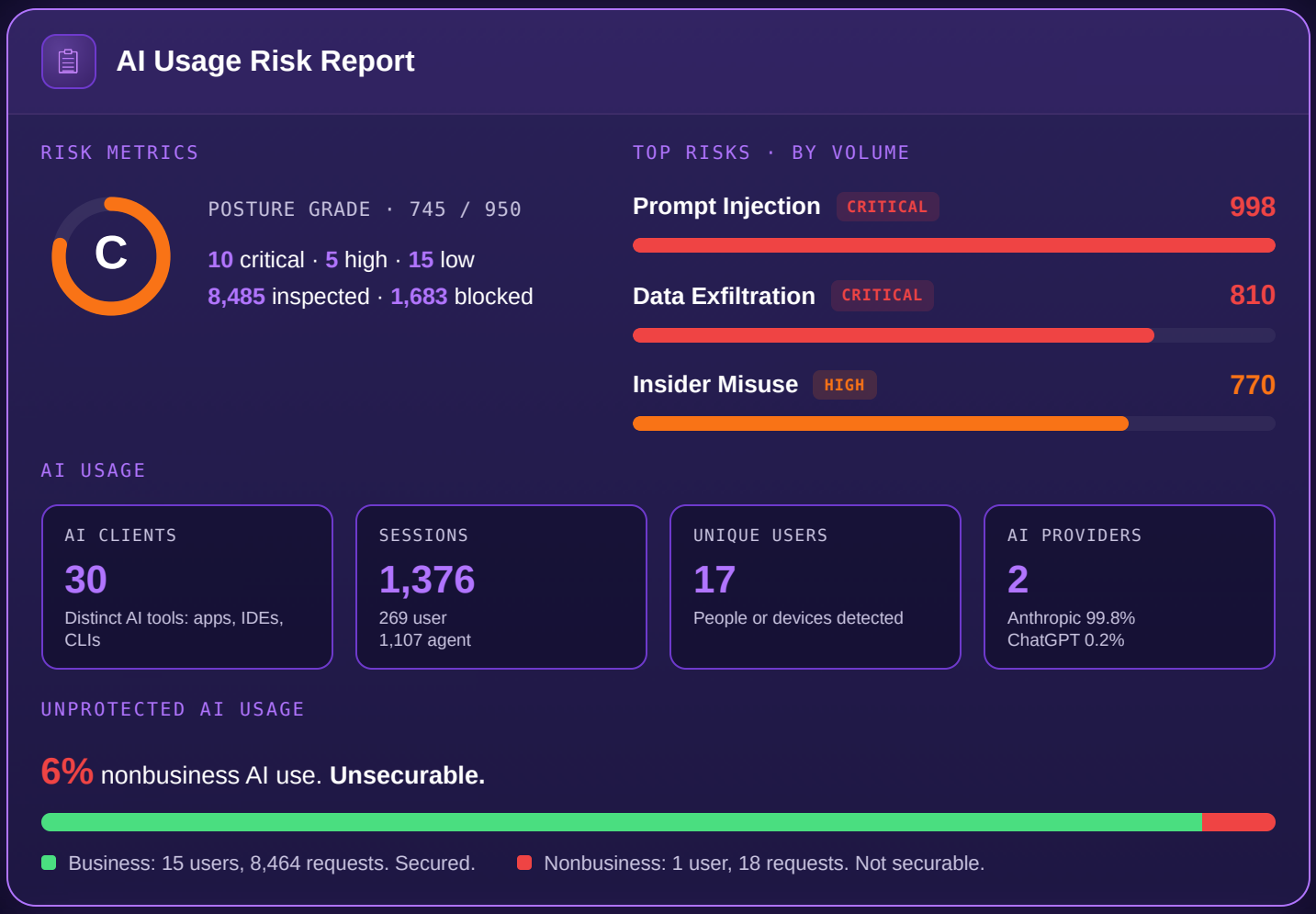
WHAT SHIPS AT EACH TIER

CAPABILITY	AISE	ENTERPRISE	NONPUBLIC
Endpoint EXE client (SSE-style)	●	●	—
SASE plug-in mode	●	●	—
Comprehensive AI client coverage	●	●	●
Edge appliance + on-board agents	●	●	●
SPM · ASM · TSM posture	●	●	●
Risk mgmt (NIST AI RMF / OWASP)	●	●	●
IAM for Agents (Entra Agent365)	●	●	●
Court-quality signed audit	●	●	●
Inline blocking	OPTIONAL	●	●
Attach to customer SASE / firewall	—	●	—
Customer-owned data plane	OPTIONAL	●	●
BYOK / customer key custody	—	●	●
Zero external dependencies	—	—	●
On-prem model serving	—	OPTIONAL	●
FIPS 140-3 validated crypto	—	OPTIONAL	●
Hardware-attested boot (TPM)	—	—	●
One-way CTI mirror (SIPRNET/WSUS)	—	—	●

● STANDARD · OPTIONAL = ADD-ON SKU · — NOT AVAILABLE

Your AI usage has blind spots. We will find them.

Point your AI traffic to Defiance AI™. It monitors every interaction, surfaces the injection and exfiltration your stack misses, names every agent, and grades your posture in a single report.



Start Your Trial →