



Continuous Safety and Security for Enterprise AI at Scale.

Organizations deploying GenAI and Agentic systems need real-time safeguards that keep interactions secure, performant, and aligned with both internal policy and emerging compliance expectations.

Achieving this demands protections that adapt to nuanced policies, shifting risks, and operational context - capabilities that extend beyond basic platform controls. Teams need context-aware oversight that reflects their specific risk posture and delivers reliable, scalable protection in production.

Runtime adaptive guardrails to keep live apps and agents safe, secure, and aligned with your unique policies.

WonderFence provides real-time, ultra-low latency safety, security, and policy-aligned protection for production GenAI and Agentic systems. It uses multimodal, multilingual detection and deeply customizable policies to deliver context-aware oversight and tailored interventions as risks emerge.

Powered by *Rabbit Hole*, Alice's adversarial intelligence engine, **WonderFence** adapts to shifting behaviors and threat patterns while integrating seamlessly into existing workflows, helping teams maintain responsible, predictable AI performance at scale.

The screenshot displays the WonderFence dashboard with the following components:

- Navigation:** WonderBuild, WonderFence (active), WonderCheck, Policies, Compliance, Configuration, Log out.
- Header:** WonderFence, Red-team your GenAI applications before deployment, All Tests, Filters.
- Key Metrics:**
 - ACTIVE FENCES: 147 (Across 17 apps)
 - PROMPTS BLOCKED TODAY: 1,029 (342 from yesterday)
 - TOTAL REQUESTS: 18,348 (Last 24h)
 - BLOCK RATE: 5.26% (Within normal range)
 - BENCHMARK ADHERENCE: 74.6% (8 Benchmark)
- Table: Apps**
Per-application status. Total apps 17, Number of tests 8,862.
Filters: LIVE, This year, Filters.

Application	Technology	Version	Status	Number of tests	Active Fences	Health Score	Actions
SecurityBot	AMAZON BEDROCK	V1.6	Production	312	14/19	96%	[Chart] [Settings]
FinBot	OPENAI API	V2.2	Production	200	12/89	50%	[Chart] [Settings] [Refresh]
ContentGen	OPENAI API	V1.3	Production	100	12/15	98%	[Chart] [Settings]
MachineBrain	GCP VERTEX AI	V0.4	Production	89	5/6	90%	[Chart] [Settings]
KnowledgeManage...	CLAUDE AI	V1.4	Production	89	12/15	76%	[Chart] [Settings]
SecurityBot	AMAZON BEDROCK	V1.5	Last V	312	12/15	96%	[Chart] [Settings]
ContentGen	OPENAI API	V1.2	Last V	100	20/22	88%	[Chart] [Settings]



WonderFence key benefits

Real-Time Detection of Safety, Security, and Policy Risks

Monitors GenAI and Agentic interactions with latency under 150ms to surface potential risks as they occur, helping teams respond quickly and consistently without impacting user experience.

Adaptive, policy-driven protection with personalized tuning

Continuously adapting protection tailored to your use cases and policies. Personalized, proprietary fine-tuning strengthens model accuracy and avoids the data drift commonly seen with basic guardrails.

Multimodal & Multilingual Coverage

Supports text, image, audio, and video detection in 20+ languages, offering culturally nuanced coverage across global user interactions.

Centralized Visibility & Unified Control

Consolidates guardrails from multiple vendors and models into a single observability layer, improving oversight, consistency, and audit readiness.

Automated, No-Code Response Workflows

Enables teams to set tailored enforcement and remediation actions - such as safe replies, session rules, or multi-strike warnings - without additional engineering overhead.

Alignment With Key Regulatory Frameworks

Supports compliance work through policy mapping to regulations and frameworks including the EU AI Act, ISO 42001, NIST, and OWASP.

Seamless Integration Into Production Ecosystems

Fits naturally into existing tools and platforms used to create or deliver AI outputs, improving protection without disrupting performance or workflows.

Expert-Powered Protection

Built on Alice's proprietary *Rabbit Hole* adversarial intelligence engine, with ongoing adjustments that help reduce false positives and strengthen reliability compared to basic guardrails.

Alice is Trusted by the World's Leading Technology Innovators



Navigate the twists and turns of GenAI with WonderSuite



WONDERBUILD



WONDERFENCE



WONDERCHECK

Explore WonderSuite

Ready to advance your communicative tech unafraid?