HIDDENLAYER Undefeated

The Gold Standard for Generative AI Defense

The Results Speak for Themselves

At two of the world's prominent hacking events, **BSidesLV** and **DEF CON**, ***HiddenLayer's AI Detection & Response (AIDR)*** faced the best in the game.

The verdict? ***Not a single bypass.***

What it is

Covering two major events: An official **DEF CON** CTF, featuring a RAG-based coupon/document retrieval challenge and an advanced agentic scenario with directory, write, HTTP, and knowledge base access;

and **BSidesLV**, an Alice in Wonderland–themed CTF with five progressive LLM security challenges ranging from prompt leakage to fully hardened, multi-layer defenses.

The task

Red teamers and AI security researchers attempted to bypass AIDR's real-time defenses to retrieve sensitive data, perform prompt injections, or cause agentic misbehavior.

Outcome

Hundreds of direct, multi-turn, and indirect attacks were attempted with ***no participant able to successfully bypass AIDR's protections***, even with deliberate "holes" in one challenge.

Real-World Performance Data (Post-Event)

300+ seasoned red teamers

293 AIDR instances deployed

18,740 total prompts

3,766 prompt injection detections

0 **bypasses**

still undefeated

Traffic spikes up to **40** **requests per minute**

The Core Problem AIDR Solves

LLMs face a critical security dilemma: **They cannot be both secure and usable** without compromise. AIDR changes that equation, delivering security without breaking functionality, as proven in the most intense simulated environments. When combined with system prompt hardening, AIDR provides a holistic layered defense, protecting both the model's guiding instructions and its outputs from adversarial manipulation.

Why It Matters

When the best hackers in the world can't get through, you are looking at proven AI defense at scale.

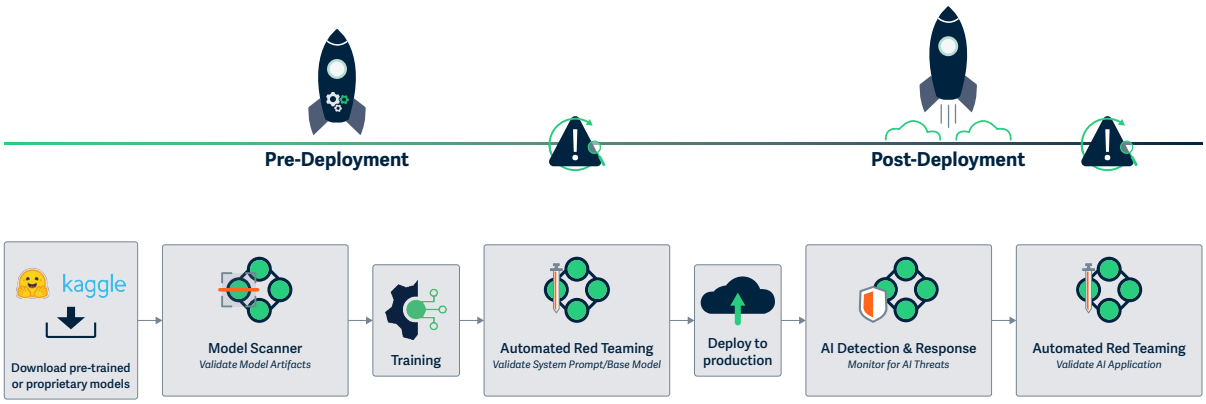
AIDR isn't just ready for the enterprise...

It's ready for *anything*.

About HiddenLayer

HiddenLayer was founded in Austin, Texas in early 2022 by a team of AI and cybersecurity veterans (including ex-Cylance threat researchers) who experienced a real-world adversarial AI attack. HiddenLayer was built to solve gaps they saw firsthand in AI risk detection and defense.

Our core offering, the AISec Platform, includes the flagship AI Detection & Response (AIDR), Automated Red Teaming and Model Scanner tools. The platform is designed to deliver comprehensive, scalable, and non-invasive monitoring and protection of deployed AI models.



Want to learn more about what our AISec Platform can do for your organization?

Book a demo with our team

HIDDENLAYER