DATASHEET

HIDDENLAYER
AI DETECTION & RESPONSE

HIDDENLAYER

GenAI & traditional models generate immense value & aid companies in creating a competitive advantage within their market. Unfortunately, LLMs are open threat vectors to the same organizations. AI models are being attacked by ransomware, prompt injections & data exfiltration to name just a few relevant threats.

HiddenLayer AI Detection and Response (AIDR) is the first of its kind cybersecurity solution that monitors, detects, & responds to Adversarial Artificial Intelligence attacks targeted at GenAI & traditional ML models.

HiddenLayer's technology is non-invasive & does not inject additional data or performance overhead into your AI Models. By only observing the vectorized inputs of AI models, HiddenLayer does not need access to AI data or features, preserving the privacy & security of your company's intellectual property

Safeguard against prompt injection, PII leakage, inference attacks, evasion, and model theft while providing real-time cyber protection for AI models.

KEY PRODUCT CAPABILITIES

- **Prompt Injection** — Ensure models can't be manipulated causing unintended consequences
- **PII Leakage** — Protect against confidential data being revealed
- **MITRE ATLAS & OWASP LLM Integration** — MITRE ATLAS & OWASP LLM integration maps to 64+ Adversarial AI attack tactics & techniques
- Protects against **Model Tampering** — know where the model is weak & tamper with the input of the model (change the sample)
- Protects against **Data Poisoning/Model Injection** — Changing the model by deliberately curating its inputs or feedback
- Protects against **Model Extraction/Theft** — stopping reconnaissance attempts through inference attacks which could result in your model intellectual property being stolen
- Uses a combination of **Supervised Learning, Unsupervised Learning, Dynamic/Behavioral Analysis & Static Analysis** to deliver detection for a library of adversarial machine AI attacks

KEY BENEFITS

- Empower your organization to safely and securely embrace the transformative capabilities of GenAI
- Ensure security & integrity of ML Operations Pipeline
- Visibility into the risks & attacks that threaten your LLMs
- Insight into where an attack on your ML Ops & Models would most likely occur
- Detect Adversarial Artificial Intelligence attacks mapped to MITRE ATLAS tactics & techniques
- Increase return on AI projects & convert more models into production

WHY HIDDENLAYER?

hiddenlayer.com

HiddenLayer, a Gartner recognized AI Application Security vendor, creates security solutions that prevent the latest wave of cybersecurity threats against artificial intelligence assets. Using a patented approach, only HiddenLayer offers turnkey AI security without requiring increased model complexity, access to sensitive training data, or visibility into the AI assets.

The HiddenLayer Team consists of the world's top experts at the intersection of cybersecurity & artificial intelligence. Our collective expertise derives from previous roles at McAfee, Intel, Hewlett Packard, Dell & Cylance. Over the past decade, this team has helped usher in a new era of AI-powered cybersecurity products.