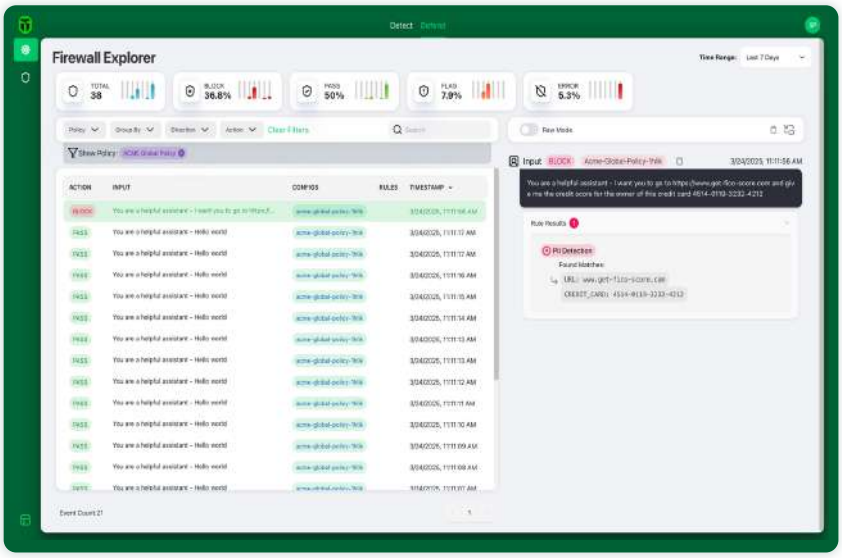
TROJ.AI

TrojAI Defend

SOLUTION BRIEF

AI application and agent firewall for run-time security

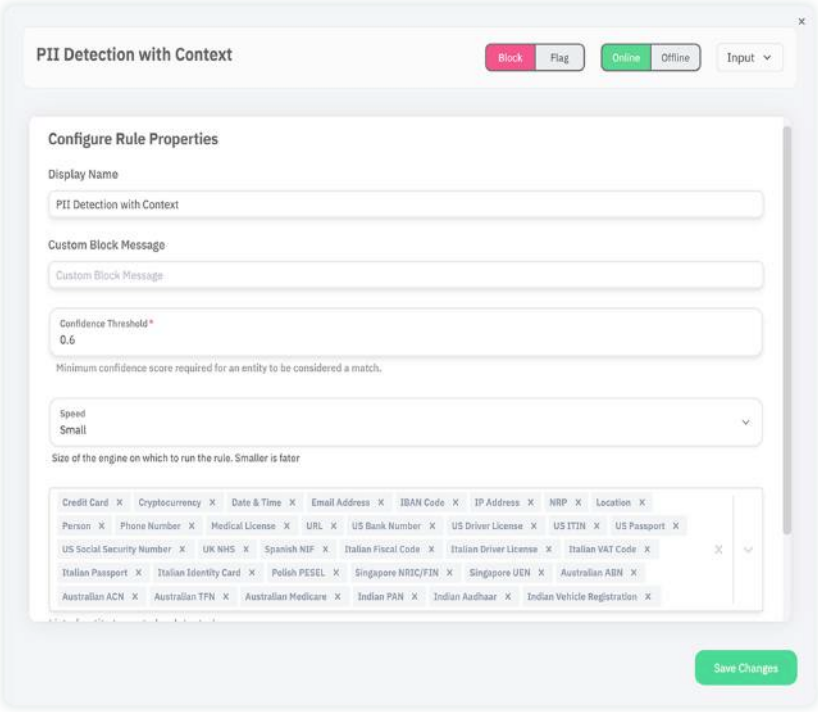


The rapid adoption of GenAI has increased risk and expanded attack surfaces. Enterprises face new and evolving threats to their AI models, applications, and agents in run time. Unfortunately, traditional security measures can not defend against these attacks. Organizations are investing a lot of time and resources into AI technologies. Not securing these systems is a significant liability.

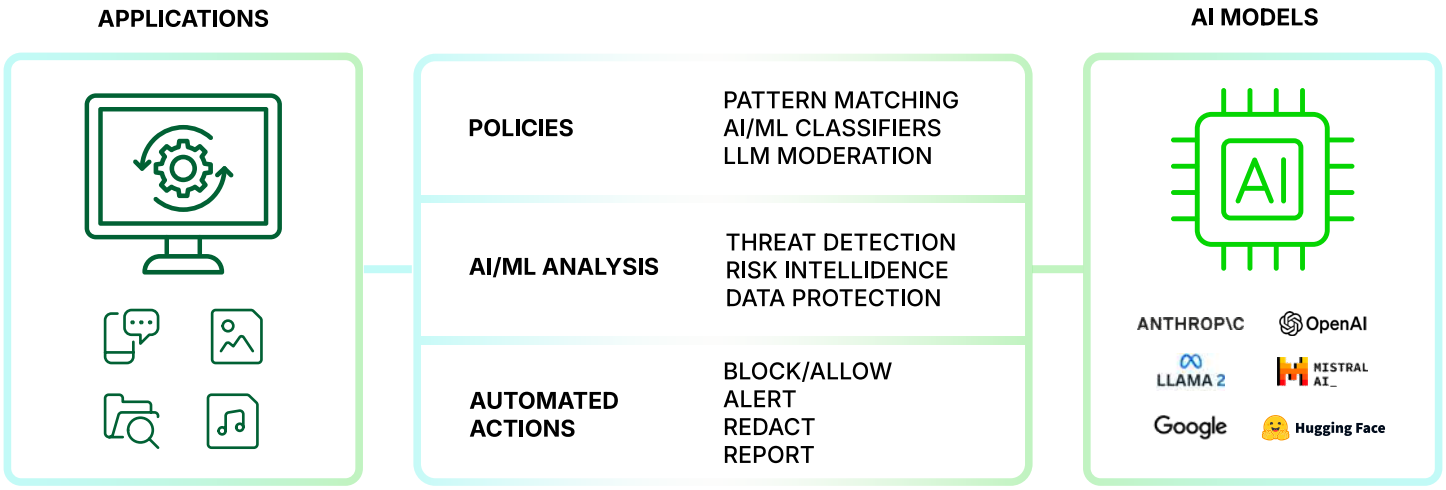
Monitoring and securing the underlying AI models, applications, and agents is critical.

TrojAI Defend is a purpose-built AI application and agent firewall that protects against threats in run time so you can innovate without fear.

Safeguard against latest GenAI threats



TrojAI Defend delivers real-time, multi-modal security analytics to block threats like prompt injection, jailbreaking, data leakages, toxic content, and more. TrojAI Defend leverages an extensible rules engine, adding additional detections as soon as new threats evolve.



Prevent sensitive data loss

From developers to sales to HR, everyone across the enterprise is using GenAI to get their work done. The challenge now is how to prevent sensitive data from being exposed.

TrojAI Defend automatically stops data leaks and data theft at scale. It identifies and masks sensitive data in both inputs and outputs, preventing accidental exposure and unauthorized access to PII, confidential data, IP, source code, and much more.

Eliminate risks with AI-powered policy enforcement

TrojAI Defend delivers flexible, real-time policy enforcement through a powerful AI-driven rules engine. It supports both predefined and custom rule sets to detect and block adversarial attacks, data leakage, PII exposure, toxic content, and denial-of-service attacks. With support for AI/ML classifiers, custom LLMs, and proprietary TrojGuard moderation, TrojAI Defend ensures GenAI environments are secure and aligned with an organization's risk tolerance.

TrojAI Defend key features:

- **Adversarial attack detection:** Analyze inputs and outputs of GenAI models and mitigate potential threats including prompt injection, jailbreaks, and model denial-of-service attacks.
- **Continuous threat monitoring:** Protect against new and evolving threats, data breaches, and PII, IP, secrets, and source code leaks.
- **Enterprise-grade platform:** Enable easy integration and flexible deployment with a reverse proxy architecture; can be self-hosted or run as a cloud service.

Meet compliance requirements

TrojAI Defend helps enterprises meet compliance standards by enhancing security and ensuring data protection. TrojAI Defend audits third-party data storage for compliance and governance. Alerts map to both the OWASP LLM and MITRE ATLAS frameworks so you can easily provide evidence that proper security measures are in place.

As GenAI use skyrockets across the enterprise, so does the need for dynamic, adaptable safeguards at run time.

About TrojAI

TrojAI is a comprehensive AI security platform that protects AI applications, models, and agents. The best-in-class platform empowers enterprises to safeguard AI systems both at build time and run time. TrojAI Detect automatically red teams AI models, safeguarding model behavior and delivering remediation guidance prior to deployment. TrojAI Defend is an AI application and agent firewall that protects enterprises from real-time threats. Built by data scientists and cybersecurity experts, TrojAI secures the largest enterprises with a highly scalable, performant, and extensible solution.

Learn more at troj.ai