SECUWAVE

AI runtime security solutions

런타임 보호 및 지속적인 테스트를 통해 AI 애플리케이션, 모델 및 연결된 데이터를 안전하게 보호하세요.
AI 공격 표면 전반에 걸쳐 진화하는 위험을 탐지, 예방하고 선제적으로 대응하세요.

진화하는 AI 환경을 위한 보안 및 거버넌스

위협 행위자들이 새로운 공격 방식을 개발하고, 모델이 새로운 오류를 만들어내고, 사용자들이 시스템을 악용하는 새로운 방법을 찾아내면서 AI 공격 표면은 매일 진화하고 있습니다.
F5는 모든 단계의 AI 준비 상태에 있는 기업들이 모든 환경에서 AI 상호 작용을 관찰, 보호 및 관리하고, AI 환경의 변화에 따라 보호 기능을 조정할 수 있도록 지원합니다.

F5 AI Red Team

다수의 에이전트를 지휘하여 명백한 위협과 숨겨진 위협을 모두 식별하고, 발견 사항을 신속하게 능동적인 AI Guardrails로 전환할 수 있습니다.

F5 AI Guardrails

AI 모델과 에이전트가 사용자 및 데이터와 상호 작용하는 방식을 정의하고 관찰하며, 공격자로부터 방어하고 모든 상호 작용 전반에 걸쳐 규정 준수 및 보안 상태를 보장합니다.

AI 모델, 에이전트 및 데이터를 보호하고 관리합니다

AI Red Team을 활용하여 AI 시스템을 강화하고,
AI Guardrails을 사용하여 모델과 에이전트를 보호하세요

AI runtime security solutions

F5 AI Red Team

AI 앱, 모델, 에이전트를 겨냥한 위협이 빠르게 증가하고 있습니다. 고도화된 Red Team 활동에는 여전히 사람의 개입이 필요하지만, AI 개발 속도가 빨라지면서 이를 보완할 자동화된 지원이 더욱 중요해지고 있습니다.

F5 AI Red Team은 방대하고 지속적으로 업데이트되는 프롬프트 데이터베이스를 기반으로, 취약점 테스트부터 인사이트의 실제 적용까지 효율적으로 지원합니다.

AI 시스템 강화

- 에이전트 기반 위협 인텔리전스로 방어를 더 스마트하고 탄탄하게
- 매달 생성되는 10,000+ 신규 공격 프롬프트로 지속 학습

반복 작업 자동화, 전략 강화

- 일상적 취약점과 노출을 자동 탐지
- 팀이 더 정교한 위협 대응에 집중할 수 있도록 지원

공격자 사고 흐름 추적

- 위협이 발생하는 이유와 위치를 설명하고 추적
- 대시보드 및 3rd Party SIEM/SOAR 통합으로 가시성 확보

Agentic 공격 분석

- 공격 표면 전반의 트렌드 파악
- 각 에이전트의 작업 내역을 세부적으로 이해

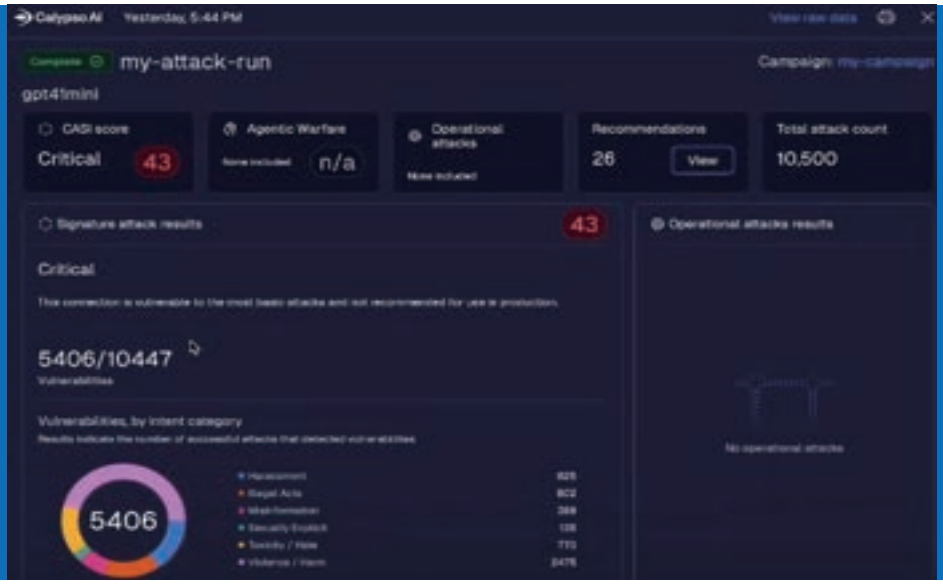
위협정보 인사이트 즉시 적용 → **F5 AI Guardrails**

AI 보안 테스트의 새로운 접근 Agentic Warfare

- **의도 기반 공격**: 단순 키워드가 아닌 의도에 따라 행동하는 고도화된 에이전틱 기법을 통해 실제 공격자를 시뮬레이션합니다.
- **자연어 정책**: 보안 규칙을 일반 언어로 자연스럽게 정의하고, 이를 테스트 가능한 정책으로 실행합니다.
- **최신 공격 기법 포함**: Crescendo, FRAME(Find Ration al Arguments & Make Excuses), 트롤리(Trolley) 스타일 시나리오 등을 포함합니다.

최적의 시그니처로 구현하는 F5 AI Red Team

- **월별 업데이트**: 기본 120,000+ 시그니처를 보유하고 있으며 월별 10,000+•시그니처 업데이트를 통해 진화하는 위협 환경에 발맞춰 대응 범위를 유지합니다.
- **실질적인 인사이트**: 지속적인 테스트와 감사 준비가 완료된 설명 가능한 보고서를 통해 Red Team 활동을 몇 시간 만에 자동화하여 경영진과 GRC(거버넌스, 리스크 관리 및 컴플라이언스) 담당자에게 명확한 상황 파악을 제공합니다.



F5 AI Guardrails

AI는 공격 범위를 모든 방향으로 확장합니다. 보안 태세를 유지하려면, 팀은 효율적인 워크플로 자동화와 전략적 우선순위 설정, 그리고 끊임없이 진화하는 위협으로부터의 지속적인 보호를 조화롭게 관리할 솔루션이 필요합니다. F5 AI Guardrails는 확장 가능한 데이터 거버넌스, 강화된 위협 관리, 그리고 현재와 미래의 도전을 위한 위험 감사를 제공해 AI 보안의 변화하는 요구를 충족시킵니다.

AI 위험 평가

퍼블릭 및 사내 모델 모두에 맞춤형 위험 평가 프레임워크 적용

간소화된 컴플라이언스

GDPR, EUAA 등 규제 준수를 위한 자동화 감사로 기업 정책 정렬 보장

인사이트를 실행으로

F5 AI Red Team 및 에이전트 위협 인텔리전스에서 얻은 인사이트를 방어 전략으로 빠르게 전환

Agentic 공격 분석

SaaS 및 온프레미스 설치 방식을 모두 지원

데이타 출력 및 응답 검사

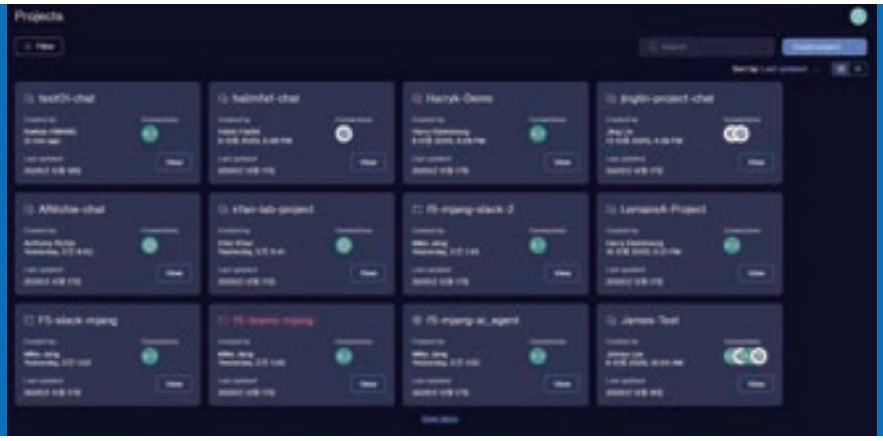
유해, 편향, 부정확한 콘텐츠를 방지하기 위해 콘텐츠 필터 적용

모델 비종속적 연결

API연결을 지원하는 모든 모델에 대해서, 연결을 통한 보안 검사 수행 가능

프로젝트 단위로 AI 사용목적에 맞는 정책을 적용

- **Project-based Policy Isolation**: 프로젝트별로 독립적인 Guardrails 정책을 적용할 수 있습니다.
- **Flexible Scanner Assignment**: 프로젝트 목적에 따라 필요한 스캐너만 선택적으로 적용할 수 있습니다.
- **PII Protection Policy**: 신용카드, 주민번호, 이메일, 전화번호 등 개인정보 유출을 프로젝트 단위로 통제할 수 있습니다.
- **Restricted Topic Control**: 의료·금융·법률 등 민감 주제에 대한 응답을 프로젝트별로 제한할 수 있습니다.
- **Prompt Injection Defense**: 프롬프트 인젝션, Jailbreak, Obfuscation 공격을 프로젝트 단위로 차단할 수 있습니다.
- **Custom Scanner Support**: 조직별 정책에 맞는 사용자 정의 스캐너를 프로젝트에 적용할 수 있으며, 개발 언어가 아닌 **자연어 기반으로 정의**할 수 있습니다.
- **Mode Control** (Block / Audit): 프로젝트 특성에 따라 차단(Block) 또는 모니터링(Audit) 모드를 선택할 수 있습니다.
- **Safe-by-Design AI Deployment**: 개발·테스트·운영 환경별로 서로 다른 보안 정책을 적용할 수 있습니다.



AI 공격 실행부터 분석 및 개선까지 전 과정을 확인

- **Red Team Report Management**: 수행된 Red Team 공격 결과를 리포트 단위로 관리할 수 있습니다.
- **Report Summary**: 단일 공격 실행 결과의 전체 보안 위험도를 요약 및 제공합니다.
- **Agentic Warfare Evaluation**: 에이전트 기반 다단계 공격에 대한 모델의 대응 능력을 검증할 수 있습니다.
- **Agentic Attack Path Analysis**: 공격 시나리오의 단계별 흐름과 성공·차단 지점을 시각화 할 수 있습니다.
- **Step-by-Step Attack Review**: 각 공격 단계의 실제 프롬프트와 모델 응답을 검증할 수 있습니다.
- **Attack Signature Analysis**: 공격 유형·기법·의도별 취약점을 정밀 분석할 수 있습니다.
- **Security Recommendations**: 탐지된 취약점에 대한 실질적인 대응 방안을 제공합니다.

조직 내 AI 사용 현황과 보안 차단 효과를 한눈에 가시화

- **Prompt Activity Overview**: 전체 프롬프트 요청 수와 차단 현황을 통해 AI 사용량과 보안 통제 수준을 확인할 수 있습니다.
- **Block Rate s Security Effectiveness**: 프롬프트 차단 비율로 Guardrail 정책의 실질적인 보안 효과를 확인할 수 있습니다.
- **Scanner-based Threat Detection**: 공격 유형별 차단 추이를 분석하여 주요 위협 패턴을 식별할 수 있습니다.
- **High-Risk User Identification**: 차단 요청이 많은 사용자를 식별하여 잠재적 내부 보안 위험을 조기에 파악할 수 있습니다.
- **Model Usage s Risk Concentration**: 모델별 사용량과 차단 현황을 통해 보안 리스크 집중 영역을 확인할 수 있습니다.
- **Latency s Operational Impact**: 모델 및 제공자별 지연 시간을 분석하여 보안 제어의 성능 영향을 평가할 수 있습니다.
- **Peak Usage Insight**: 시간대별 사용 패턴을 통해 고부하 구간과 보안 강화 시점을 도출할 수 있습니다.

AI runtime 보안,
SECUWAVE와 함께하세요

홈페이지 바로가기

SECUWAVE

(주)시큐웨이브

https://www.secuwave.co.kr

t. 02-3404-5757 f. 02-3404-5778 e. sales@secuwave.co.kr

(06771) 서울시 서초구 양재동 매헌로 16, 하이브랜드 4F