AI 보안:
귀사가 AI 경쟁에서 우위를 유지하는 방법

위험을 관리하면서 AI 도입을 가속화하기 위한 CxO 가이드

서론

조직들이 AI의 혁신적인 힘을 활용하기 위해 앞다투어 움직이는 가운데, 그 도입 속도가 보안 대비를 앞질러가는 경우가 많습니다. 공식적으로 승인된 AI 프로젝트 한 건마다 개인이나 팀 단위로 이루어지는 수많은 비공식 섀도우 AI 사례들이 존재합니다. 실제로 IT 의사 결정자의 85%는 직원들이 IT 팀이 AI 도구를 평가하기도 전에 먼저 도입하고 있다고 말합니다.[1] 그리고 직원의 93%가 승인 없이 AI 도구에 정보를 입력했다고 인정합니다.[1]

한편, 기존 보안 도구가 이를 따라잡는 데 어려움을 겪으면서 AI 네이티브 공격 벡터가 급증하고 있으며, 이미 작년 대비 47% 증가했습니다.[2] 새로운 형태의 데이터 노출과 규제 준수의 공백이 기존의 거버넌스 체계에 도전을 가하고 있습니다. 경영진은 중요한 문제에 직면해 있습니다. AI가 가져다주는 혁신을 저해하지 않으면서, AI로 인해 발생하는 위험을 어떻게 관리해야 할까요?

요구 사항은 명확합니다. 조직은 다음과 같은 환경을 조성해야 합니다.

- 모든 AI 사용이 파악되는 환경
- 모든 AI 통신이 안전하게 보호되는 환경
- 모든 AI 정책이 시행될 수 있는 환경
- 모든 AI 모델이 악용으로부터 보호되는 환경

Cloudflare AI Security Suite는 조직이 위험에 대한 불확실성을 제거하여 AI를 통해 더 빠르게 혁신할 수 있다는 자신감을 갖도록 해줍니다. Cloudflare의 통합 플랫폼은 AI 사용 현황을 파악하고, Zero Trust 액세스를 적용하며, 사전 예방적 위협 방어를 통해 웹 및 API 엔드포인트를 보호함으로써 AI 수명주기 전반을 안전하게 지켜줍니다. 이 통합 데이터 거버넌스를 통해 팀은 보안을 저해하지 않고 자유롭게 혁신할 수 있습니다.

85%
직원들이 IT 팀이 AI 도구를 평가하기도 전에 해당 도구를 도입하고 있다고 말한 IT 의사 결정권자 비율.[1]

AI 보안 문제

AI 도입의 위험성으로 인해 AI 보안 솔루션 시장이 빠르게 성장하고 있습니다. 예를 들어, 클라우드 네이티브 애플리케이션 보호 플랫폼(CNAPP)은 AI 개발 워크플로우에 집중해 왔습니다. 이러한 초기 솔루션들도 필수적인 방어선이지만, 그것만으로는 충분하지 않습니다. 기업은 AI 시스템이 실제 운영 환경에서 가동된 이후까지 포함하여, AI의 전체 수명주기를 안전하게 보호할 필요가 있습니다.

AI 보안 과제의 전체 범위가 점점 명확하게 드러나고 있습니다. 오늘날, 개발자들은 AI 기능을 개발하고, 직원들은 외부 AI 도구를 사용하며, 고객들은 AI 기반 애플리케이션과 상호 작용하고 있습니다. 이러한 다양한 환경을 수동으로 보호하는 것은 복잡하며, 일관성 있게 처리하는 것은 훨씬 더 어렵습니다.

직장 내 AI 오용 방지

조직 인력이 승인된 AI 도구와 미승인 AI 도구를 모두 도입함에 따라, 민감한 데이터와 운영이 위험에 노출됩니다. 보안 팀은 다음 사항을 처리해야 합니다.

- **AI 사용에 대한 가시성 부족:** 리더는 직원들이 어떤 AI 도구를 사용하는지, 어디에서 민감한 데이터가 처리되는지, 그리고 이러한 AI 시스템이 비즈니스 애플리케이션에 어떻게 연결되는지에 대한 완전한 그림을 파악하지 못하는 경우가 많습니다. 이러한 사각지대는 상당한 위험 노출을 야기합니다.
- **데이터 보안 및 규제 준수 위험:** AI는 조직 내 데이터 흐름 방식을 변화시킵니다. 개인 정보, 독점 데이터 및 고객 기록은 규제 준수 위반을 야기하거나 경쟁 정보를 노출하는 방식으로 AI 시스템에 저장될 수 있습니다.
- **에이전틱 AI 워크플로우에 대한 접근 제어:** 관리해야 하는 것은 AI 도구에 대한 인간의 접근만이 아닙니다. MCP 서버 및 기타 중요 시스템에 대한 에이전틱 AI 액세스도 관리해야 합니다. 이를 위해서는 ID 관리 및 액세스 제어에 대한 새로운 접근 방식이 필요합니다.

공개 AI 애플리케이션 및 모델 보호

AI 시스템은 내부 개발되었는지, 혹은 제3자를 통해 제공되는지에 관계없이 고객 및 사용자 경험의 필수적인 요소가 되고 있으므로, 반드시 보호되어야 합니다. LLM 애플리케이션에 대한 OWASP 상위 10가지 취약점은 기존 도구로 해결할 수 없는 다음과 같은 위협을 강조합니다.

- **무제한 자원 소비 공격:** 기존 DoS 공격과 유사하게, 무제한 자원 소비 공격은 리소스 집약적인 요청으로 LLM을 과부하 상태로 만드는 것을 목표로 합니다. 종량제 클라우드 모델에서는 이러한 공격이 발생할 경우 비용이 급격히 증가할 수 있으며, 정상 사용자들은 서비스 품질 저하를 겪게 됩니다.
- **모델 포이즈닝:** 공격자는 모델 학습을 위해 개발자가 사용하는 공개 데이터 세트 또는 리포지토리에 손상된 데이터를 주입하여 LLM에 백도어, 편향 또는 취약점을 심습니다. 이러한 방식으로 오염된 모델은 특정 트리거가 유해한 동작을 활성화할 때까지는 정상적으로 작동합니다.
- **프롬프트 삽입 공격:** AI 인터페이스를 통해 데이터를 유출하는 데 널리 사용되는 방법인 프롬프트 삽입은 악성 명령을 사용자 프롬프트나 외부 콘텐츠에 삽입하여 LLM 입력을 조작합니다. 이를 통해 모델이 원래 지침을 무시하고 공격자가 내린 명령을 대신 실행하도록 합니다.
- **탈옥:** 역할극 시나리오, 명령어 재정의, 다중 턴 전략과 같이 정교하게 만들어진 프롬프트는 LLM의 안전 장치를 우회하여 금지된 콘텐츠를 생성하거나 민감한 정보를 추출할 수 있습니다.

조직을 위험에 빠뜨리지 않으면서 빠른 AI 혁신을 실현하기 위해 리더는 전체 요구 사항 범위에 걸쳐 포괄적인 AI 보안 접근 방식을 도입하는 것이 좋습니다.

- 직원들의 생성형 AI 사용 보호
- AI 기반 애플리케이션 및 워크로드 보호
- 설계부터 안전한 AI 기반 애플리케이션 빌드

AI 개발 및 학습 워크플로우 보호

AI 프로젝트는 대규모 데이터 세트, 고비용 리소스 및 반복적인 실험을 포함하며, 이는 모두 새로운 공격면 및 취약점을 생성합니다. 보안 팀은 다음 사항을 해결해야 합니다.

- **학습 데이터 보안 및 무결성:** 손상된 학습 데이터는 프로덕션 모델에 지속적인 편향, 백도어 또는 취약점을 유발할 수 있습니다. 조직은 학습 데이터 세트에 대한 접근을 안전하게 보호하고, 무단 변경을 방지하며, 모델 수명주기 동안 데이터 출처를 보장해야 합니다.
- **자격 증명 및 비밀 관리:** AI 개발 워크플로우에서는 데이터 세트용 클라우드 스토리지, 학습용 컴퓨팅 클러스터, 모델 레지스트리 및 타사 서비스를 포함한 시스템에 대한 액세스가 필요합니다. 비밀 또는 API 키가 제대로 보호되지 않으면 중요한 자격 증명이 노출되어 독점 모델, 학습 데이터 또는 프로덕션 시스템에 대한 무단 액세스를 허용할 수 있습니다.
- **개발 환경 액세스 제어:** AI 엔지니어는 모델을 실험하고 민감한 데이터에 접근하기 위해 높은 권한이 필요한 경우가 많습니다. 적절한 액세스 제어가 없으면 내부자 위협, 실수로 인한 데이터 노출 또는 무단 모델 추출로 이어질 수 있습니다.

Cloudflare와 보안 통합

성공적인 AI 도입은 생산성을 제한하는 것이 아니라 생산성을 높이는 데 중점을 둡니다. 팀이 적절한 안전 장치가 마련되어 있다는 것을 확신하고 AI를 사용할 수 있을 때, 혁신은 더욱 빨라지고 더 야심찬 프로젝트를 수행할 수 있습니다.

Cloudflare AI Security Suite는 AI 혁신을 위한 안전한 환경을 조성합니다. 최고 경영진은 보안 액세스 서비스 에지(SASE)와 웹 애플리케이션 보안 역량을 통합하여 다음의 두 가지 핵심 영역을 연결하고 보호할 수 있습니다.

- 외부의 퍼블릭 AI 기반 애플리케이션
- 내부의 프라이빗 AI 시스템과 워크로드

전체 AI 수명주기에 초점을 맞춘 Cloudflare AI Security Suite는 검색 및 위험 관리부터 데이터 보호, 사용자 액세스 보호, AI 지원 애플리케이션 및 개발 워크플로우 보호에 이르기까지 모든 보안 요구 사항을 해결합니다. Cloudflare 글로벌 네트워크는 실시간 프로덕션 보안의 핵심적인 계층을 제공하기 위해 인라인으로 작동하여 모든 AI 상호 작용을 검사 및 필터링하고, 모든 사용자와 애플리케이션에서 데이터를 보호합니다.

Cloudflare는 문제가 발생한 후에 문제를 감지하는 대신, AI 모델에 문제가 도달하기 전에 이를 예방하고 위협을 차단합니다. 보안 팀은 새로운 위협에 앞서 나가기 위해 필요한 가시성을 확보합니다.

Cloudflare의 주요 기능

Cloudflare AI Security Suite는 포괄적인 모니터링, 실시간 보호, 사전 예방적 위험 관리를 하나의 플랫폼으로 통합하여, AI 보안에 대한 전체적 접근 방식을 제공합니다.

포괄적인 AI 검색 및 가시성

효과적인 AI 보안은 승인 및 미승인 AI 리소스와 사용 현황에 대한 완전한 실시간 인벤토리를 확보하는지 여부에 달려 있습니다. Cloudflare AI Security Suite의 기반은 모든 유형의 환경(퍼블릭, 프라이빗, 내부)에서 모든 AI 모델, 어시스턴트, 에이전트 및 섀도우 AI 배포를 식별하기 위한 지속적인 모니터링 및 자동 검색입니다.

사전 예방적 AI 위험 관리

Cloudflare AI Security Suite는 조직이 AI 관련 취약점, 잘못된 구성 및 공격 경로를 탐지하고 완화하여 공격을 방지하도록 지원하며, 여기에는 LLM에 대한 OWASP 상위 10가지 취약점도 포함됩니다. 애플리케이션 신뢰도 점수 시스템은 팀이 가장 중요한 위험을 먼저 해결할 수 있도록 해결 우선순위를 지정하는 데 도움이 됩니다.

AI 기반 앱 애플리케이션 보안

Cloudflare AI Security Suite는 SecOps가 최신 AI 위협 벡터에 대한 정보를 지속적으로 업데이트할 수 있도록 AI 파이프라인 내의 AI 관련 취약점, 잘못된 구성 및 공격 경로에 대한 사전 예방적 위협 감지 및 완화 기능을 제공하며, 프롬프트 주입, 데이터 포이즈닝 및 모델 오용으로부터 보호합니다.

특수 AI 방화벽은 생성형 또는 에이전틱 AI 및 API 엔드포인트를 탐지 및 분류하고, 개인 식별 정보 유출 시도를 감지하며, 악성 프롬프트가 AI 모델 성능에 영향을 주거나 유해 콘텐츠 또는 허위 정보로 모델을 오염시키기 전에 차단합니다.

생성형 AI 및 에이전틱 AI 워크플로우에 대한 Zero Trust 액세스

최소 권한 등의 Zero Trust 원칙은 직원과 AI 에이전트 모두에게 적용됩니다. Cloudflare AI Security Suite는 인간 대 AI 및 AI 대 AI 상호 작용 모두에 대해 Zero Trust 네트워크 액세스 정책(ZTNA)을 적용할 수 있습니다. MCP 서버에 대한 중앙 집중식 로그 기록 및 제어를 통해 에이전틱 AI가 즉시 실행 워크플로우를 포함하여 권한이 부여된 대상에만 액세스하도록 보장합니다.

AI 인식 데이터 보호

효과적인 AI 보안을 위해서는 데이터 손실 방지 기능이 필요합니다. 이는 여러 언어 모델을 활용하여 프롬프트 내용과 그 이면에 숨겨진 의도까지 파악할 수 있는 기능입니다. Cloudflare AI Security Suite는 학습, 프롬프트, 응답 전반에 걸쳐 데이터 손실 방지(DLP) 기능을 통합하여 AI 모델과 파이프라인 내의 개인 식별 정보 노출, 데이터 유출, 무단 액세스를 방지합니다. 인라인으로 배포되는 API 중심 런타임 보안은 빠르고 간단한 첫 번째 방어 계층 역할을 하여 CNAPP를 통해 지원되는 Shift-Left 접근법을 보완합니다.

데이터 현지화

LLM 및 AI 애플리케이션은 다른 데이터 환경과 마찬가지로 동일한 규정을 준수해야 합니다. Cloudflare AI Security Suite는 학습 데이터와 추론 요청을 승인된 지역 내에 유지하는 정책을 통해 조직에서 AI 워크로드가 지리적 및 관할권 경계를 준수하도록 보장합니다.

AI 개발을 위한 설계 단계부터 고려된 보안

다른 모든 소프트웨어와 마찬가지로 AI 애플리케이션 보안 역시 개발 주기 초기에 내장되어야 하며, 뒤늦게 추가되어서는 안 됩니다. Cloudflare AI Security Suite는 개발자에게 설계 단계부터 안전한 AI 기반 애플리케이션을 구축할 수 있는 도구 및 프레임워크를 제공합니다.

Cloudflare AI Security Suite의 비즈니스 영향

AI의 부상은 단순한 진화가 아니라 산업화 또는 컴퓨터화와 같은 근본적인 혁신입니다. 따라서 빠르고 효과적인 도입은 성장 촉진뿐만 아니라 조직의 생존 가능성과도 관련이 있습니다. 느리거나 안전하지 못한 도입은 이제 모든 산업 분야의 조직에게 생존과 관련된 위협이 되고 있습니다.

Cloudflare AI Security Suite는 치열한 경쟁 환경에서 가속화되는 고객 기대치 및 시장 요구 사항을 충족하기 위해 안전하고 제어된 상태에서의 효율적인 변환을 지원합니다.

- **더욱 빨라지는 AI 혁신:** 보안이 도입을 막기보다는 안전한 사용을 가능하게 할 때, 직원과 팀은 새로운 AI 애플리케이션을 탐색하여 일상 업무에서 생산성을 향상시킬 수 있습니다. 적절한 보안 프레임워크는 개발자가 민감한 데이터나 시스템을 위태롭게 하지 않고도 야심찬 AI 기능을 구축할 수 있도록 자신감을 줍니다.
- **AI 관련 위험 감소:** 조직은 AI 지원 웹 애플리케이션에 내재된 AI 관련 위험을 포함하여 AI 도입에 따른 모든 위험을 관리할 수 있습니다. 개발 중인 AI 애플리케이션은 보안을 강화하여, 직원들이 민감한 데이터를 유출하거나 AI 학습용 데이터 세트에 노출하지 않도록 보호할 수 있습니다. SecOps는 AI 관련 특정 위협 및 취약점을 사전에 식별하고 완화하여 공격면을 최소화하고 중요한 데이터와 모델을 보호할 수 있습니다.
- **보안 운영 간소화:** AI 보안 태세에 대한 중앙 집중식 가시성 및 제어를 통해 관리 효율성을 높이고 사고 대응을 간소화합니다. SecOps 팀은 AI 관련 사고에 계속 대응하는 대신 전략적 이니셔티브에 집중할 수 있습니다.
- **강력한 데이터 거버넌스 및 규제 준수:** AI 맞춤형 데이터 보호 제어는 AI 수명주기 전반에서 중요한 정보를 보호하고 변화하는 규제 요구 사항을 충족하도록 지원합니다.
- **총 소유 비용(TCO) 절감:** 기존 보안 투자를 통합하는 통합 플랫폼을 활용하는 것이 각 AI 보안 문제에 대해 별도의 포인트 솔루션을 구현하는 것보다 경제적입니다.

Cloudflare AI Security Suite의 모델 사용 사례

Cloudflare AI Security Suite는 AI 도입과 관련된 핵심 요구 사항을 해결하는 데 적합합니다.

- **직원 AI 도구 사용 보안:** ChatGPT와 같은 퍼블릭 생성형 AI 도구와 내부 개발 AI 기반 애플리케이션을 사용하는 직원의 액세스에 대해 Zero Trust 정책을 적용합니다.
- **퍼블릭 AI 기반 애플리케이션 보호:** 챗봇 및 추천 엔진과 같이 AI 모델을 통합하는 웹 애플리케이션 및 API를 중요한 데이터 노출 또는 모델 남용 공격으로부터 보호합니다.
- **섀도우 AI 관리:** 조직 전체에서 무단 AI 도구를 자동으로 검색하고 적절한 제어를 적용하여 관리되는 위험 범위 내에서 지속적인 혁신을 지원합니다.
- **AI 상호 작용을 위한 AI 기반 DLP:** AI 프롬프트 또는 응답에서 민감한 데이터가 노출되지 않도록 방지하여 PII 및 기밀 정보를 보호합니다.
- **AI 개발 보안:** 엔지니어링 팀에 보안 개발이 기본 접근 방식이 되도록 프레임워크를 제공하여, 보안을 저해하지 않으면서 AI 기능을 신속하게 구축할 수 있게 합니다.

구현 관련 고려 사항

AI 보안은 기본적인 보안 전략으로서 신중하게 접근해야 합니다.

현재 인프라부터 시작	가장 성공적인 AI 보안 구현은 기존 SASE 및 애플리케이션 보안 도구를 대체하기보다는 기존 SASE 및 애플리케이션 보안 도구를 기반으로 구축되는 것입니다. 이 접근 방식은 기존 투자를 활용하면서 AI 특화 위험을 포괄하도록 보호 범위를 확장합니다.
통합된 인라인 보호 기능 배포	악성 활동이 네트워크 에지에서 발생하는 즉시 이를 차단하는 능력은 AI 보안에 매우 중요합니다. 실시간 인라인 제어 기능을 구축하고 API 기반 모니터링으로 보완하세요.
완벽한 커버리지 보장	AI 보안 전략은 직원의 생성형 AI 도구 사용, AI 기반 애플리케이션 및 워크플로우 보호, 에이전틱 AI 워크플로우 보안, 그리고 AI 개발 워크플로우 보안을 포함한 전체 요구 사항을 충족해야 합니다.
엔터프라이즈 규모를 위한 계획	AI 도입과 함께 확장 가능한 솔루션을 선택하세요. 하나의 파일럿 프로그램에서 효과를 본 방법은 AI 활용이 확대됨에 따라 조직 전체로도 확장 적용되어야 합니다.
성공 기준의 유효성 검사	전면 도입에 앞서 실제 환경에서 솔루션을 검증해야 합니다. 엔터프라이즈 기능을 무료 셀프 서비스로 활성화할 수 있는 플랫폼을 선택하세요. 이를 통해 단일 팀 또는 애플리케이션과 같이 작은 규모에서 전체 보안 제품군을 테스트하여 가치를 빠르게 입증하고 성공 기준을 충족하는지 확인할 수 있습니다.

Cloudflare AI Security Suite로 다음 단계로 나아가기

전 세계적으로 AI 도입이 가속화됨에 따라 AI를 안전하게 확장할 수 있는 조직은 상당한 경쟁력을 확보할 것입니다. 핵심은 AI 보안이 AI 사용을 막는 것이 아니라, AI를 지능적으로 활용할 수 있도록 지원하는 데 있다는 점을 인식하는 것입니다.

Cloudflare AI Security Suite를 통해 조직은 안심하고 혁신할 수 있습니다. 널리 사용되는 Cloudflare의 SASE 및 애플리케이션 보안 플랫폼을 기반으로 확장된 AI Security Suite는 통합 AI 검색, Zero Trust 액세스 제어, 인텔리전스 기반의 사전 예방적 위협 방어 및 강력한 데이터 거버넌스를 제공합니다. 모든 측면에서 AI 사용과 모델을 안전하게 보호함으로써 조직은 최종 사용자 경험을 저해하지 않으면서 개발자가 더 빠르게 개발하고 직원의 생산성을 향상시킬 수 있도록 지원할 수 있습니다.

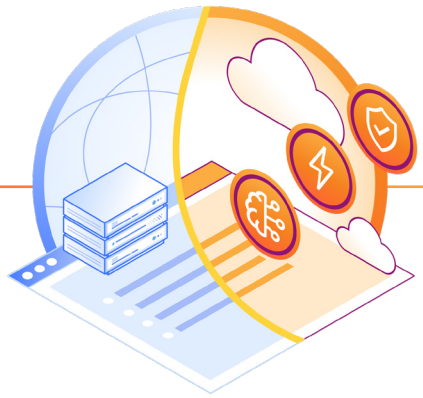
자문 예약

Cloudflare AI Security Suite가 안전한 AI 도입을 위한 조직의 접근 방식을 어떻게 혁신할 수 있는지 알아보세요.

007-9814-2030-192

enterprise@cloudflare.com

www.cloudflare.com/ko-kr



1. ManageEngine. 기업 내 섀도우 AI 급증: 미국 및 캐나다 인사이트
2. Check Point 연구. Check Point Software에서 발표한 2025년 1분기 글로벌 사이버 공격 보고서: 약 50% 급증한 전 세계 사이버 위협, 126%나 증가한 랜섬웨어 공격