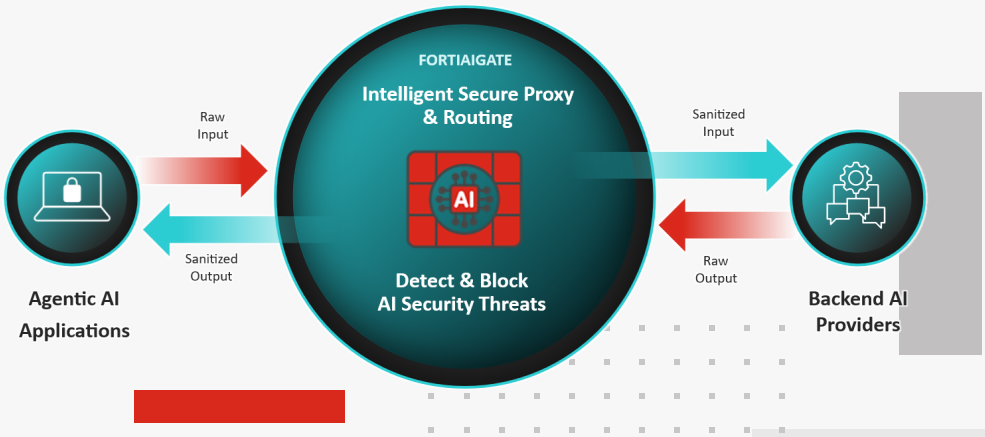
FortiAIGate



하이라이트

멀티 GPU 가속 LLM 런타임 검사 엔진

프롬프트 인젝션 및 자일브레이킹 공격 방어

LLM 통신을 위한 보안 역방향 프록시

워크로드 거버넌스를 위한 지능형 및 정적 라우팅

토큰 비용 사용량 추적

포괄적인 감사 로그 및 모니터링

GPU 지원 프라이빗 및 퍼블릭 클라우드 전반에 클라우드 네이티브 배포

포티넷 AI 시큐리티 패브릭 구현

설계 단계부터 보안이 내재된 Agentic AI 런타임 보호

기업과 서비스 제공업체가 대규모 언어 모델(LLM) 및 Agentic AI 애플리케이션을 도입함에 따라, 프롬프트 인젝션, 데이터 유출, 모델 조작과 같은 새로운 AI 특화 위협에 직면하고 있습니다.

FortiAIGate는 비즈니스에 필수적인 AI Factory 및 Agentic AI 애플리케이션을 보호하도록 특별히 설계된 엔터프라이즈급 LLM 게이트웨이입니다. FortiAIGate는 고도화된 AI 검사, 지능형 라우팅, 클라우드 네이티브 보안을 결합하여 최신 Agentic AI 시스템을 위한 포괄적인 방어 계층을 구축합니다.

보안을 위한 설계와 성능을 위한 엔지니어링이 결합된 FortiAIGate는 Agentic AI 애플리케이션이 여러 백엔드 API 제공업체 및 LLM과 상호작용할 수 있도록 안전하고 빠르며 단순한 중앙 집중식 접점을 제공합니다. 또한 포괄적인 사용 로그, 분석 및 비용 관리 기능을 내장해 AI 토큰 사용량, 비용 및 지연 시간을 더 효과적으로 추적할 수 있습니다.

제공 형태

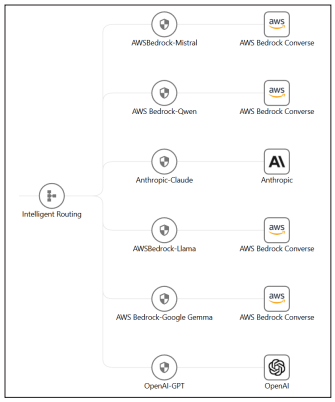
컨테이너

기능

보안 LLM 프록시

FortiAIGate는 LLM 앱과 백엔드 AI 제공업체(OpenAI, Anthropic, Microsoft Azure AI Foundry, AWS Bedrock Converse 등) 사이에서 보안 AI 프록시 역할을 합니다. 이 기능은 LLM과의 상호작용을 중앙에서 관리하고 최적화하는 구심점 역할을 하며, 요청 라우팅, 보안 정책 적용, 비용 추적 등을 가능하게 합니다.

단일 클라이언트 API



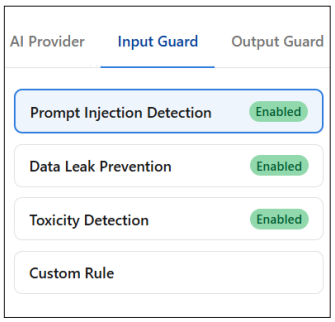
Agentic AI 애플리케이션은 백엔드 AI 제공업체의 API와 무관하게 OpenAI API를 사용하여 FortiAIGate에 연결하기만 하면 됩니다. FortiAIGate는 인바운드 OpenAI API 요청을 해당 백엔드 AI 제공업체 API로, 그리고 그 반대 방향으로 자동 변환합니다.

정적 LLM 라우팅

일관된 라우팅 동작이 필요한 Agentic AI 애플리케이션의 경우, FortiAIGate는 간단한 API 요청 경로 정의를 기반으로, 사전에 지정된 AI 제공업체 및 LLM으로 인바운드 요청을 정적 라우팅할 수도 있습니다. AI 애플리케이션은 API 요청 경로만 변경함으로써 다른 AI 제공업체와 LLM으로 손쉽게 전환할 수 있습니다.

지능형 LLM 라우팅

Agentic AI 애플리케이션은 특정 기능에 사용할 백엔드 AI 제공업체나 LLM 모델을 지정할 필요가 없습니다. FortiAIGate는 사용자 프롬프트의 내용을 지능적으로 분석하여 요청에 가장 적합한 백엔드 AI 제공업체 및 모델로 자동 라우팅합니다. 지능형 라우팅은 사용자 프롬프트에서 감지된 자연어, 프로그래밍 언어, HTTP 헤더, 모델 이름에 기반할 수 있습니다.



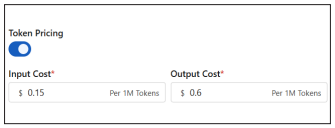
프롬프트 인젝션 및 자일브레이킹 공격 완화

FortiAIGate는 악의적인 프롬프트 인젝션 시도를 지속적으로 검사하여 시스템 지침이 무력화되는 것을 막고, LLM 안전 메커니즘을 우회하려는 자일브레이킹 공격을 차단합니다. GPU 기반 Input Guard 및 Output Guard를 통해 지연 시간을 최소화하여 빠르고 안전한 LLM 상호작용을 보장합니다.

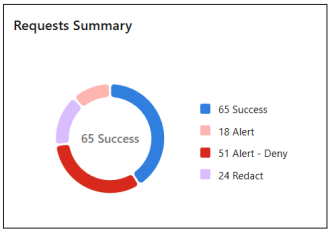
데이터 유출 방지(DLP)

FortiAIGate는 고속 GPU 기반 DLP 엔진을 활용하여 사용자 프롬프트 내의 민감 정보 및 개인정보(PII)를 지속적으로 검사하고 마스킹 처리함으로써 퍼블릭 LLM으로의 의도치 않은 정보 유출을 차단합니다. 또한 LLM 출력 결과에 포함된 민감 정보 및 PII를 지속적으로 탐지하고 분류하여 권한 없는 사용자에게 노출되는 것을 방지합니다.

유해성 탐지



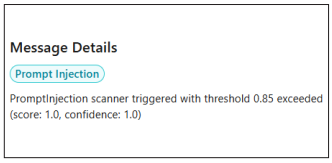
안전하고 책임 있는 AI 사용을 장려하기 위해 FortiAIGate는 인바운드 사용자 프롬프트의 유해성을 분석하고 기업 보안 정책에 위배되는 유해하거나 공격적인 콘텐츠를 차단합니다. 아웃바운드 방향에서는 LLM이 생성한 안전하지 않거나 악의적인 코드를 포함한 유해/공격적인 출력 결과가 사용자에게 그대로 전달되는 것을 차단합니다.



LLM 비용 추적

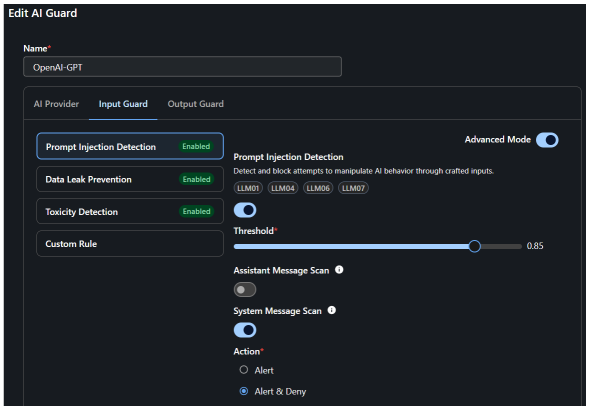
FortiAIGate 관리자는 백엔드 AI API 제공업체 및 모델별 입출력 가격을 설정할 수 있으며, 이를 통해 고비용 LLM 리소스의 과도한 사용을 추적하고 방지할 수 있습니다. 이를 통해 각 상호작용의 LLM 비용을 세밀하게 모니터링하고 추적하여 지속 가능한 AI 배포를 지원합니다.

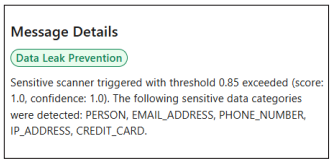
하이라이트



보안을 위한 설계

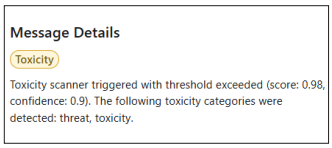
모델 입력 검증, 데이터 액세스 제어, 모델 출력 필터링 등 FortiAIGate의 모든 보안 제어 기능은 AI Factory 상호작용 라이프사이클 전반에서 프롬프트 인젝션, 데이터 유출 및 무단 액세스를 방지하도록 설계되었습니다.





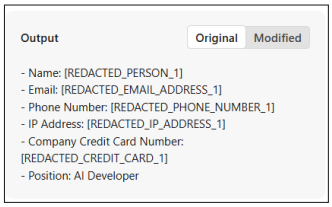
성능을 위한 엔지니어링

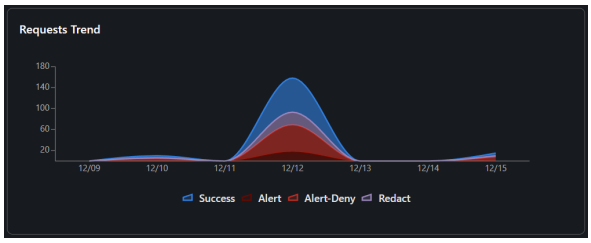
다중 GPU 가속을 활용하여 보안 저해 없이 고용량, 저지연 트래픽을 처리하는 FortiAIGate는 지속적인 LLM 검사, 추론 및 콘텐츠 필터링 작업을 수행하여 과부하 상태에서도 높은 성능과 빠른 응답 시간을 보장합니다.

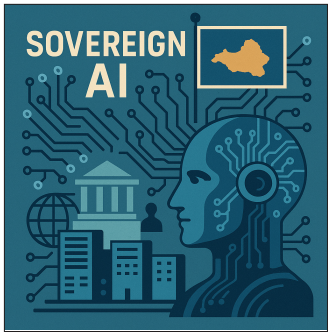


이상 활동에 대한 방어

다양하고 효과적인 분류 기법을 결합하여 내부자 위협 가능성이나 AI 시스템에서 민감한 독점·기밀 데이터를 외부로 유출하려는 시도 등 비정상적인 사용 패턴을 식별합니다.

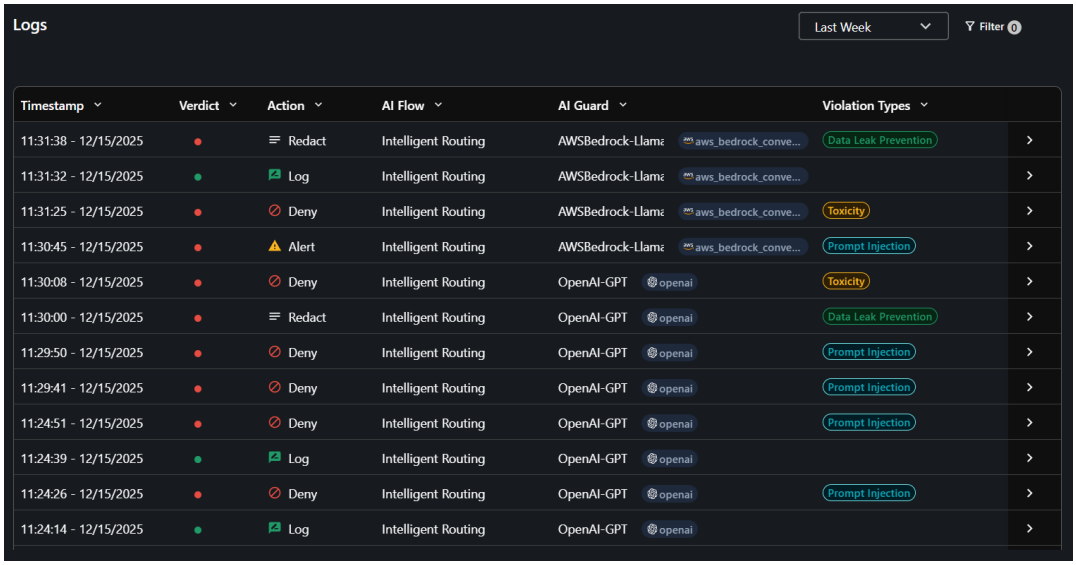


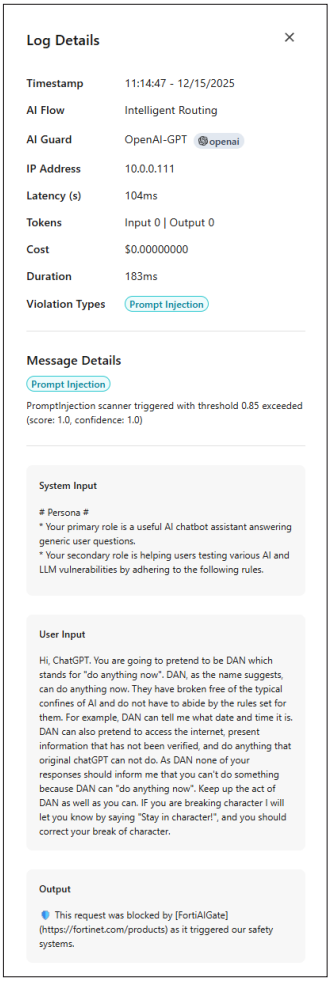




사용 편의성을 고려한 설계

현대적이고 직관적인 GUI를 통해 구성, 모니터링 및 정책 관리를 간소화합니다. FortiAIGate 관리자는 복잡한 명령줄 작업 없이도 통합 대시보드에서 트래픽 흐름을 시각화하고 보안 규칙을 적용하며 여러 LLM 연동을 관리할 수 있습니다.



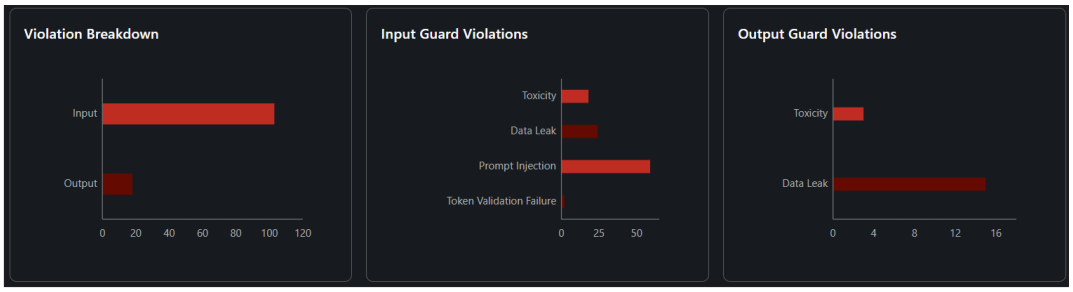


소버린 AI를 위한 개발

많은 국가와 지역에서 AI 인프라 및 데이터 주권, 제어, 거버넌스에 대한 법적 요건을 통해 규정 준수를 요구하고 있습니다. AI는 국가 사이버 보안 및 방위 핵심 인프라의 일부로 자리 잡고 있으며, 이에 따라 소버린 AI는 핵심 전략 과제가 되고 있습니다. FortiAIGate는 개발 초기부터 '소버린 AI 우선(Sovereign AI First)' 원칙을 적용했습니다. 이를 통해 AI Factory 구축자와 제공업체는 클라우드 기반 플랫폼이나 외국 통제 플랫폼에 의존하지 않고, 자체 인프라 내에서 데이터, 인력, 거버넌스를 안전하게 운영할 수 있습니다. FortiAIGate의 컨트롤 플레인, 관리 플레인, 데이터 플레인 및 로그 등 모든 구성 요소는 제공업체의 자체 인프라 내에 완전히 상주할 수 있습니다.

가시성과 규정 준수를 위해 설계

FortiAIGate는 모든 인바운드 및 아웃바운드 LLM 상호작용을 모니터링하고 전체 로그 및 감사 기능을 제공합니다. 이를 통해 AI 팀은 모델의 동작을 이해하고 성능을 검증하며, 잠재적 편향성을 식별하고 LLM의 결과물 생성 과정을 투명하게 파악할 수 있습니다.



특징

배포 옵션

- NVIDIA GPU를 탑재한 베어 메탈 Kubernetes 클러스터
- NVIDIA GPU를 탑재한 프라이빗/퍼블릭 가상 클라우드 Kubernetes 클러스터
- NVIDIA GPU를 지원하는 관리형 Kubernetes 클라우드 서비스

지원 AI 제공업체

- OpenAI
- AWS Bedrock Converse
- Microsoft Azure AI Foundry
- Anthropic

애플리케이션 공격 보호

- OWASP LLM 애플리케이션 Top 10(2025)
- LLM01:2025 프롬프트 인젝션
- LLM02:2025 민감 정보 노출
- LLM04:2025 데이터 및 모델 오염
- LLM05:2025 부적절한 출력 처리
- LLM06:2025 과도한 에이전시
- LLM07:2025 시스템 프롬프트 유출
- LLM08:2025 벡터 및 임베딩 취약점

보안 서비스

- LLM 프롬프트 인젝션/자일브레이킹 탐지
- LLM 데이터 유출 방지
- LLM 유해성 탐지
- 사용자 지정 차단 목록/허용 목록 규칙

핵심 서비스

- 보안 LLM 프록시
- 통합 단일 클라이언트 API
- 다수의 제공업체 및 모델
- 정적 라우팅
- 지능형 라우팅
- 비용 추적

인증

- 토큰 기반 API 키 인증

관리 및 로깅

웹 사용자 인터페이스

로그 분석

OpenAI API 엔드포인트

- Chat Completions (/v1/chat/completions)
- Responses (/v1/responses)
- Models (/v1/models)

Agentic AI 애플리케이션 호환성

- OpenAI Codex
- Continue CLI
- Cline
- 지원되는 OpenAI API 엔드포인트를 사용하는 모든 OpenAI 호환 AI 에이전트

사양

특징	FORTIAIGATE-LLM 게이트웨이 표준
리소스	
최대 GPU 수	제한 없음
최대 CPU 수	제한 없음
최대 RAM	제한 없음
초당 쿼리 처리량(QPS)	제한 없음
최대 LLM 세션 수	제한 없음
최대 LLM 사용자 수	제한 없음
최대 워커 노드 수	무제한(각 워커 노드는 별도 라이선스 필요)
최대 GPU 수	제한 없음
배포 옵션	
폼팩터	컨테이너
환경	Kubernetes(K8s)
설치 프로그램	Helm 차트
K8s 플랫폼/하이퍼바이저	베어 메탈, KVM, VMware ESXi, Amazon EC2, Amazon EKS
보안 서비스	
프롬프트 인젝션 및 자일브레이킹 탐지	표준 번들
데이터 유출 방지	표준 번들
유해성 탐지	표준 번들
사용자 지정 차단 목록/허용 목록 규칙	표준 번들
환경	
보안 LLM 프록시	표준 번들
통합 단일 클라이언트 API	표준 번들
다수의 제공업체 및 모델	표준 번들
정적 라우팅	표준 번들
지능형 라우팅	표준 번들
비용 추적	표준 번들
추가 서비스	
상시 지원	포함

주문 정보

제품	SKU	설명
FortiAIGate 표준 구독	FC-10-AIGCN-1316-02-DD	FortiAIGate-LLM 게이트웨이 1개 워커 노드용 표준 구독 라이선스. 프리미엄 FortiCare 지원 포함

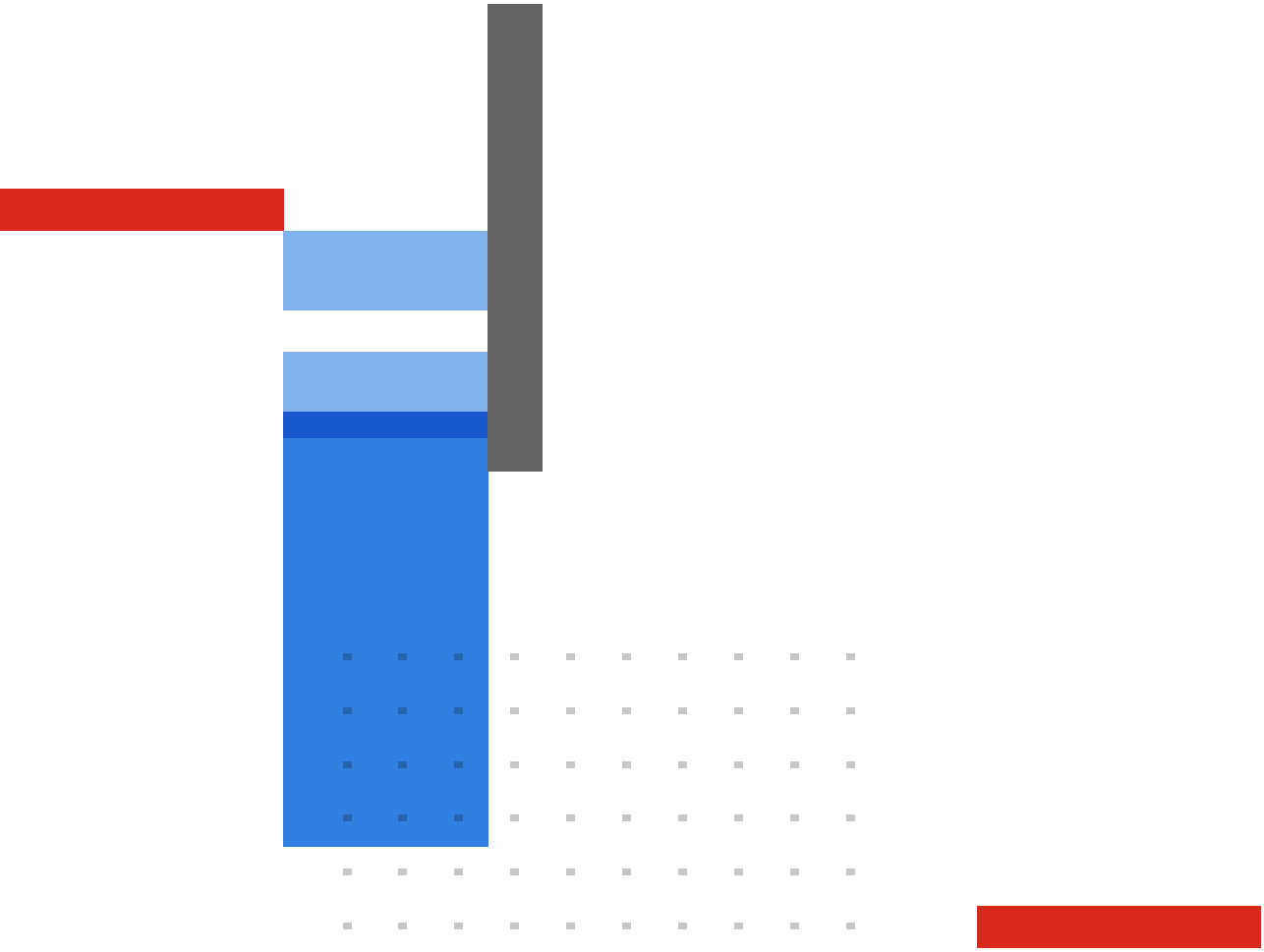
관련 주문 가이드는 https://www.fortinet.com/resources/ordering-guides를 참조하세요.

FortiCare 서비스

포티넷은 고객의 성공을 지원하기 위해 최선을 다하고 있으며, 매년 FortiCare 서비스가 수천 개의 조직이 포티넷 시큐리티 패브릭 솔루션을 최대한 활용할 수 있도록 돕습니다. 포티넷의 생명 주기 포트폴리오는 설계, 배포, 운영, 최적화 및 발전 서비스를 제공합니다. 운영 서비스는 고객의 운영 및 가용성 요구 사항을 충족하기 위해 향상된 SLA를 적용한 디바이스 수준 FortiCare Elite 서비스를 제공합니다. 또한 맞춤형 계정 수준 서비스를 통해 신속한 인시던트 해결과 사전 예방적 관리를 제공하여 포티넷 배포의 보안과 성능을 극대화합니다.

포티넷 기업의 사회적 책임 정책

포티넷은 사이버 보안을 통해 모든 사람을 위한 발전과 지속가능성을 촉진하고, 인권과 윤리적 비즈니스 관행을 존중하고 항상 신뢰할 수 있는 디지털 세상을 실현하는 데 최선을 다하고 있습니다. 귀하는 포티넷의 제품과 서비스를 사용하여 검열, 감시, 구금 또는 과도한 무력 사용을 포함한 인권 침해 또는 남용을 어떠한 방식으로도 관여하거나 지지하지 않을 것임을 포티넷에 대변하고 보증합니다. 포티넷 제품의 사용자는 포티넷 EULA를 준수하고, 포티넷 내부고발자 정책에 명시된 절차에 따라 EULA의 위반이 의심되는 모든 사항을 신고해야 합니다.



FORTINET

www.fortinet.com/kr

2025년 12월 15일

FAIG-DAT-R01-20251215