SOMANSA

GS인증(1등급) 보안기능확인서

WebKeeper AI DLP NEW

생성형 AI 서비스를 통한 개인·기밀·민감정보 유출방지 솔루션

WebKeeper AI DLP는 생성형 AI를 통해 유출되는 **개인·기밀·민감정보를 의미기반 정밀탐지(LLM)으로 탐지·통제**하며
내부자의 실수와 무단 전송을 모두 차단하고 접속 내역을 모두 로깅합니다.
국세청, 국가유산청 등 공공기관에서 선제적으로 도입, 안정화까지 완료하여 성능을 인정받았습니다.
N2SF 국가망 보안체계 최적화 및 제1금융권 최초 도입으로 증명된 생성형AI 기밀유출 통제의 표준 솔루션입니다.

1 30년 간 네트워크 DLP 시장 점유율 1위 기업 '소만사' 에서 선제적으로 개발

- **생성형 AI 전체 상호작용 중 39.7%에서 민감한 기업데이터가 노출**
  생성형 AI는 업무 생산성을 향상시키지만 동시에 기업 보안 리스크가 발생할 수 있음
- **생성형 AI 서비스 통제 및 감사로그 확보하여 임직원의 생성형 AI 사용현황 가시성 확보**
  30년간 누적된 네트워크 DLP 기술력을 기반으로 외부로 유출되는 개인·기밀정보 정밀 식별

2 기존 AI DLP의 패턴/키워드 기반 탐지의 한계점을 모두 보완

	기존 패턴기반 DLP	NEW LLM 분석 엔진 적용 AI DLP
탐지 방식	• 패턴/키워드 기반 탐지 • 복잡한 문맥, 맥락 이해 불가	• LLM 분석모델 기반으로 의미기반 분석, 프롬프트 맥락 이해 • 비정형 주소, API키, 민감정보, 인젝션, 유해 프롬프트 탐지 정확도 향상
이미지 분석	• 지도, 설계도, 도면 등 이미지 판단 어려움	• 도면, 이미지, 스캔파일의 내용과 구조 기준으로 카테고리 분류 • 이미지 내 텍스트 추출하여 맥락에 기반하여 분류 정확도 향상
의도 분석	• '고의적 유출시도' 구분 불가 • 불법/범죄목적 프롬프트 구분 불가	• 대화 맥락을 파악하여 '고의적 유출시도' 입력의도 정교하게 구분 • 불법/범죄/위협 의도 파악 및 구분가능

주요 고객사

국세청 | 국가유산청 | 서울특별시 | 우리은행 | KB국민은행 | LG유플러스 외 다수

특장점

국내최초

프롬프트 및 리스폰스 메신저 형태로 로깅/재현

대화내역 메신저 형태로
모두 재현하여 사후감사 시
전반적인 대화의도, 고의유출여부
판단지표 활용

국내최초

생성형 AI 접속 선별차단 및 기록

생성형 AI 차단 카테고리 제공
특정 생성형AI 서비스만 허용하고
나머지는 모두 차단하여
오남용 최소화

프롬프트·이미지·문서파일 벡터, LLM 기반 차단

프롬프트 입력의도를
맥락에 기반하여 통제
탐지된 데이터는 기밀·개인·민감정보,
인젝션, 유해 데이터로 분류

기능

생성형 AI를 통한 개인정보 유출통제 및 감사로그 확보

- 프롬프트 및 업로딩 파일 내 개인정보·기밀정보 포함 여부 실시간 탐지 및 차단
  차단내역은 마스킹 처리되어 2차 유출방지
- 파일원본 및 상세이력은 모두 로그로 저장되어 사후 감사자료 확보
- 프롬프트/리스폰스 값은 구문분류하여 메신저 형태로 재현, 비전문가도 직관적으로 감사자료 파악 용이
- 특정 키워드 포함 및 반복 사용/업로딩 파일 사이즈 기준으로 차단 등 정책 기반 차단 지원

개인·기밀·민감정보 LLM 기반 실시간 차단

- 정규식, 키워드 딕셔너리, 임베딩 벡터, LLM 분석 기반의 다층 분석 체계 제공
- 다층화된 분석을 기반으로 개인·기밀·민감정보 및 인젝션 정보에 대한 외부 유출을 실시간 모니터링
- 사전 학습된 데이터를 기반으로 형태, 키워드, 맥락, 작성의도 등을 읽고
  이미지·문서·문단 유사도 판단하여 실시간 차단 및 로그저장

문서 카테고리 자동 분류화

- 20개 대분류, 100여개 소분류 항목 기준으로 문서 성격에 따라 카테고리별로 분류하는 자동화 프로세스 자체 개발
- 기밀 수준, 위험도에 따라 문서 등급을 자동분류하여 정교한 보안체계 구현 가능

상용화된 대부분의 생성형 AI 서비스를 통제합니다.
WebKeeper AI DLP의 생성형 AI 서비스 커버리지는 지속적으로 확장되고 있습니다.