xLLM
제품 소개서 v1.1

2025.12.26

z3soft

xLLM G/W 개요

xLLM GW – AI-LLM 보안 가드레일 (보안특화 필터링)

Z3soft - xLLM은 Z3 기술을 확장 개발하여 웹 전용 보안 모듈을 포함한 1차 서비스를 제공합니다.
한국어 특화 및 보안 기능을 강화한 LLM으로, 기업 및 공공 시장의 특정 요구사항을 충족시키는 데 중점을 둡니다.

xLLM 기술 요소

한국어 특화 기반 모델
3B~7B 수준의 한국어에 최적화된 기반 모델

보안 프롬프트 필터ing
악의적인 프롬프트나 민감 정보 유출 시도를 사전 차단

민감 데이터 감지·차단 개인/민감/ 기업대외비
기업 내부 민감 데이터 자동 감지 및 외부 유출 차단

LLM 접근 통제 (Z3 연동)
Z3 보안 솔루션과 연동하여 접근 권한 세밀 통제

기업별 튜닝 (P-Tuning, LoRA)
고객사 데이터에 맞춰 모델 미세 조정으로 성능 최적화

브라우저별 대응가능

Firefox(Gecko) 보안 확장 개발 (2~3개월)

- API Wrapper Layer 구현
- clipboard API 차단 기능 별도 구현
- Firefox 권한 구조에 따른 백그라운드 스크립트 조정
- Firefox 전용 보안 플러그인 패키징

Safari(WebKit) 보안 확장 개발 (3~5개월)

- Chrome Extension → Safari Extension 변환 (X code converter 사용)
- WebKit 보안 정책에 맞춘 Native Messaging 추가
- “보안 모드”를 OS 레벨 정책과 연동 (MDM 가능)
- 기업용 Safari Extension 인증서 발급

사업 확장 방향

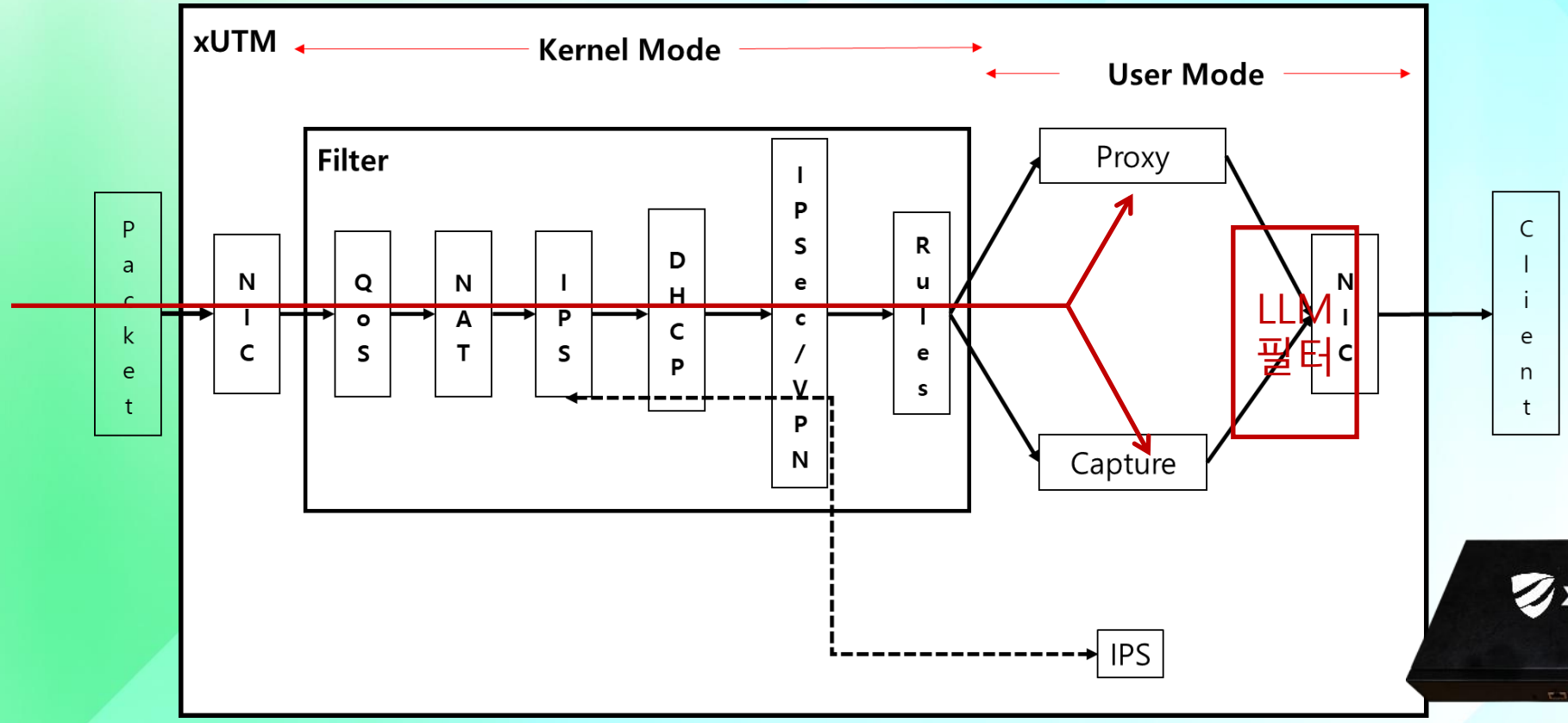
공공/금융 국책사업 참여

- AI 안전성·보안성 LLM 개발
- 공공 RAG 플랫폼 고도화
- 산업부 스마트팩토리 LLM
- 중기부 AI 바우처 적용

z3soft

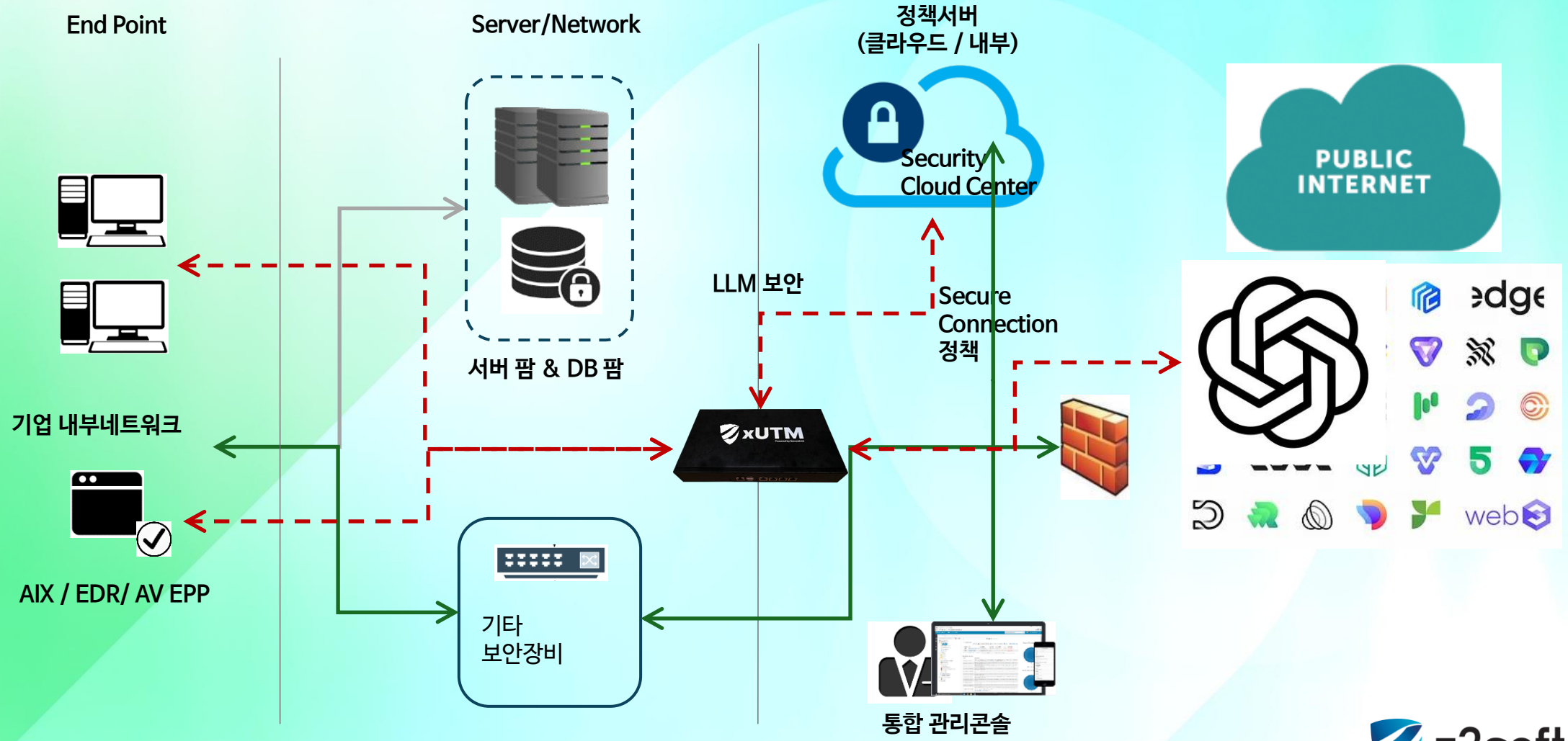
xLLM – SW 및 장비형 GW

기업에 설치되는 에이전트 및 브라우저 확장형 한계 존재 (사용자 우회 및 에이전트 설치 부담 / 클라우드 등)
LLM 사용시, 지정된 패턴 (개인정보 및 기업 키워드) 검출 차단 (모니터링 운영 가능)
공공 및 중소기업 통합 클라우드 서비스 등



z3soft

xLLM – SW 및 장비형 GW



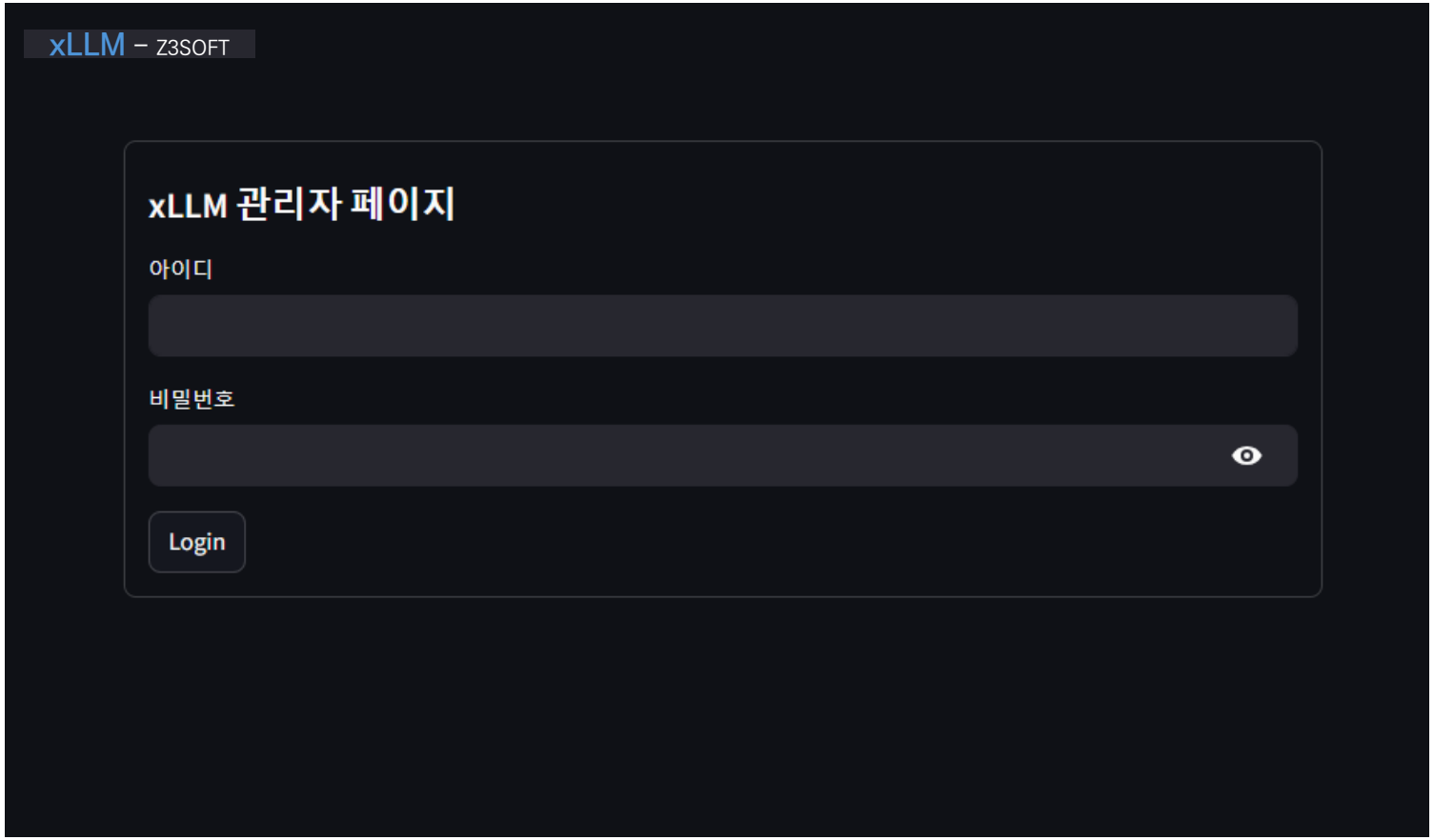
z3soft

Contents

01.

관리자 페이지

관리자 페이지 로그인

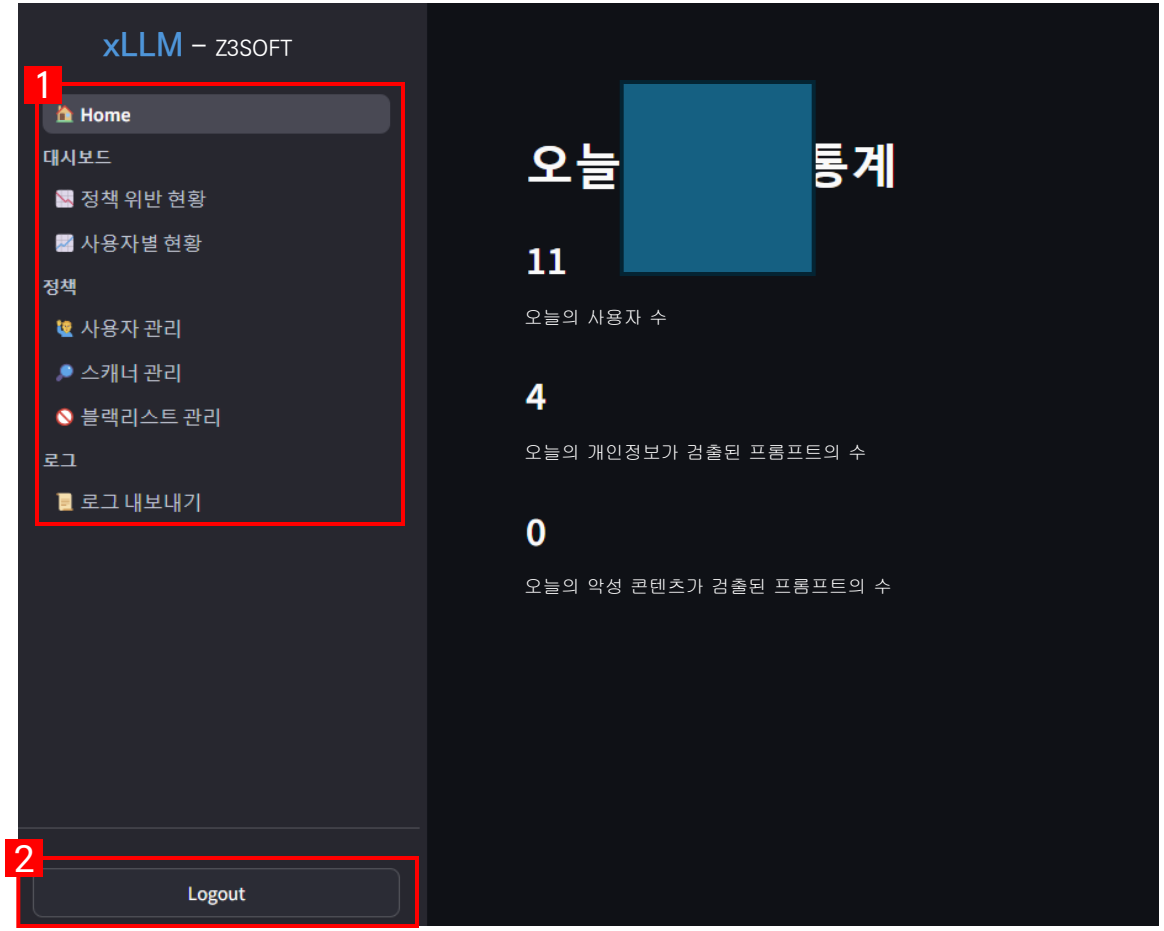


관리자 페이지에서는 xLLM이 탐지한 민감 정보와 위반 항목들을 실시간으로 모니터링할 수 있으며, 이를 통해 시스템 내 보안 상황을 즉시 파악하고 필요한 조치를 취할 수 있습니다. 또한, 위반 항목과 관련된 탐지 로그를 내보내는 기능을 제공하여 효율적인 보안 관리를 지원합니다.

관리자 페이지 링크:
아이디: (아이디)
비밀번호: (비밀번호)
(꼭 비밀번호 변경 후 사용해주세요.)

해당 아이디와 비밀번호는 관리자 계정이며,
xLLM 사용은 관리자가 사용자 계정 생성 후 사용 가능합니다.

관리자 페이지 메뉴



1 Home: 오늘의 xLLM 통계
오늘 날짜 기준으로 통계 확인

대시보드
정책 위반 현황: 위반 항목 기준 탐지 개수 확인
사용자별 현황: 생성된 계정 기준 위반 항목별 탐지 개수 확인

정책
사용자 관리: 사용자 계정 생성, 수정, 삭제
스캐너 관리: 특정 스캐너 선택 및 테스트
블랙리스트 관리: 차단 키워드 등록

로그
로그 내보내기: 아이디, 프롬프트 일시, 위반 항목, 탐지 결과, 처리 시간 로그 엑셀(csv)파일 다운로드

2 로그아웃

02.

대시보드

오늘의 xLLM 통계

1 **오늘의 사용자 수**: 오늘 날짜로 xLLM을 사용한 사용자의 수

2 **오늘의 개인정보가 검출된 프롬프트의 수**: 오늘 날짜로 개인정보가 검출된 프롬프트의 수

3 **오늘의 악성 콘텐츠가 검출된 프롬프트의 수**: 오늘 날짜로 악성 콘텐츠가 검출된 프롬프트의 수

정책 위반 현황

1 **날짜 선택**:
통계를 보고자 하는 날짜를 선택할 수 있습니다.
날짜는 최대 31일 선택 가능합니다.
선택된 기간에 쌓인 데이터가 없다면 '출력 할 데이터가 없습니다.'로 표시됩니다.

2 각 위반 항목에 대한 탐지 개수를 확인할 수 있습니다.

개인 식별 정보 위반: 개인식별정보(PII)를 포함한 모든 개인 정보가 프롬프트에 포함된 경우입니다.
악성 콘텐츠 위반: 악성 URL이나 코드 등 보안에 위협이 될 수 있는 콘텐츠가 포함된 경우입니다.
기밀 정보 위반: 회사 또는 조직의 비공개 정보가 포함된 경우입니다.
온라인 자격 증명 노출 위반: Git 토큰이나 API key 같은 온라인 서비스에 접근할 수 있는 자격 증명이 포함된 경우입니다.
유해 표현 정책 위반: 불쾌하거나 공격적인 표현이 포함된 경우입니다.

사용자별 현황

	개인 식별 정보 위반	악성 컨텐츠 위반	기밀 정보 위반	온라인 자격 증명 노출 위반	유해 표현 정책위반	총 탐지 개수
admin	615	9	69	9	3	134
extension	939	2	342	34	1	187
111108	20	15	10	8	0	8

1 아이디
사용자에게 부여된 계정이 표시됩니다.

2 선택된 기간에 각 위반 항목에 대한 탐지 개수를 확인할 수 있습니다.

개인 식별 정보 위반: 개인식별정보(PII)를 포함한 모든 개인 정보가 프롬프트에 포함된 경우입니다.
악성 콘텐츠 위반: 악성 URL이나 코드 등 보안에 위협이 될 수 있는 콘텐츠가 포함된 경우입니다.
기밀 정보 위반: 회사 또는 조직의 비공개 정보가 포함된 경우입니다.
온라인 자격 증명 노출 위반: Git 토큰이나 API key 같은 온라인 서비스에 접근할 수 있는 자격 증명이 포함된 경우입니다.
유해 표현 정책 위반: 불쾌하거나 공격적인 표현이 포함된 경우입니다.

3 총 탐지 개수
위반 항목의 총 개수를 확인할 수 있습니다.

03.

정책

사용자 관리

아이디	이름	이메일	마지막 로그인	총 사용량	계정 생성일
111108	윤서연	mortar@estsecurity.com	2024-10-23	10	2024-09-23 03:55:21
reyes2	신동훈	reyes2@estsecurity.com	2024-09-10	1,507	2023-03-16 03:00:00
preview	한미래	preview@estsecurity.com	2024-09-09	1	2024-09-04 00:00:00
admin	장현우	admin@estsecurity.com	2024-10-29	2,082	2023-10-11 00:00:00
kdb	김지원	kdb@estsecurity.com	2024-09-09	98	2024-09-04 00:00:00
hyun1	이서준	hyun1@estsecurity.com	2024-09-07	78	2024-09-04 00:00:00
postman	박민지	postman@estsecurity.com	2024-09-06	0	2024-09-05 00:00:00
000001	정우진	user1@estsecurity.com	2024-09-05	7	2024-09-05 00:00:00
sarah	최유나	sarah@estsecurity.com	2024-09-11	545	2024-09-05 00:00:00
extension	강태호	extension@estsecurity.com	2024-10-22	591	2024-04-21 21:00:00

사용자 계정을 추가, 수정, 삭제할 수 있습니다.
사용자는 해당 계정으로 ChatGPT 화면에서 xLLM에 로그인하여 사용할 수 있습니다.

1 일괄 추가
엑셀(CSV)파일을 이용하여 사용자 정보를 일괄로 추가 또는 수정할 수 있습니다.

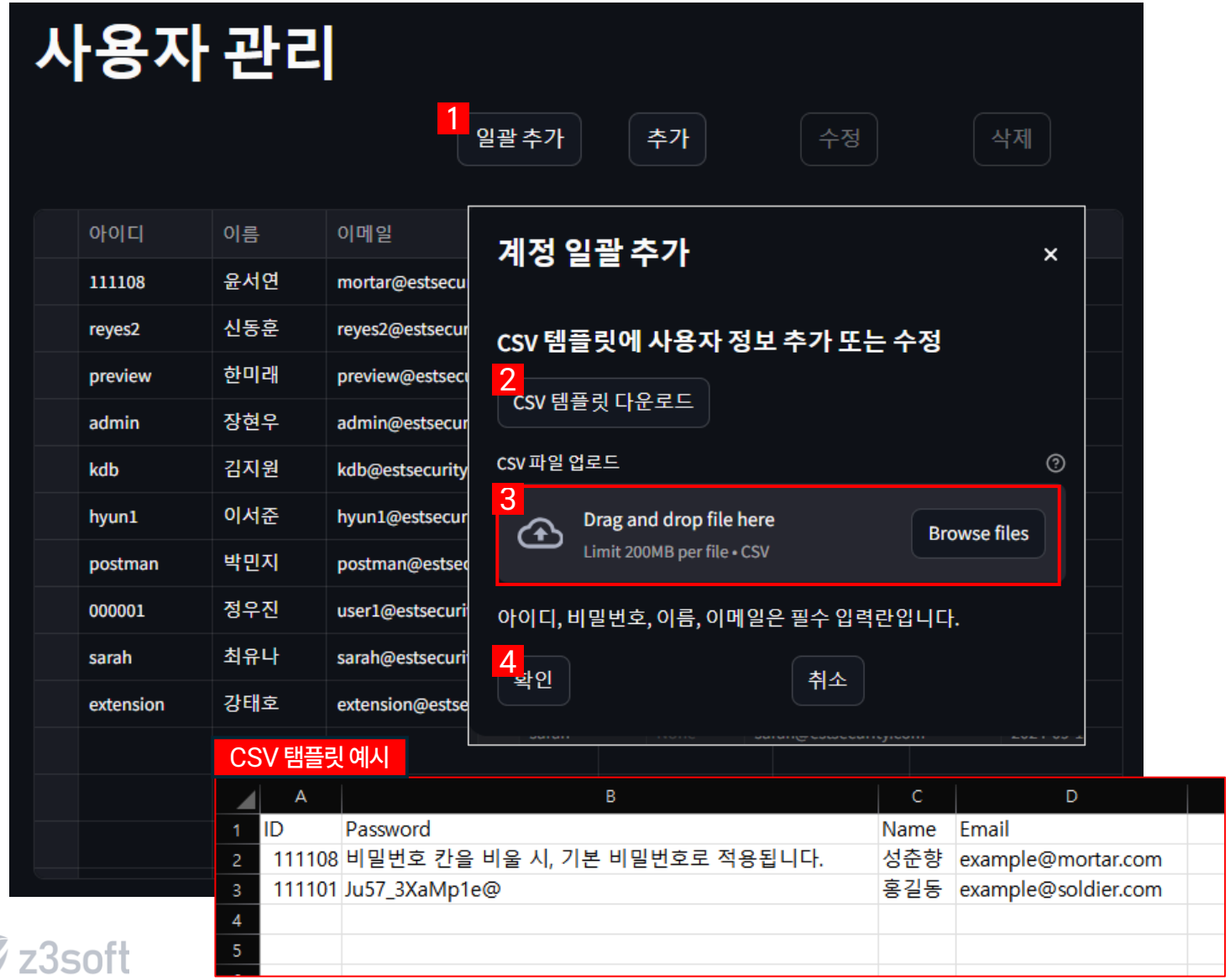
2 추가
사용자 계정을 추가할 수 있습니다.

3 수정
등록된 사용자 계정을 수정할 수 있습니다.

4 삭제
등록된 사용자 계정을 삭제할 수 있습니다.

5 등록된 사용자 계정 목록을 확인할 수 있습니다.
아이디: 사용자에게 부여된 계정입니다.
이메일: 사용자가 사용하는 이메일 주소입니다.
마지막 로그인: 사용자가 마지막으로 로그인한 날짜입니다.
총 사용량: 사용자가 입력한 Prompt 개수입니다.
계정 생성일: 사용자 계정이 생성된 날짜입니다.

사용자 관리 – 일괄 추가



1 일괄 추가
엑셀(CSV)파일을 이용하여 사용자 정보를 일괄로 추가 또는 수정할 수 있습니다.

2 CSV 탬플릿 다운로드
CSV 탬플릿 예시 파일을 다운로드 받을 수 있습니다. ID, Password, Name, Email을 모두 채워야 합니다. 이때 ID와 Email은 중복으로 사용할 수 없습니다.

3 CSV 파일 업로드
작성된 CSV파일을 업로드할 수 있습니다.

4 확인 버튼
CSV파일 업로드 후 '확인'버튼을 눌러야 정상적으로 계정이 등록 및 수정됩니다.

사용자 관리 – 추가

1 추가
사용자 계정을 추가할 수 있습니다.

2 계정 추가 팝업
아래 정보를 빠짐없이 입력해야 합니다.
이때 아이디와 이메일 주소는 중복이 불가합니다.

아이디: 사용자가 로그인할 계정을 입력합니다.
비밀번호: 사용자가 로그인할 비밀번호를 입력합니다.
비밀번호는 영문 대/소문자, 숫자, 특수문자를 포함한 8자 이상이어야 합니다.
이름: 사용자 이름을 입력합니다.
이메일: 사용자의 이메일 주소를 입력합니다.
이때 이메일 주소는 '아이디@주소'의 형태를 이뤄야 합니다. (ex. xIIm@estsecurity.com)

3 확인 버튼
계정이 추가됩니다.

사용자 관리 – 수정

1 수정
등록된 사용자 계정을 수정할 수 있습니다.
이때 계정은 한 개만 선택할 수 있습니다.

2 계정 수정 팝업
아이디는 수정이 불가합니다.
비밀번호: 수정할 비밀번호를 입력합니다.
비밀번호는 영문 대/소문자, 숫자, 특수문자를 포함한 8자 이상이어야 합니다.
이름: 수정할 이름을 입력합니다.
이메일: 수정할 이메일 주소를 입력합니다.
이때 이메일 주소는 '아이디@주소'의 형태를 이뤄야 합니다. (ex. xllm@estsecurity.com)

3 확인 버튼
계정을 수정합니다.

사용자 관리 – 삭제

1 삭제
등록된 사용자 계정을 삭제할 수 있습니다.
이때 계정은 한 개 이상 선택할 수 있으며, **삭제된 계정은 복구가 불가합니다.**

2 계정 삭제 팝업
삭제될 계정 명이 표시됩니다.

3 확인 버튼
계정을 삭제합니다.

스캐너 관리

1 스캐너 선택 목록
현재 선택된 스캐너 목록을 확인할 수 있습니다.

2 목록 펼침 버튼
선택할 수 있는 스캐너 목록을 확인할 수 있습니다.

3 정책 저장 버튼
선택된 스캐너가 적용될 수 있도록 정책을 저장할 수 있습니다.

4 Prompt 테스트/파일 첨부 테스트 탭
적용된 스캐너를 기준으로 Prompt 텍스트와 파일 첨부 내의 기밀 정보를 어떻게 탐지하는지 미리 테스트할 수 있습니다.

5 Prompt 테스트 box
Prompt에 입력할 텍스트를 넣어줍니다.

6 스캔 버튼
Prompt에 입력된 텍스트를 스캔합니다.

7 스캔 결과 box
적용된 스캐너를 기준으로 Prompt의 민감 정보를 탐지한 결과를 보여줍니다.

블랙리스트 관리

등록된 키워드는 프롬프트 수/발신이 차단되며, 사용자 전체에 적용됩니다.

1 키워드 추가
키워드 입력 후 엔터(Enter)를 눌러 키워드를 등록할 수 있습니다.

2 적용된 키워드 목록
현재 적용되어 있는 블랙리스트 키워드 목록을 확인할 수 있습니다. [x] 버튼을 눌러 등록된 키워드가 차단되지 않도록 적용 해제할 수 있습니다.

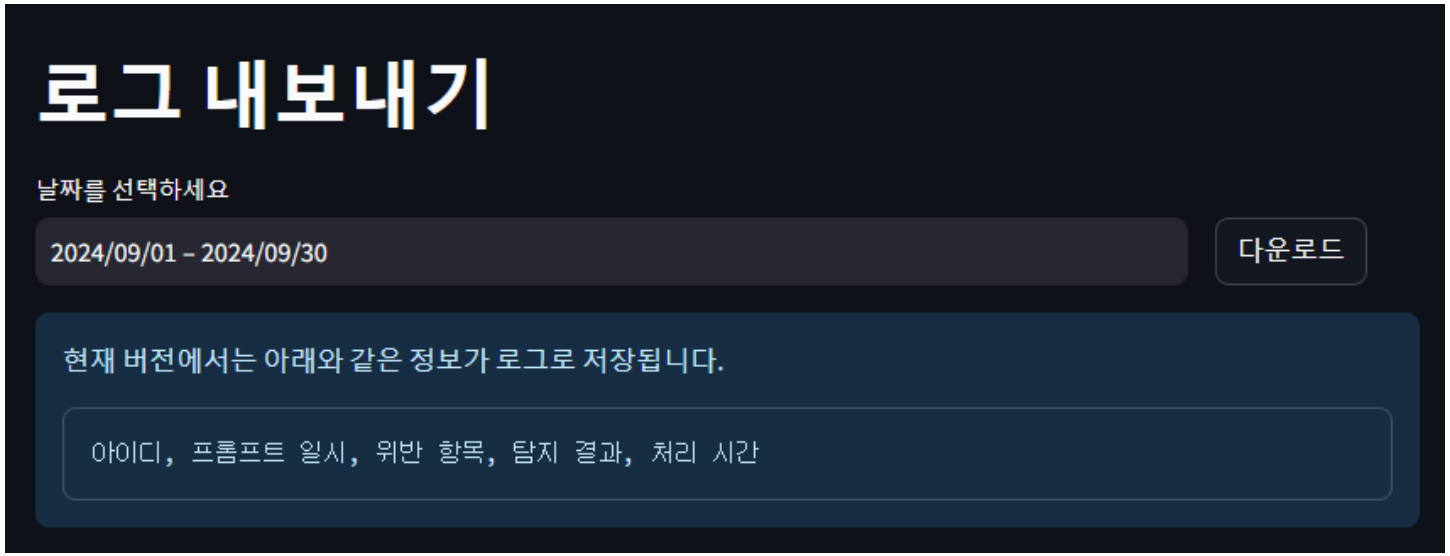
3 전체 키워드 목록
현재 적용되지 않은 블랙리스트 키워드 목록을 확인할 수 있습니다. 키워드를 클릭해 적용할 수 있습니다.

4 정책 저장 버튼
키워드 목록을 저장하고 적용할 수 있습니다.

04.

로그

로그 내보내기



로그 예시

1	admin	2024-09-05 04:54:56.327978+00:00	pii_info	[KR_SSN] 아래 내용을 참고해서 부드러운 느낌의 이력서 초안을 작성해줘. 이름은 [PERSON]입니다. 홈페이지는 estsecurity.com이며, 현재 알약 LLM Security 제품을 담당하고 있습니다. 제 핸드폰번호는 [PHONE_NUMBER]이며, 주민등록번호는 [KR_SSN]입니다. 현재 [LOCATION]에 거주하고 있습니다.	Nones
1	admin	2024-09-05 04:54:56.515638+00:00	pii_info	[KR_SSN] 아래 내용을 참고해서 부드러운 느낌의 이력서 초안을 작성해줘. 이름은 [PERSON]입니다. 홈페이지는 estsecurity.com이며, 현재 알약 LLM Security 제품을 담당하고 있습니다. 제 핸드폰번호는 [PHONE_NUMBER]이며, 주민등록번호는 [KR_SSN]입니다. 현재 [LOCATION]에 거주하고 있습니다.	Nones
1	admin	2024-09-05 04:56:31.778823+00:00	pii_info	SLXadmin1! [LOCATION][LOCATION]	Nones

선택된 기간의 로그를 csv파일로 다운로드 받을 수 있습니다.

1 **날짜 선택**:
로그를 보고자 하는 날짜를 선택할 수 있습니다.
날짜는 최대 1년 선택 가능합니다.

2 **로그 예시**
아래 항목에 대한 내용을 확인할 수 있습니다.
아이디: 사용자가 로그인하는 계정입니다.
프롬프트 일시: 해당 프롬프트가 수/발신된 일시입니다.
위반 항목: 프롬프트가 위반한 항목을 표시합니다.
개인 식별 정보 위반: 개인식별정보(PII)를 포함한 모든 개인 정보가 프롬프트에 포함된 경우입니다.
악성 콘텐츠 위반: 악성 URL이나 코드 등 보안에 위협이 될 수 있는 콘텐츠가 포함된 경우입니다.
기밀 정보 위반: 회사 또는 조직의 비공개 정보가 포함된 경우입니다.
온라인 자격 증명 노출 위반: Git 토큰이나 API key 같은 온라인 서비스에 접근할 수 있는 자격 증명이 포함된 경우입니다.
유해 표현 정책 위반: 불쾌하거나 공격적인 표현이 포함된 경우입니다.
탐지결과: 프롬프트에 탐지된 민감 정보를 익명 처리한 결과를 볼 수 있습니다.
처리 시간: 프롬프트에서 민감 정보를 탐지하기까지 걸린 시간을 확인할 수 있습니다.

05.

서버 트러블슈팅

API 서버 문제 및 대응

1. 상태 확인

1-1. systemctl status xllm_api
- disable → 서버가 죽은 상태
  서비스 재시작: systemctl restart xllm_api
- enable → 정상, 그래도 장애상황이면 서버는 살아있으나 응답을 제대로 보낼 수 없는 상황, journalctl -exf로 로그 추가 확인

2. journalctl -exf 확인

2-1. 서버가 올라온 직후에 /sanitizer/health_check 에 get 요청했을 시에 변화가 없으면 정상. (로그를 꺼놓음) 에러가 뜬다면 로그가 찍힘.

2-2 . DB가 내려갔을 때
에러문: connection to server on socket "/var/run/postgresql/.s.PGSQL.5432" failed: No such file or directory
해결: sudo docker restart xllm_posgres (혹은 sudo docker ps -a 했을 때 나오는 posgres 이름)
확인: - 로그인 시도: sudo docker exec -it xllm_posgres psql -U kotraxllm xllm
- DB로그 확인: sudo docker logs xllm_posgres
다른 DB 에러는 직접 문의.

2-3. 변화가 없거나, 일반적인 로그에서 계속 멈춰있는 경우 -> 외부 AI모델 다운로드 시도, 계속 멈춘 상태가 됨.
에러문: 딱히 없음. DEBUG 레벨로 설정해야 보임.
해결: xllm_app/xllm_api/config.yaml 파일의 input_scanners에 ScanHate 부분을 주석처리(3줄)
이후 서비스 재시작(systemctl restart xllm_api)

대시보드 문제 및 대응

1. 상태 확인
   1-1. systemctl status xllm_dashboard
   - disable → 서버가 죽은 상태
     서비스 재시작: systemctl restart xllm_ dashboard
   - enable → 정상, 그래도 장애상황이면 서버는 살아있으나 응답을 제대로 보낼 수 없는 상황