



Keep
Your Value in
SaferZone

AI기반 업무환경의 MCP보안 위협과 개인정보 유출 방지 전략

Presentation : 장성민
Chief Technology Officer of SaferZone

✦ PIS FAIR 2025

Presentation Summary

MCP를 통한 개인정보 유출 위험

API Call/MCP 활용에 대한 유출 경로 분석

AI 기반 서비스를 통한 유출 시나리오

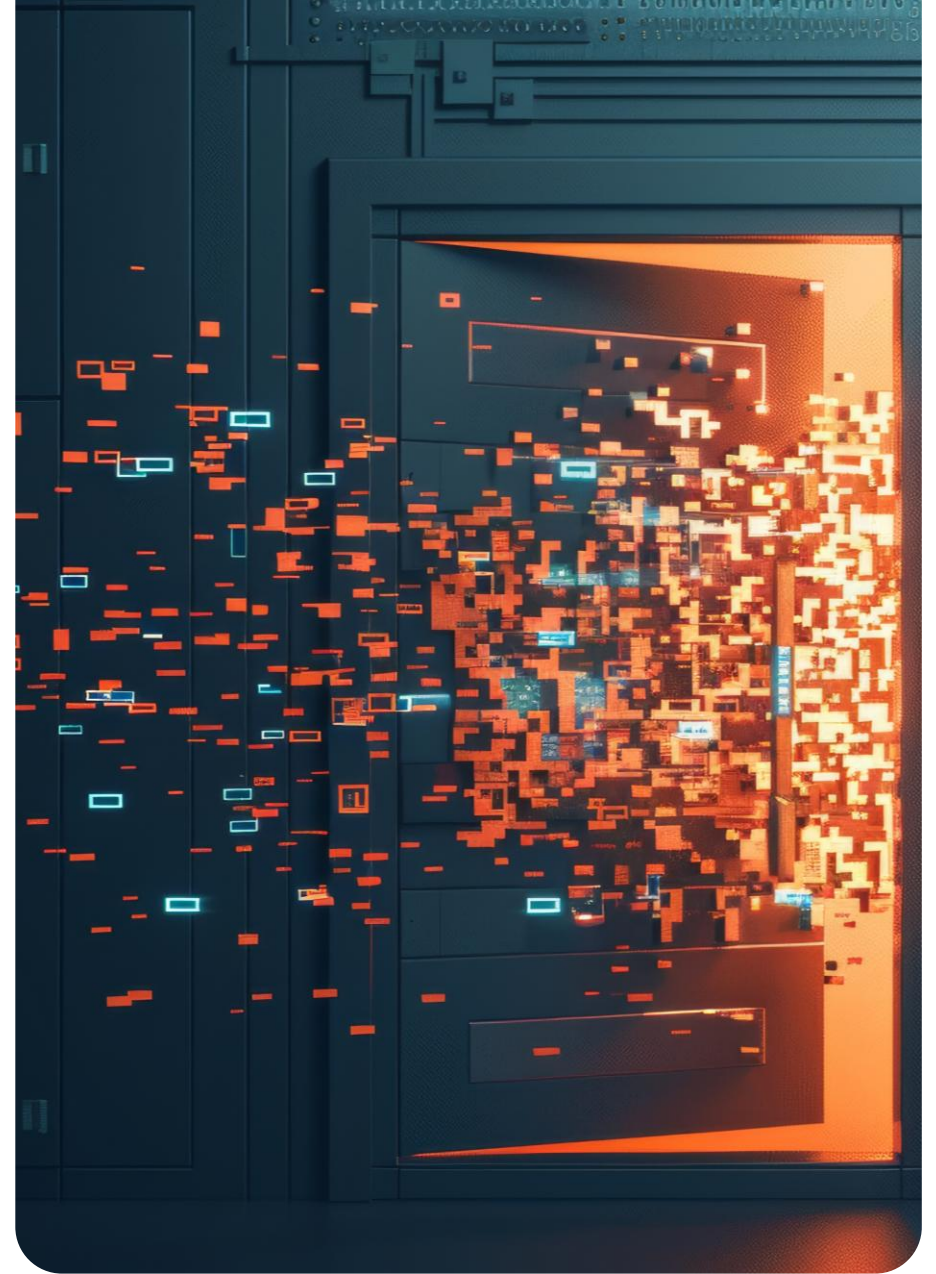
서비스 사용을 통한 유출 시나리오 정의, 기타 유출 경로

정보 유출 대응전략

Zero-Trust 기반 접근 방식

전략 실행 방안과 기대 효과

핵심 가치, Zero Trust, 감사, 기술 적용



AI의 업무 활용 : API Call, Prompt

AI를 통한 업무활용 사례 증가

AI를 활용하여 생산되는 업무데이터 증가

AI를 활용한 데이터 정리를 위해 API Call을 하거나 프롬프트에 직접 입력

→ 개인정보 데이터의 AI를 통한 경로 유출 확대



ChatGPT



사업장 챗GPT 허용 20일, 정보 유출 사고 3건 발생
제조·수출 데이터, 미국 기업에 고스란히 전송...회수 불가
해당 임직원 징계...사내 전용 AI 서비스 구축 검토



이코노미스트 - [단독] 우려가 현실로...삼성전자, 챗GPT 빗장
풀자마자 '오남용' 속출

Model Context Protocol : 새로운 활용처?

Anthropic 에서 표준화한 MCP

로컬 환경에 Server 형태의 플러그인을 설치 및 실행

AI 에이전트가 로컬에 설치된 MCP Server 를 통해 File I/O 를 수행하여 로컬데이터 접근

→ 로컬 데이터에 직접 접근하므로 새로운 유출 경로가 발생

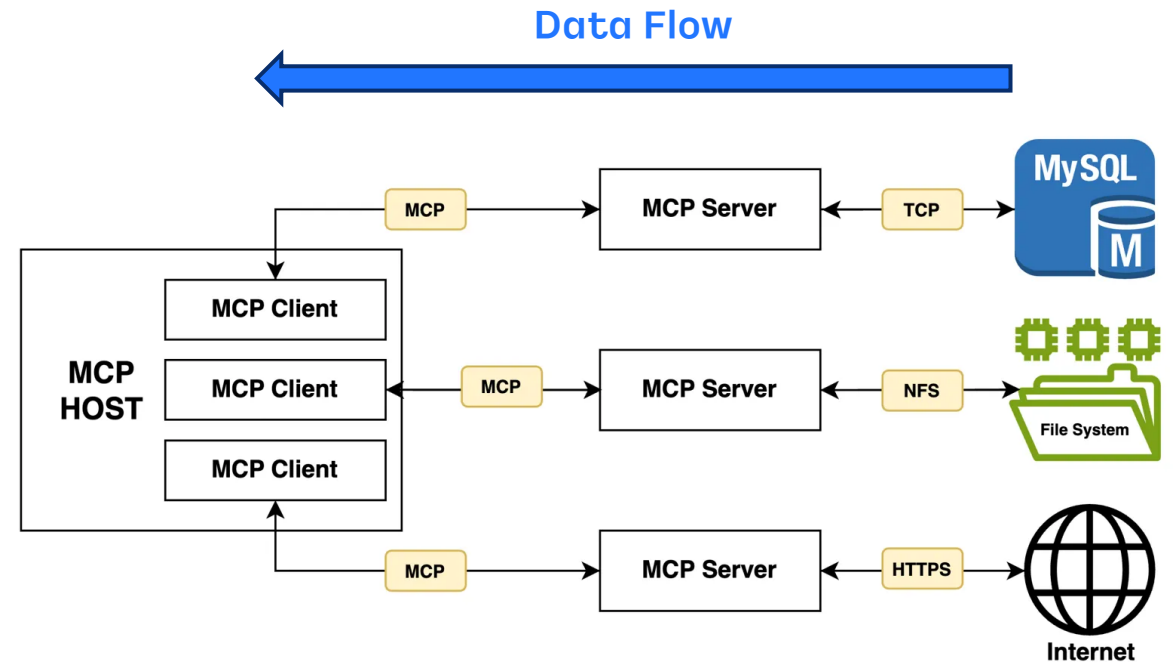
로컬데이터 무단 접근에 대한 통제가 어려움

개인정보 처리시스템 개발환경의 경우 선별적 통제가 어려운 부분이 발생함

→ MCP Server에서 사용하는 Node.js, Python에 대한 제약

개인정보 생산 시 임의 접근할 수 있는 MCP Server 이 존재함

→ 일괄 통제하지 않으면서 개인정보에 대한 핀셋접근이 필요



Model Context Protocol : 새로운 활용처?

smithery.ai

Global 에 등록된 MCP Server를 활용 할 수 있는 Community
Local에서 접근가능한 MCP부터 Remote MCP 까지 다양한 Server 공유

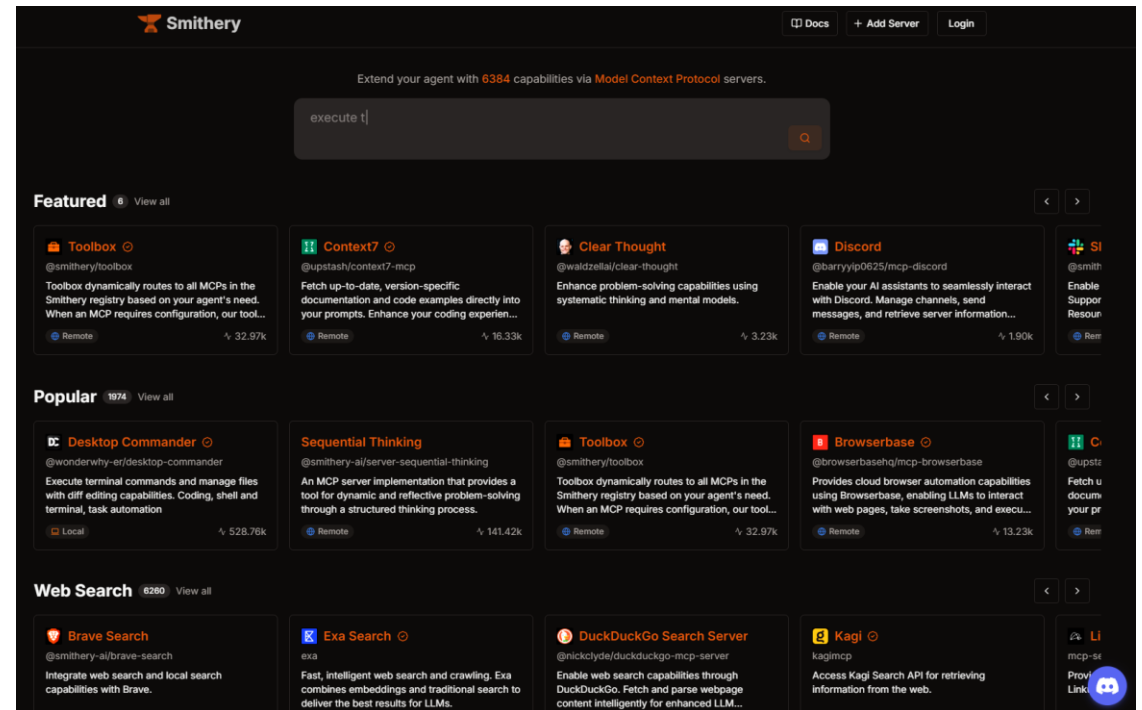
→ 출처가 불분명한 개인이 만든 MCP도 업로드 가능한 문제 존재

→ Local Filesystem에 접근 가능한 MCP Server 도 수백가지에 이름

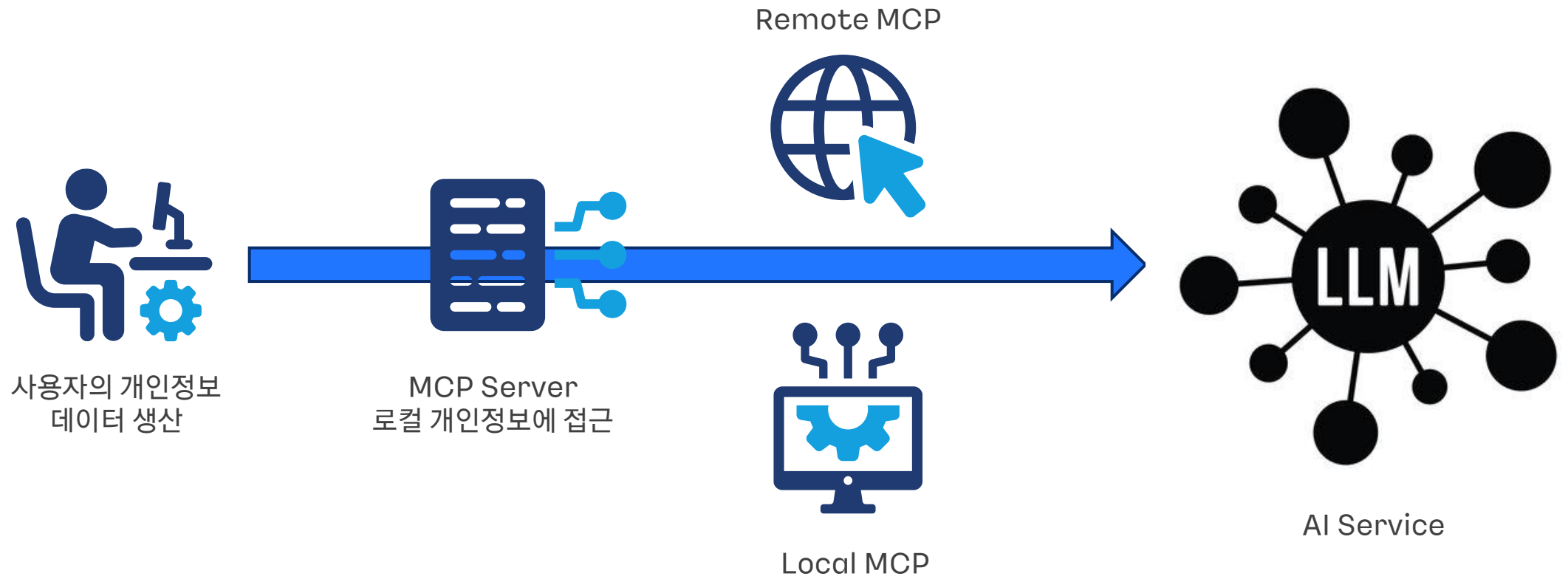
생산성 향상이 최고일까?

생산성 향상이 증가하는 것은 사실이나 보안통제범위가 불분명함

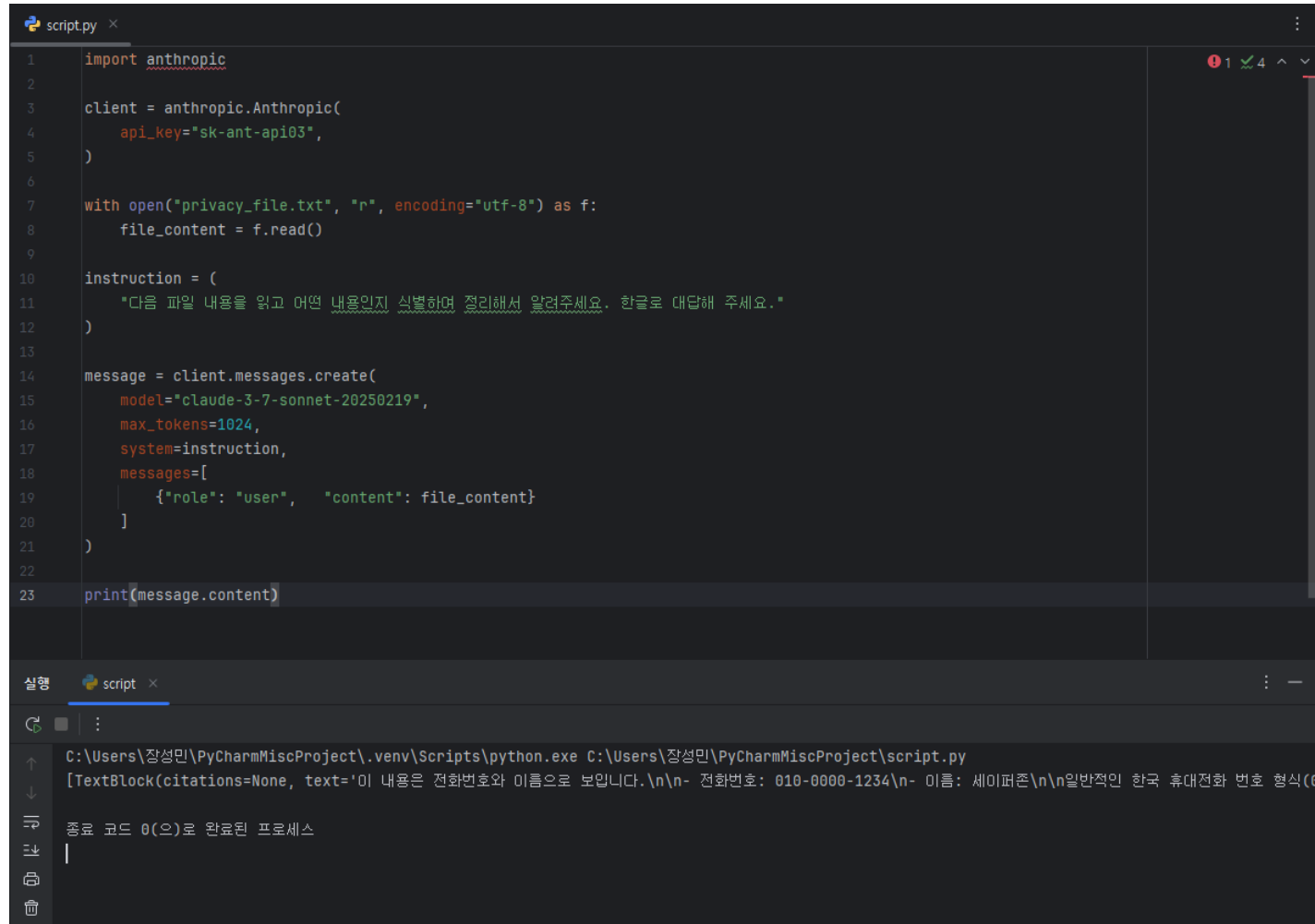
→ 어떤 기능을 사용해서 어떻게 유출되는지 확인하기 어려움



유출시나리오 : MCP



유출시나리오 : API Call



The image shows a screenshot of a code editor with a Python script named 'script.py' and its execution output in the console.

```
1 import anthropic
2
3 client = anthropic.Anthropic(
4     api_key="sk-ant-api03",
5 )
6
7 with open("privacy_file.txt", "r", encoding="utf-8") as f:
8     file_content = f.read()
9
10 instruction = (
11     "다음 파일 내용을 읽고 어떤 내용인지 식별하여 정리해서 알려주세요. 한글로 대답해 주세요."
12 )
13
14 message = client.messages.create(
15     model="claude-3-7-sonnet-20250219",
16     max_tokens=1024,
17     system=instruction,
18     messages=[
19         {"role": "User", "content": file_content}
20     ]
21 )
22
23 print(message.content)
```

The console output shows the execution of the script, displaying the content of 'privacy_file.txt' as a text block. The output is in Korean and includes a phone number and a name, which are redacted with underscores.

```
실행 script ×
C:\Users\장성민\PyCharmMiscProject\.venv\Scripts\python.exe C:\Users\장성민\PyCharmMiscProject\script.py
[TextBlock(citations=None, text='이 내용은 전화번호와 이름으로 보입니다. \n\n- 전화번호: 010-0000-1234\n- 이름: 세이퍼존\n\n일반적인 한국 휴대전화 번호 형식(0
종료 코드 0(으)로 완료된 프로세스
```

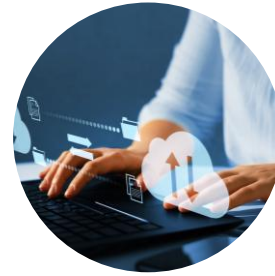
유출시나리오 : N가지의 경우



USB를 통한 유출



Email을 통한 유출



클라우드를 통한 유출

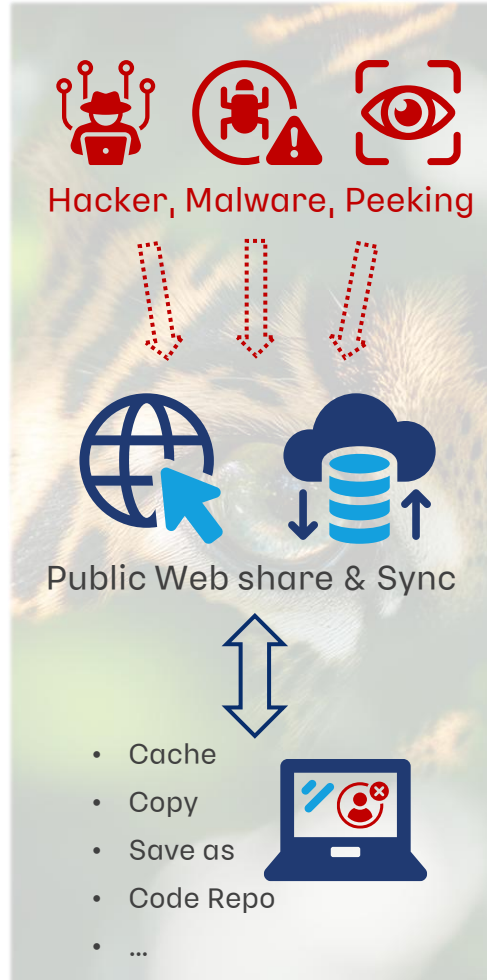
다양한 유출 경로

유출 우회방법은 매우 많음

전문적인 유출 방법이 아닌
대부분 일반인이 행할 수 있는
방법과 경로로 유출됨

일원화된 개인정보 처리가 필요함

유출시나리오 : N가지의 경우



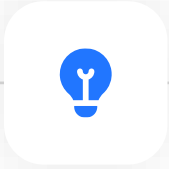
우리는 "보통"
사용자이고 싶다.

하지만,

개인 PC에서 업무를 수행하여 감염에 의한 유출
클라우드 접근 권한 설정 미숙으로 데이터 임의 공개
캐시된 화면을 통해 고의로 정보 유출
내부 코드 저장소를 악용한 유출
임의로 저장한 업무파일 무단 복사 등...

보안 관리자의 업무

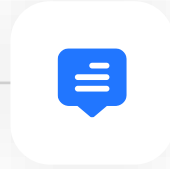
보안 의식 부족



정보보호 인식 부족

직원들의 정보보호 인식과 책임감 결여로
인한 유출 발생, 보안 교육의 부재

정책 미비



보안정책과 절차의 느슨함

명확한 보안정책이 없거나 절차가 느슨한
상황에서 발생하는 기술 유출

기술 통제 부족



유출 통제 기술 미비

유출케이스를 정의하여 관련된 모든 통제
기술 적용 미비

지속적인 보안교육

체계적인 정책 및 절차 수립

유출 경로 트렌드 파악 후 통제기술 적용 등등등등....

개인정보 유출 대응전략

인라인 DLP

- 전사/부서/사용자별 정책에 따른 경로 통제
- AI서비스 및 모든 서비스 환경에 대한 통제
- 오프라인 환경에서도 지속 가능한 검증

실시간 탐지 및 격리

- 중요 정보 데이터 실시간 탐지
- 정보 이동 전 개인정보 데이터 자동 격리 (MCP를 통한 유출 보호)

RBAC/Zero Trust

- 사용자 권한에 따라 최종 반출 결정
- 결재에 따른 개인정보 처리
- Zero Trust 기반 사용자 증명

개인정보 암호화

- 탐지된 중요 개인정보를 암호화
- 유출에 대한 실시간 보호

All-In-One 솔루션

SafeZone

클래스 신청

보안 USB

패치웨어

비저장 장치 예외

개인정보

DLP

보유등록

오입등록

조회

개인정보 보유 등록 화면입니다.

보유 등록할 파일을 신청할 수 있습니다.

파일 목록

파일검색

파일삭제

<input checked="" type="checkbox"/>	No.	개인정보개수	개인정보유형	파일명	상세 정보
<input checked="" type="checkbox"/>	1	10	주민등록번호	주민.txt	

사용자

부서

직위

사용자ID

김영찬1

KYC_test

test1

ycyh1

결재자1

부서

직위

결재자ID

김영찬1

KYC_test

test1

ycyh1

신청 기간 (시작일)

신청 사유 (0 / 100)

2025-05-15

0시

0분

~

2025-05-30

23시

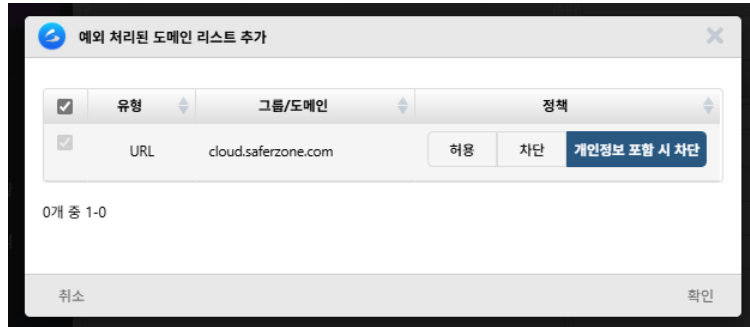
59분

신청 사유를 입력하세요.

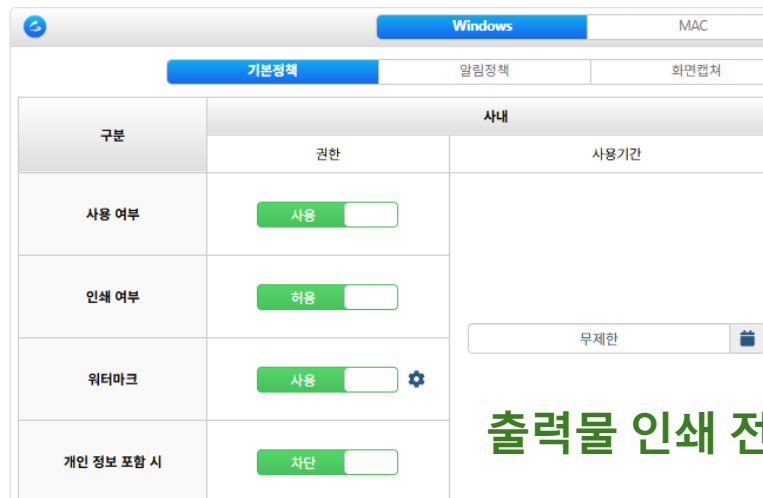
[직접 입력]



개인정보 유출 대응전략 : Zero Trust



인터넷 파일전송 시 개인정보 검사



출력물 인쇄 전 개인정보 검사

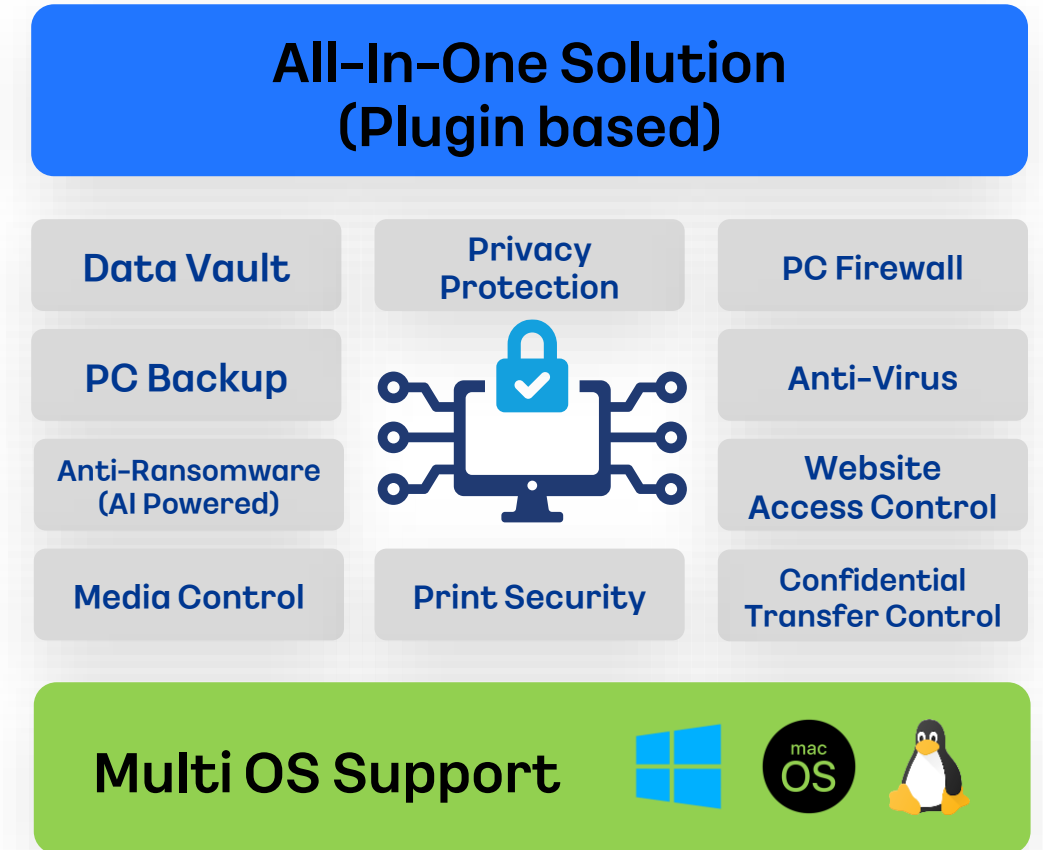


개인정보 실시간 검사 및 처리

운영 환경에서의 대응전략

통합 보안 관리 : 엔드포인트-네트워크 일괄 통제

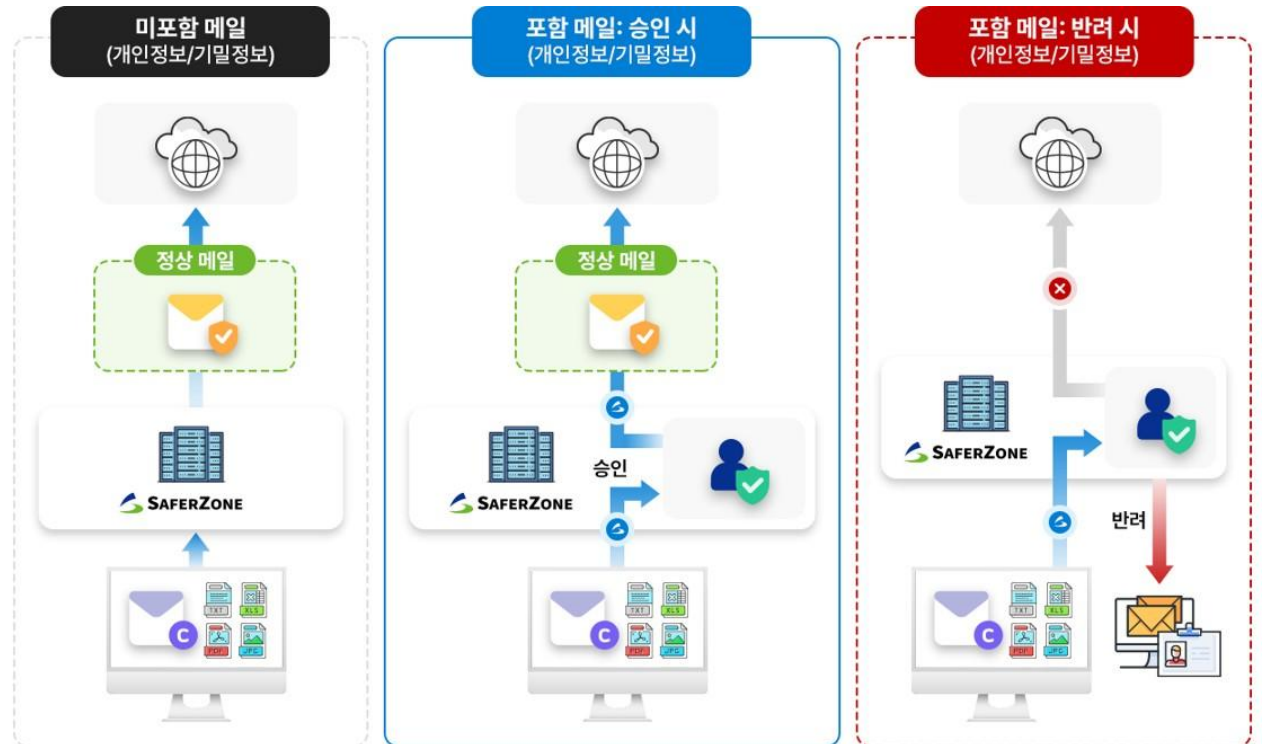
- 회사 로컬환경과 클라우드간 데이터 이동 모니터링 및 자동 차단
⇒ 이동 데이터 로그 감사 및 전사 보안정책 적용
- 개인정보 통제 일원화
⇒ 개인정보 생산자(임직원, 협력사) 구분 없이 Zero Trust 접근
⇒ OS 환경과 상관없는 일괄적인 보안 정책 적용



운영 환경에서의 대응전략

이메일 발신 및 외부 반출 승인 강화

- 외부로 나가는 이메일의 경우 첨부파일 및 본문 검사
⇒ 보안 정책상 민감 정보 판별 후 권한자에게 승인
- 개인정보 반출 시 다단계 승인 절차 도입
- 로그 감사를 통해 주기적인 내부 보안정책 모니터링



개인정보 유출방지 실행 로드맵

1단계 : 데이터 식별 및 분류

보호가 필요한 자산 식별,
데이터검사 후 기밀도에 따라 분류 체계 수립
⇒ 고객 정보, 설계도, 소스코드 등 기밀 분류

2단계 : 권한 부여 및 접근 관리

Role에 따른 폴더/파일 접근 권한 제한
외부 반출 권한 대외 업무 이외 최소화
⇒ 최소 권한 원칙 적용

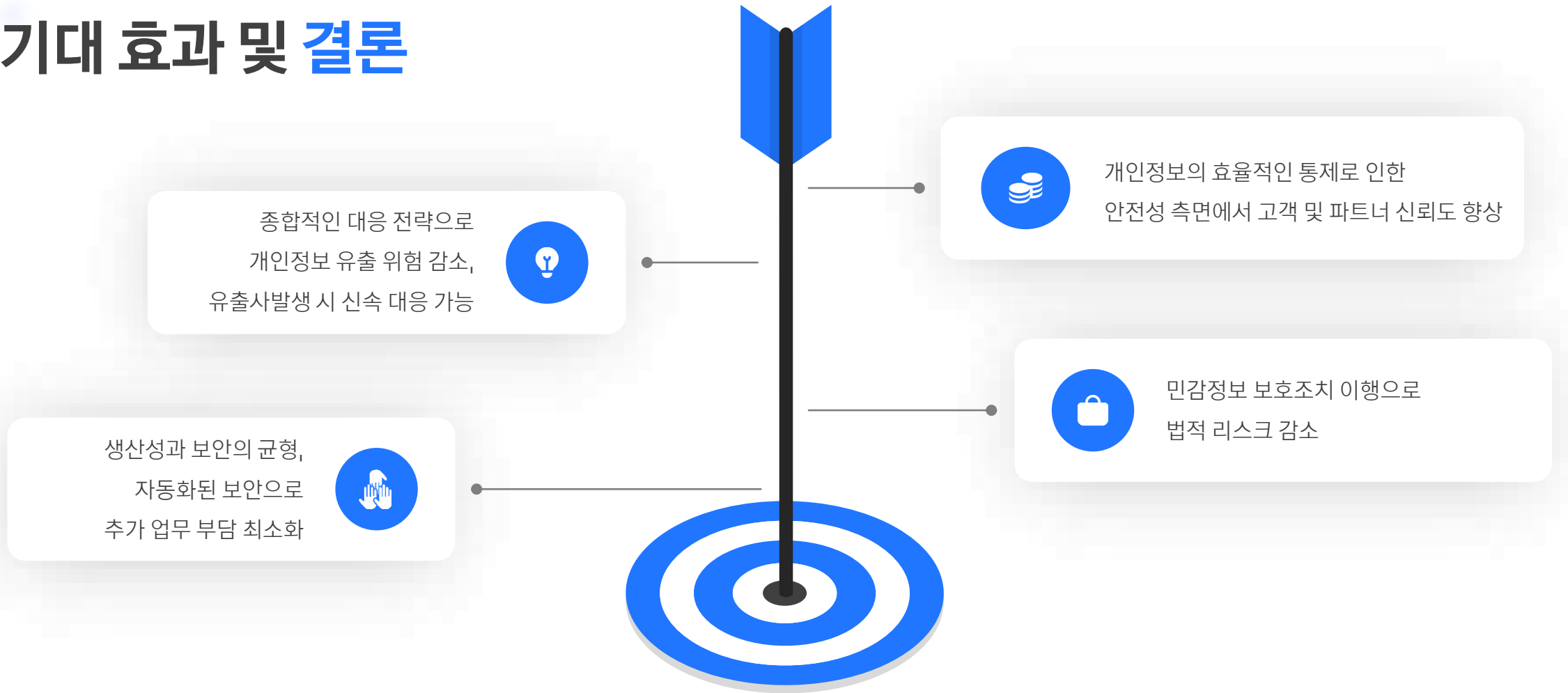
3단계 : 모니터링 및 자동 차단

매체 삽입, 대량 파일 이동, 이메일 첨부 감사
정책 위반시 알림 및 즉시 자동 차단
⇒ 실시간 이상 행동 감시

4단계 : 파일럿 운영 후 확대

특정부서, 그룹 대상 시범 운영
문제점 보완 후 전사로 점진적 확대
⇒ 사용성과 보안성의 균형

기대 효과 및 결론



"개인정보 **Zero-Leak** 구현"

→ Keep Your Value In SaferZone

Thank You

Email : sales@saferzone.com

TEL : 1577-3110