인공지능 시대의 사이버보안 패러다임 변화

2018. 4. 26(목)

정보보호R&D기술공유센터

강필용 센터장

KISA
Korea Internet & Security Agency

목차

1 사이버보안 대응환경 변화

사회 전반의 ICT 융합 가속화

※ 스마트시티 · 자동차 · 의료 · 공장 · 농장 등 전통산업과 ICT 융합으로 다양한 서비스 등장

신규기기(IoT) 연결로 보안위협 대상의 폭증 및 복잡

※ 인터넷에 연결되는 기기가 기하급수로 증가

사이버보안을 위한 전통적 대응방식은 한계에 도달

※ 유사/변종 악성코드 급증(일평균 190만개, AV-TEST 2017) 등으로 기존 시그니처(탐지규칙) 기반 탐지로는 대응불가

참고

하루에 발생한 침해사고(2017년)

악성코드 활동 : 23,883건

(안랩 탐지 8,717,194건)

신규 취약점 : 4.32건

(KISA 탐지·신고·분석 1,577건)

랜섬웨어 피해 : 16건

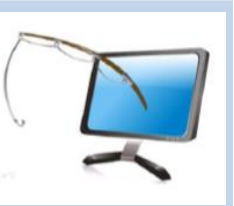
(KISA 피해상담 5,825건)

DDoS 공격 : 1.25건

(KISA 탐지·신고 551건)

홈페이지 변조 : 5건

(KISA 탐지 1,724건)



피싱·파밍 사이트 : 35건

(KISA 탐지·조치 12,683건)

홈페이지 악성코드 유포 : 37건

(KISA 탐지·조치 13,347건)

스미싱 메시지 : 1,376건

(KISA 탐지·조치 502,027건)

2.1 인공지능 혁명 : 알파고 vs. 이세돌

※ 출처 : 바둑TV

참고 알파고 시리즈의 성능비교

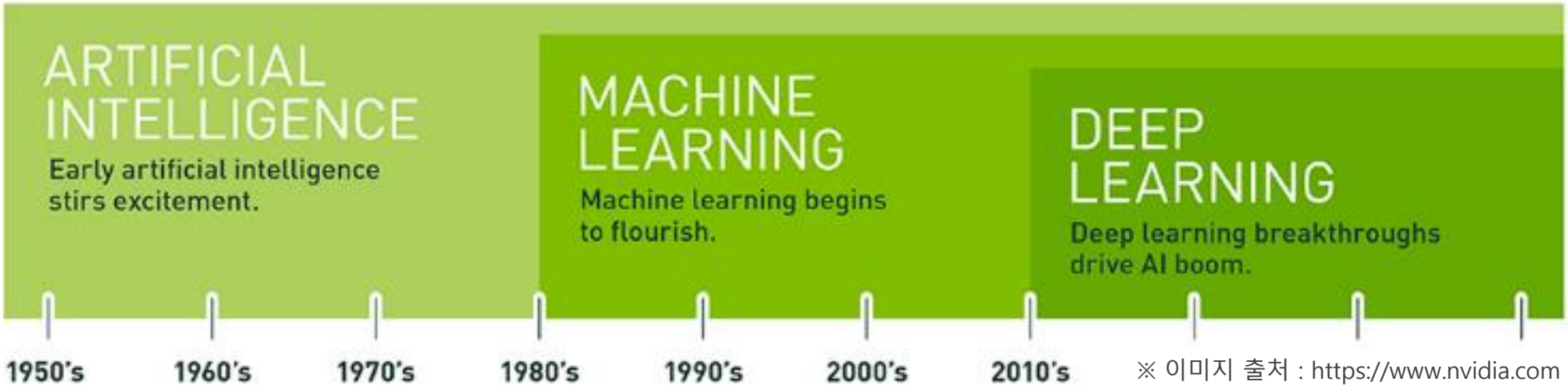
※ 출처 : 동아일보

2.2 인공지능 발자취

정의: **인간 지능**의 본질을 규명하고, 이를 **인공적으로 재현**하려는 기술

인지 · 학습 · 추론 · 행동

※ 스스로 지능을 발전시키는 **학습단계(강화학습) 수준**이며, 새로운 상황을 추론하고 행동하는 단계로 발전中



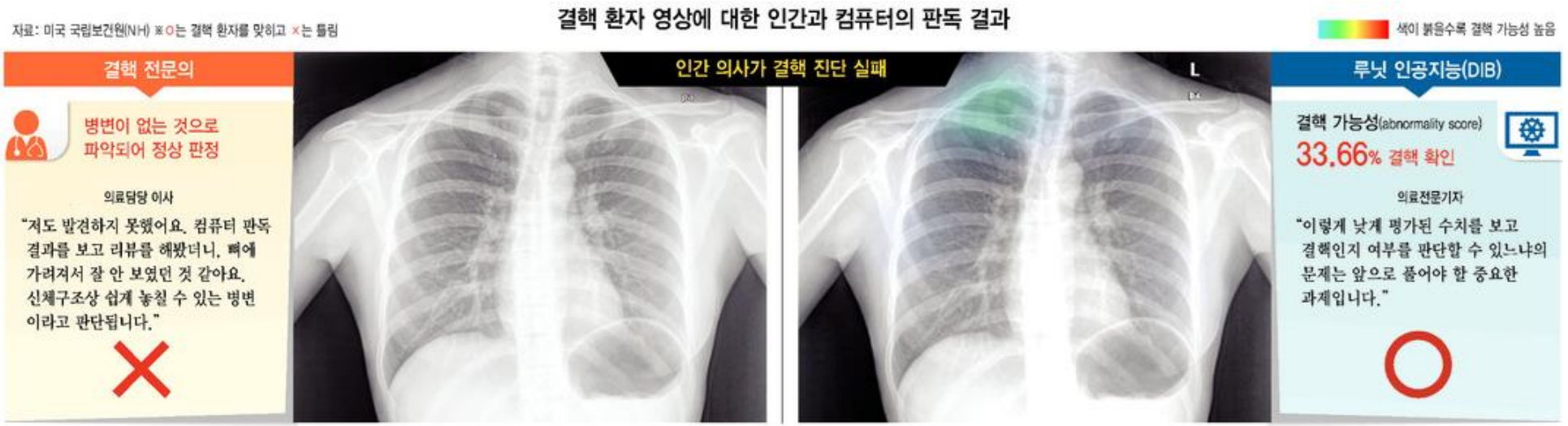
※ 이미지 출처 : https://www.nvidia.com

Othello('80) Chess('97) Jeopardy('11) Atari('15) ImageNet AlphaGo('16) Poker SkinCancer('17)

참고 전문분야 인공지능 활용사례

정보의 홍수 속에 사는 의사들의 딜레마

※ 출처 : 위키트리



※ 출처 : 한계레

참고

朝鮮日報

2018년 04월 13일 금요일 A14면 사회

1호 '인공지능 의사' 탄생… 진단서도 척척

美 식품의약국 최종 판매 승인

당뇨 망막병증 진단 의료기기
1분내 90% 가까운 정확도로 분석
실명 원인 1위 질환 조기에 발견

유방조영술부터 암치료법까지
글로벌 IT 기업들 개발 뛰어들어
"향후 많은 영역서 의사 대신"

의사 도움 없이 단독으로 안과 질환인 당뇨 망막병증을 진단할 수 있는 인공지능(AI) 의료 기기 'IDx-DR'의 모습. 인공지능 의료 기기로는 처음으로 미국식품의약국(FDA) 승인을 받았다.

의사의 도움 없이 스스로 병을 진단할 수 있는 인공지능(AI) 의료기기가 미국에서 처음 판매 허가를 받았다. 전문의처럼 환자에게 진단서를 발급할 수 있는 '인공지능 의사'가 탄생한 것이다. 세계 첫 의료용 인공지능인 IBM 왓슨이 의사를 보조해 암 진단을 했다면, 이번 인공지능 의료기기는 한발 더 나가 사람을 대신하는 단계로 발전했다.

미국 식품의약국(FDA)은 미국 의료기기업체 IDx가 개발한 안과용 인공지능 의료기기 'IDx-DR'에 대해 최종 판매 승인을 내렸다고 11일(현지 시각) 밝혔다. IDx-DR은 환자의 눈 영상을 분석해 당뇨 망막병증을 진단한다. 당뇨 망막병증은 고(高)혈당으로 인해 망막 혈관이 손상돼 시력이 떨어지는 질환이다. 심할 경우 실명(失明)까지 할 수 있다. 이 의료기기가 본격 보급될 경우 환자는 병원에서 오랜 시간 전문의 진료를 기다리지 않고 일반 의사나 간호사의 도움을 받으면서 간편하게 검사를 할 수 있을 전망이다.

당뇨 망막병증은 발병 초기에 별다른 증상이 나타나지 않아 치료 시기를 놓치는 경우가 많다. FDA에 따르면 미국에서 매년 2만4000명이 이 질환에 걸리는데, 절반 이상이 제때 검사를 받지 않은 것으로 조사됐다. 안과 전문의 진료 예약이 쉽지 않고, 대기 시간이 길어 정밀 검사를 생략하는 경우가 많기 때문이다. 당뇨 망막병증은 국내에서 실명 원인 1위 질환으로 꼽힌다.

IDx-DR은 환자의 망막 영상만으로 1분도 안 되는 짧은 시간에 당뇨 망막병증을 진단할 수 있어 이런 문제를 상당 부분 해소할 것으로 기대된다. IDx-DR은 우선 망막 디지털 카메라로 양쪽 눈의 망막 영상을 두 장씩 촬영한다. 이어 망막 이미지를 인공지능 소프트웨어에 입력해 기존 당뇨 망막병증 환자의 눈 영상과 비교해 진단을 하는 식이다. IDx는 머신러닝(기계학습)을 통해 전문의보다 빠르게 병을 진단할 수 있다고 설명했다. 질환이 발견될 경우 인공지능은 수술·치료가 필요하다는 소견과 함께 한 장짜리 진단서를 안과 전문의에게 전달한다. IDx-DR은 지난해 900명의 당뇨 환자를 대상으로 한 미국 임상 시험에서 87.4%의 정확도로 당뇨 망막병증 환자를 가려냈다. 음성인 경우 진단 정확도는 89.5%였다.

글로벌 IT(정보 기술) 기업들도 의료용 인공지능 개발에 뛰어들고 있다. 바둑 인공지능 알파고로 유명한 구글 자회사 '구글 딥마인드'는 지난 2월 눈의 영상 자료를 분석해 녹내장과 당뇨 망막병증, 시력 감퇴 등 주요 눈 질환 진단이 가능한 인공지능을 개발했다. 딥마인드는 향후 방사선 치료와 유방조영술(유방 X선 검사)까지 인공지능 진단 영역을 확대한다는 계획이다. 마이크로소프트는 인공지능을 이용한 환자 맞춤형 암 치료법을 개발하고 있다.

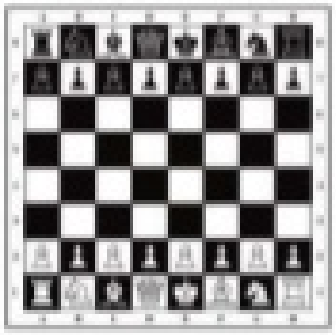
최재순 서울아산병원 의공학과 교수는 "인공지능은 아직 사람에 비해 미흡한 점이 있지만 영상을 분석하는 알고리즘이 정교해지고 있고 임상 경험을 통해 데이터가 쌓일수록 정확도가 높아지기 때문에 향후 많은 영역에서 의사를 대신해 병 진단을 할 것으로 보인다"고 말했다.

최인준 기자

2.3 인공지능 능력 및 현황

- Artificial Narrow Intelligence (ANI, Weak AI) : 인간의 지능을 모방하여 특정한 문제를 해결
- Artificial General Intelligence (AGI, Strong AI) : 인간처럼 생각(사고·감정·창의)하여 문제를 해결
- Artificial Super Intelligence (ASI, Super AI) : 인간을 초월한 지능 보유

현재



ANI: Chess Program

2040년

AGI: Human Level Intelligence

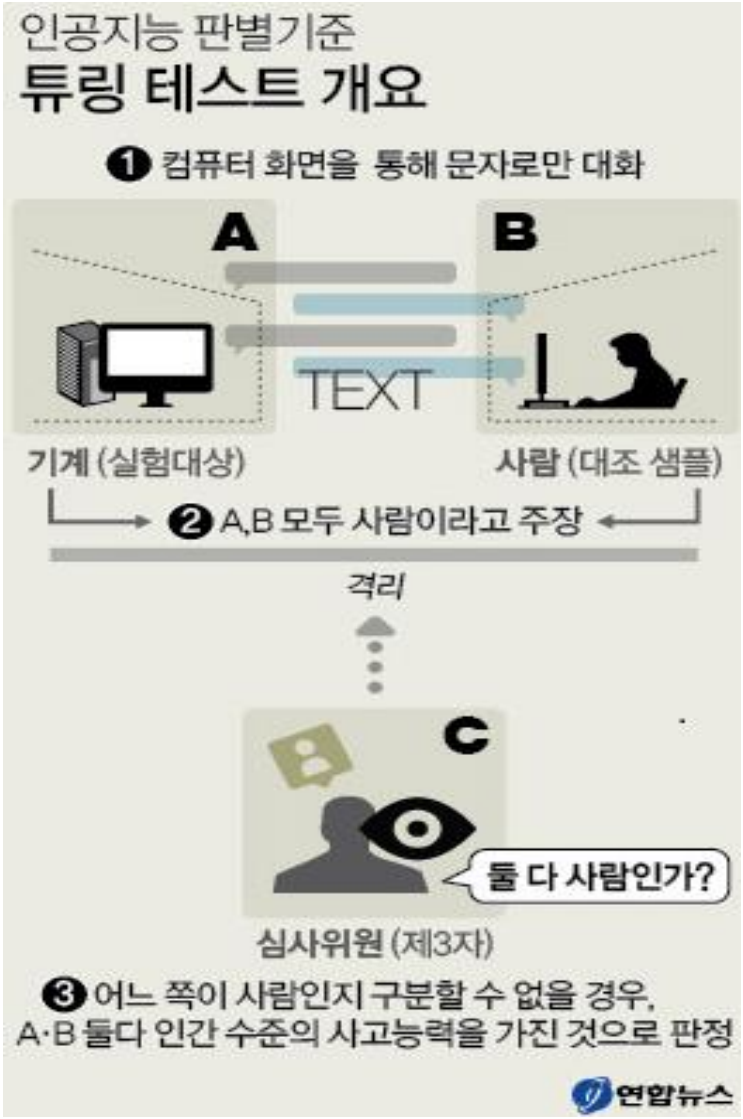
2060~70년

ASI: Superintelligence

※ 출처 : Superintelligence: Paths, Dangers, Strategies (2014)

참고

튜링 테스트를 통과한 인공지능은?



2.4 인공지능 활용분야

개인 맞춤형 상품 추천 서비스

자율주행자동차

금융·법률·의료 및 사이버보안 등
최신 정보 파악·조언

자동 통·번역 서비스

공장에서 제조공정 통합관리

인공지능 전투로봇

산불·지진·홍수 및 사이버 침해사고 등
재해·사고 예측 서비스

에너지 효율을 높인 스마트 빌딩

3.1 인공지능 기반 사이버보안 개선분야

인공지능 활용으로 **새로운 공격에 대응**하고, **단순반복 및 수작업 대체**

인텔리전스

보안위협 **정보 급증 및 복잡**

각종 보안위협
분석 및 연계를 통해
의사결정 지원

대응시간

사람에 의한 **수동 분석 및 대응**
(수주~수개월 소요)

침해사고 **대응시간 단축**

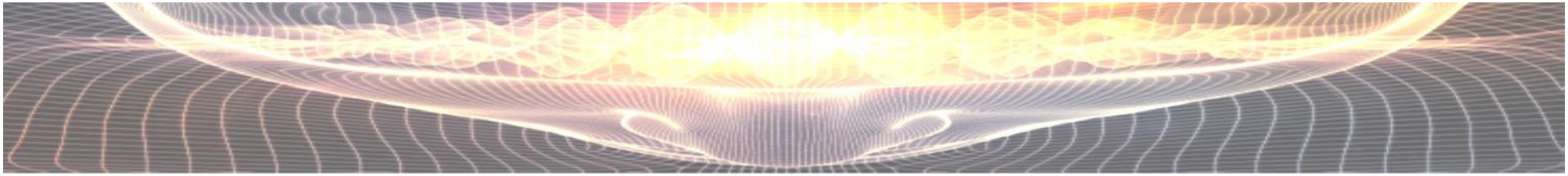
정확도

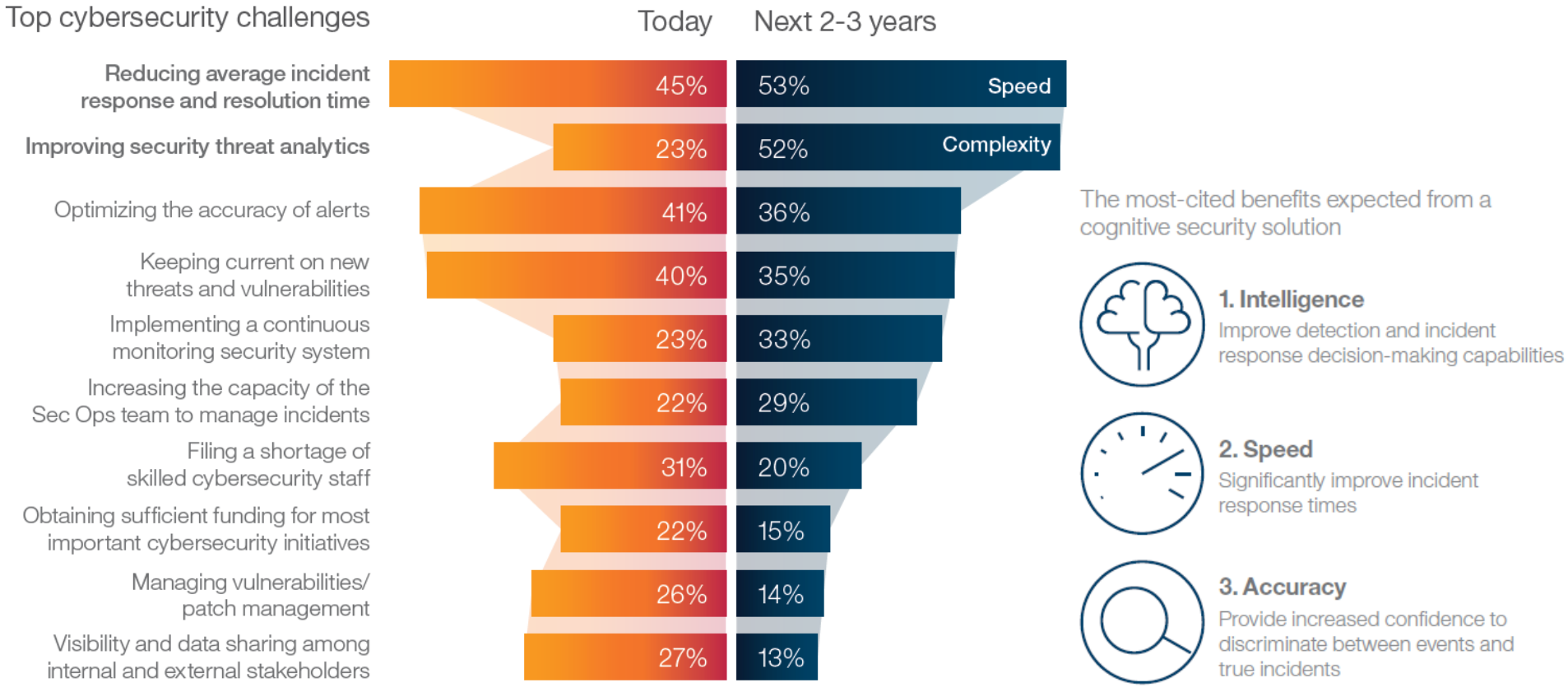
제로데이 취약점 및 변종 확산 및
규칙(시그너처) 기반 탐지 한계

이상징후·공격시도 등
탐지 정확도 향상

참고

인공지능에 기대하는 사이버보안 난제





※ 출처 : Cybersecurity in the Cognitive Era (IBM, 2016)

3.2 인공지능 기반 사이버보안 전망

23%

plan to invest in cyber AI

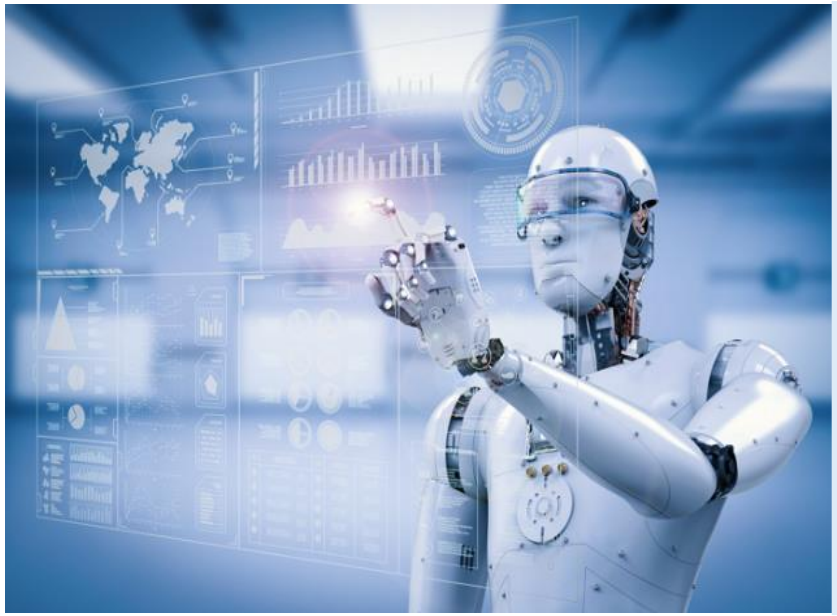
85%

of all cyber attacks predicted by using AI

36.1%

is the expected yearly market growth between 2016 and 2024 for AI

※ 출처 : Global Opportunity Report 2017



- 非정상 · 악성 행위 탐지 및 공격 저지
- 제로데이 취약점 제거
- 사람의 분석 능력 및 결과 보강
- 보안 반복작업 자동화

3.3 적용분야 : 이상금융거래탐지(FDS)

Better fraud targeting

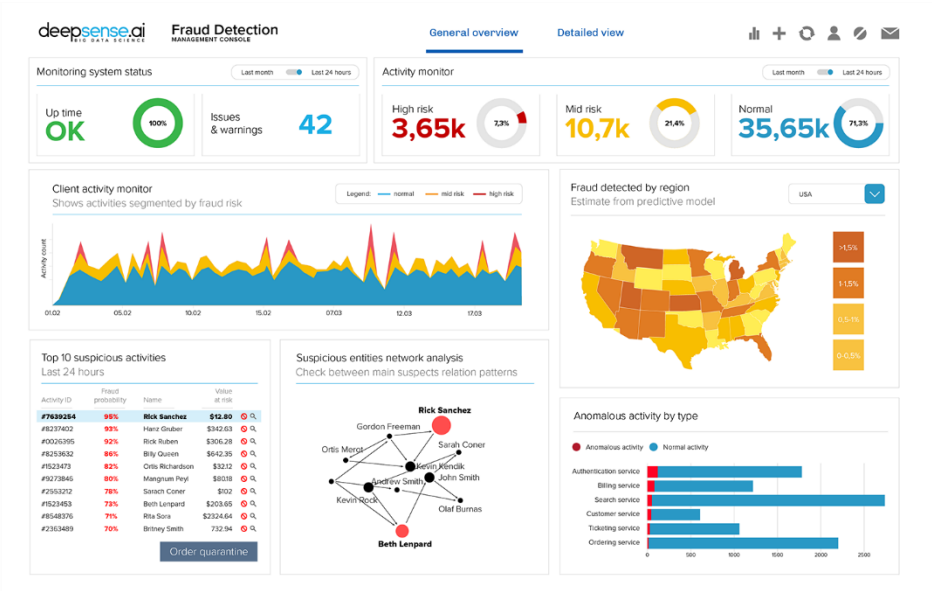
up to 50% more fraud detected

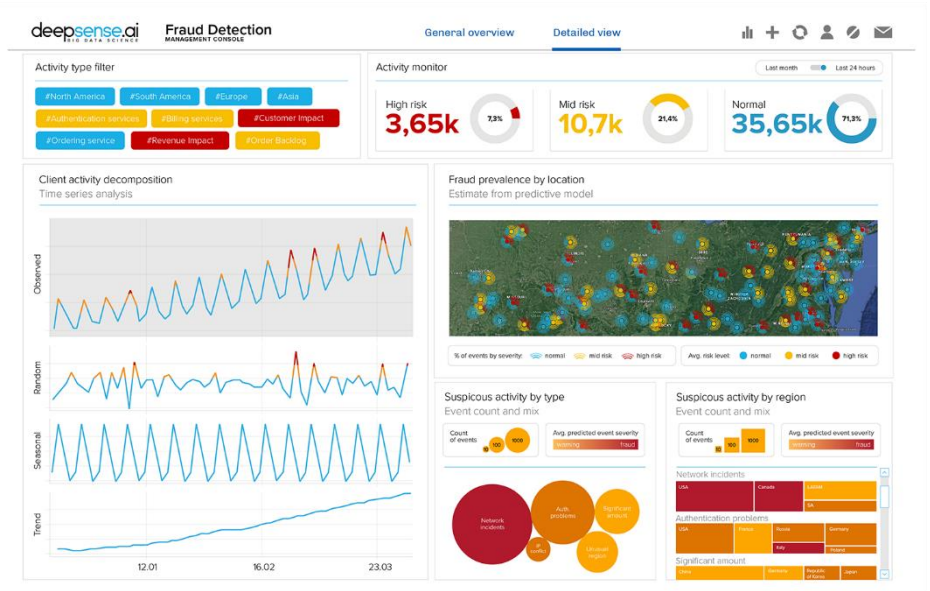
Increased organizational efficiency

by **smarter targeting of frauds**

Reduced false alarms &

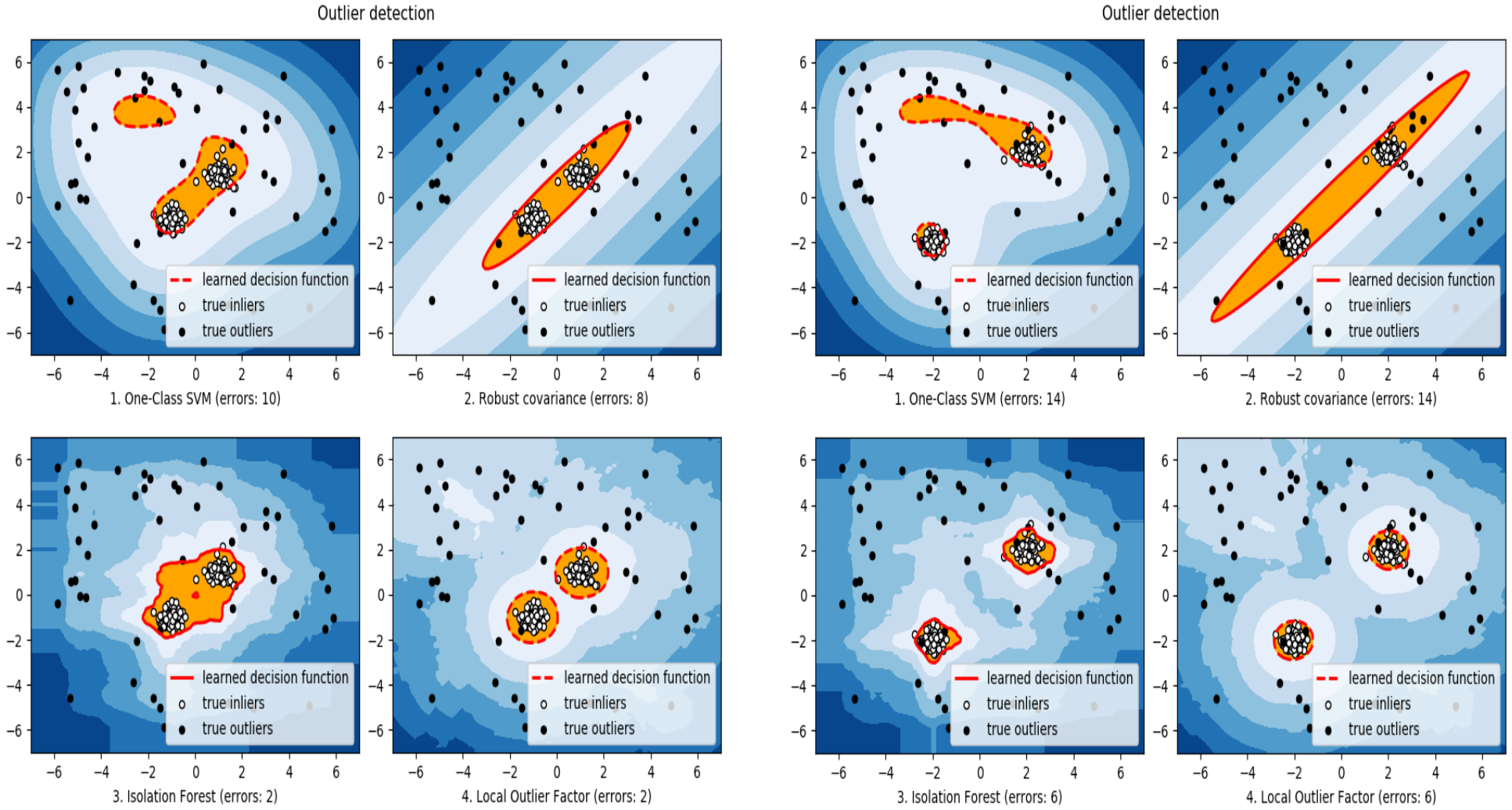
decreased cost of handling issues





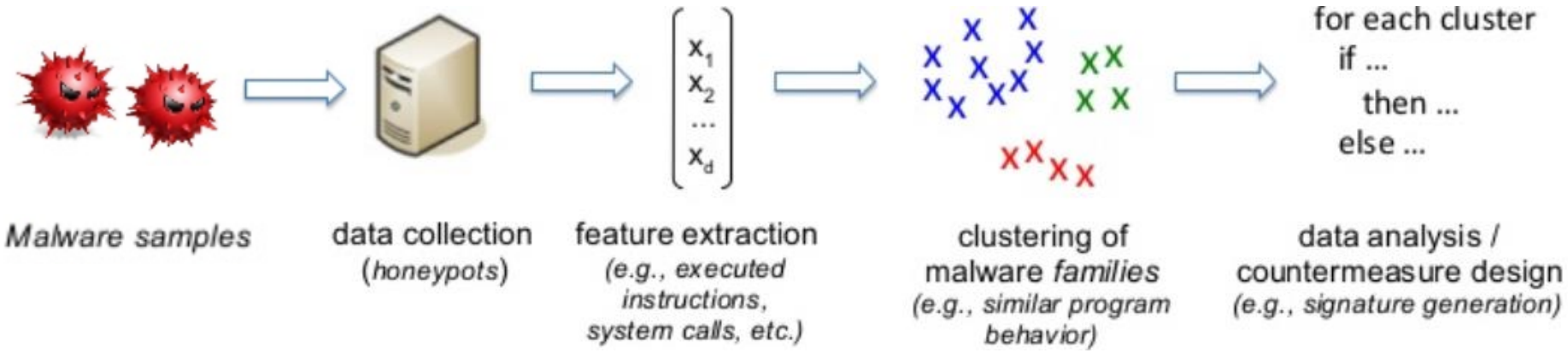
https://deepsense.ai/fraud-detection

3.3 적용분야 : 네트워크 침입탐지

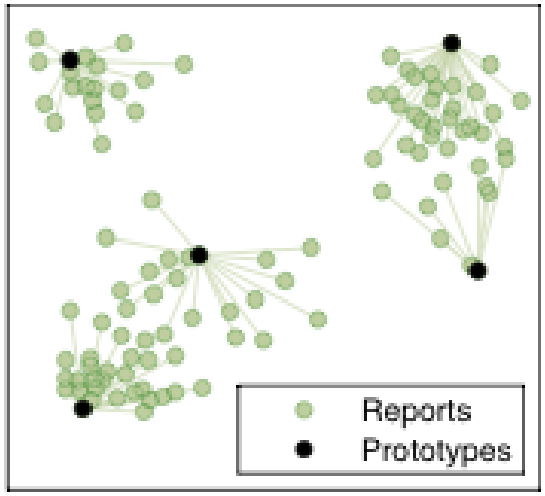


http://scikit-learn.org/stable/modules/outlier_detection.html

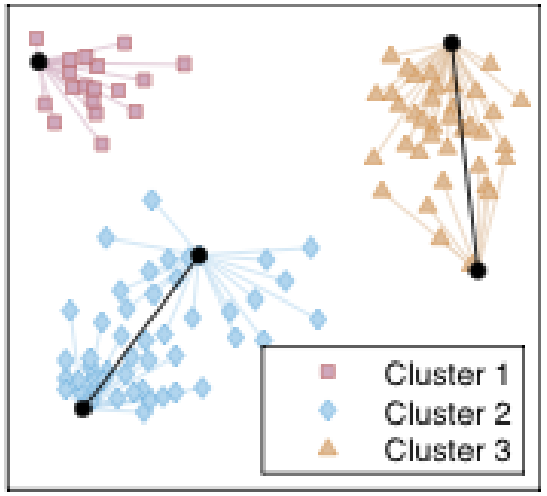
3.3 적용분야 : 악성코드 분석



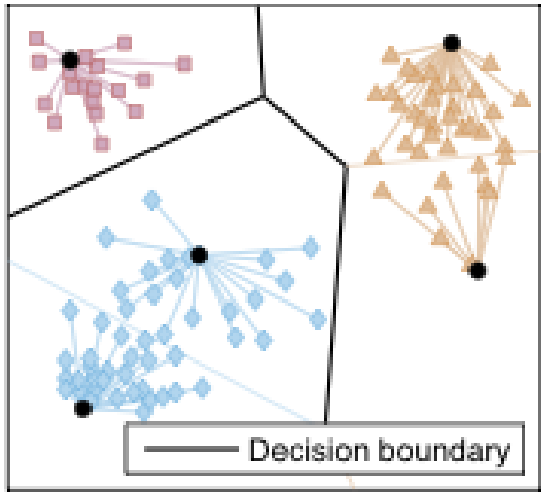
※ Poisoning Behavioral Malware Clustering (AISec 2014)



(1) Prototypes



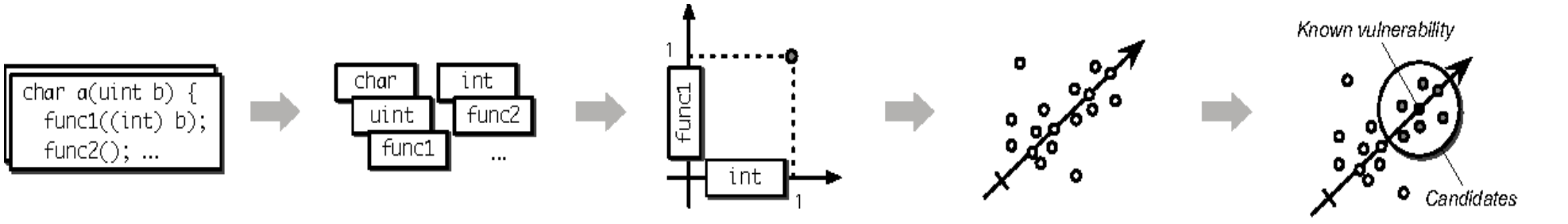
(2) Clustering



(3) Classification

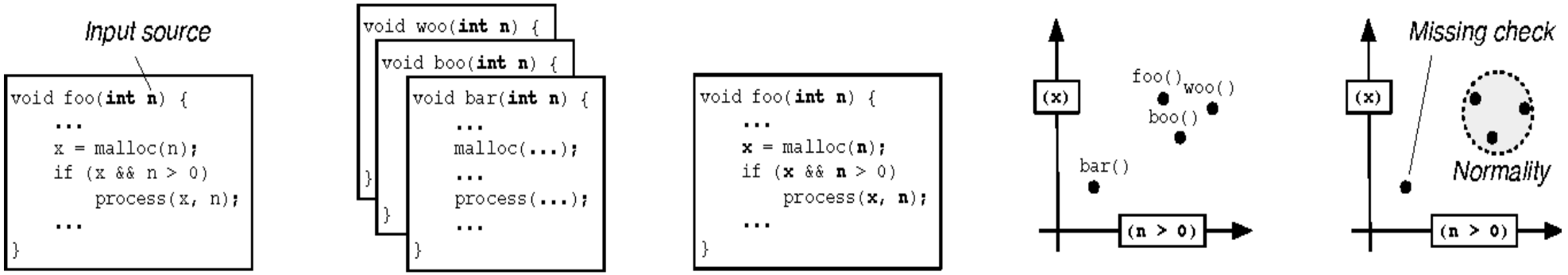
http://www.mlsec.org/malheur

3.3 적용분야 : SW 취약점 분석



Code base of functions | *(1) Extraction of API symbols* | *(2) Embedding in vector space* | *(3) Identification of API usage patterns* | *(4) Assisted vulnerability discovery*

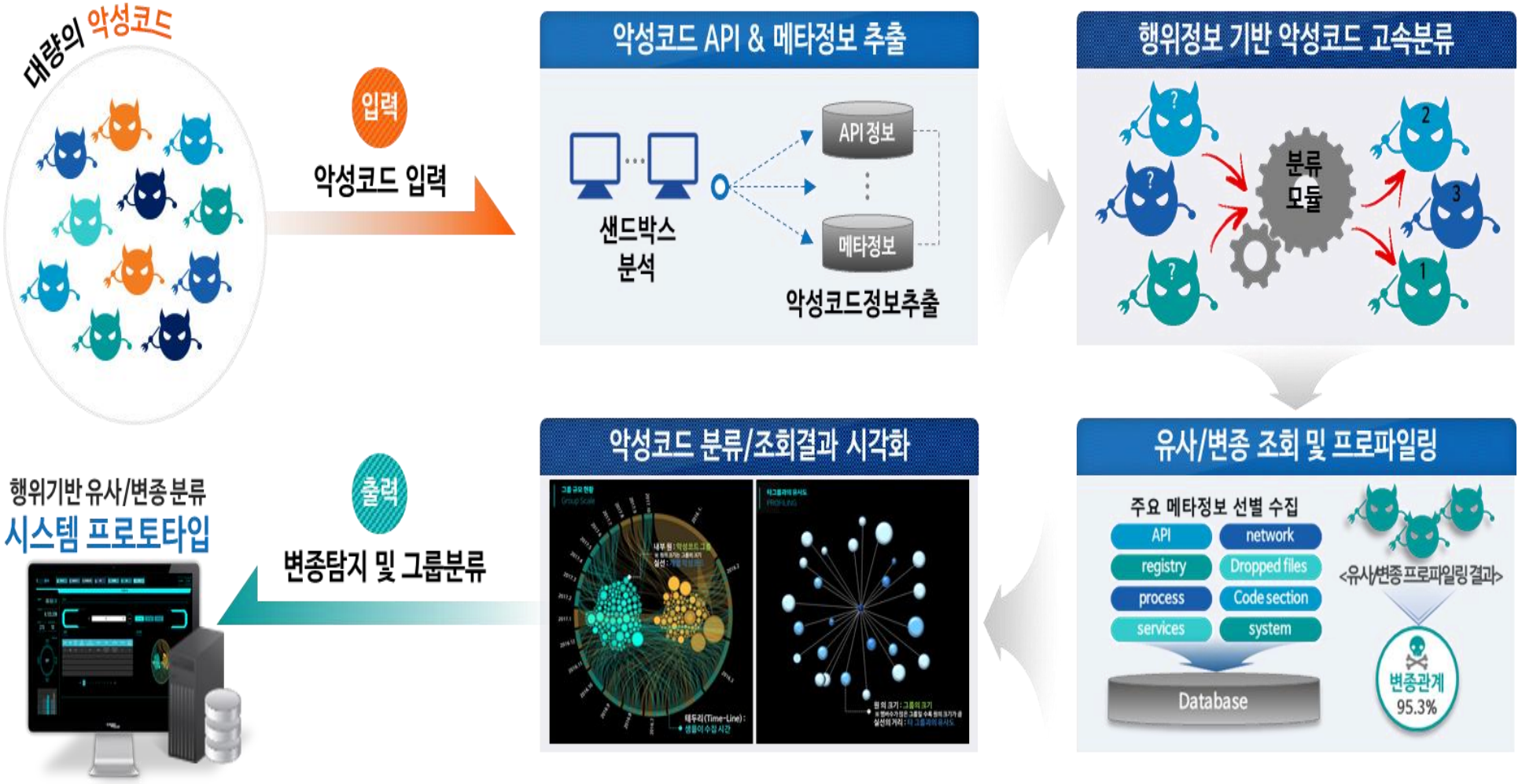
※ Vulnerability Extrapolation: Assisted Discovery of Vulnerabilities Using Machine Learning



(1) Identification of sinks and sources | **(2) Neighborhood discovery** | **(3) Lightweight tainting** | **(4) Embedding of functions** | **(5) Anomaly detection**

※ Chucky: Exposing Missing Checks in Source Code for Vulnerability Discovery

참고 행위기반 유사/변종 악성코드 분류(KISA)(1/2)

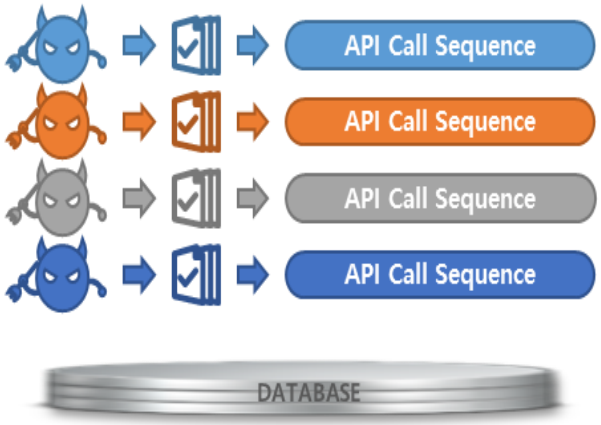


참고 행위기반 유사/변종 악성코드 분류(KISA)(2/2)

악성코드 행위정보 추출

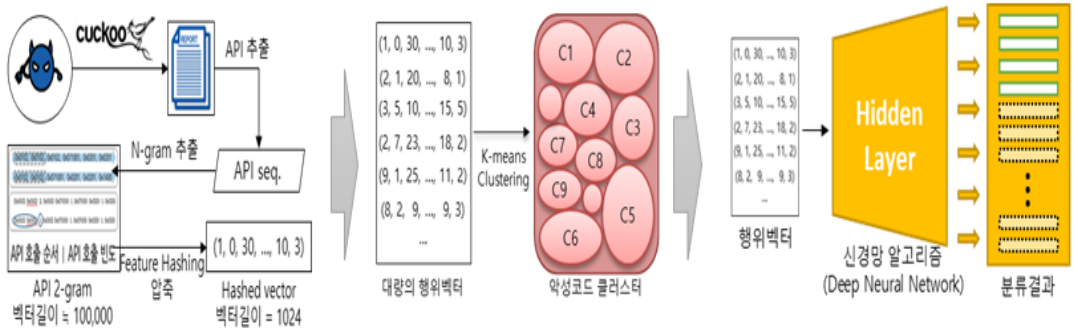
① 오픈소스 샌드박스 Cuckoo를 이용한 악성코드 정보생성

② 악성코드 고속분류를 위한 실행과정의 API 시퀀스 추출



악성코드 행위기반 클러스터링 및 고속분류

③API 시퀀스 기반 행위벡터 생성 ④유사도 기반 클러스터 생성 ⑤악성코드 분류모델 및 학습기반 고속분류



<38,984개 악성코드 분류결과>

측정 내역	측정결과
소요시간	약 10분
클러스터 수	371개
평균크기	105개
평균 유사도	98.63%

(클러스터 생성 결과)

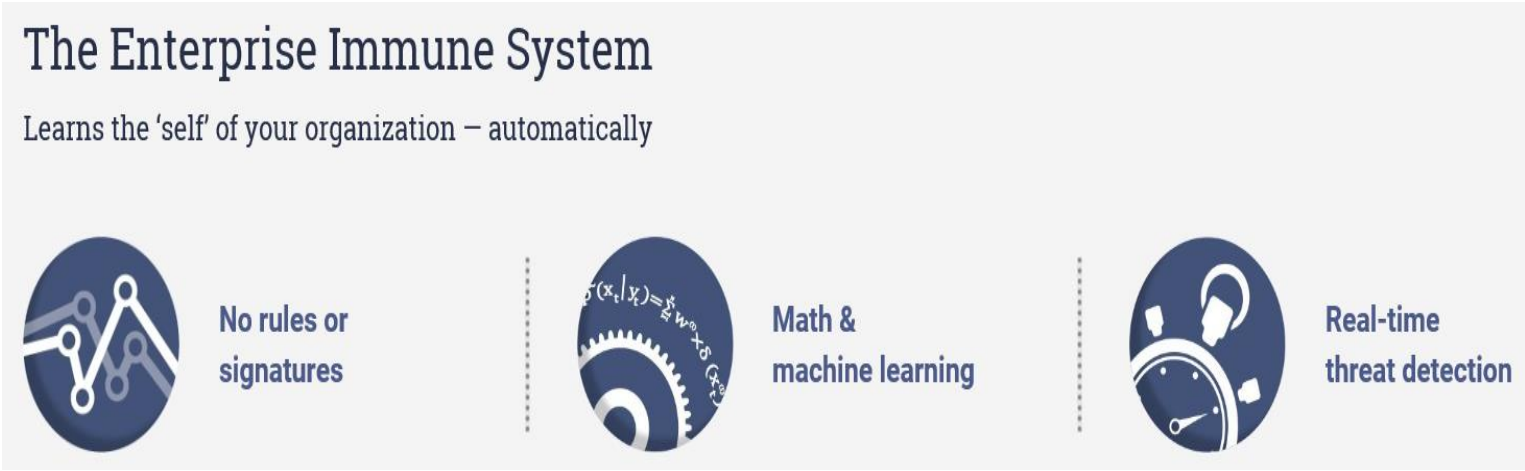
측정 내역	측정결과
소요시간	약 100분
학습회수	100 epoch
학습결과	94.57%

(악성코드 분류 결과)

3.4 인공지능 기반 보안제품(1/2)

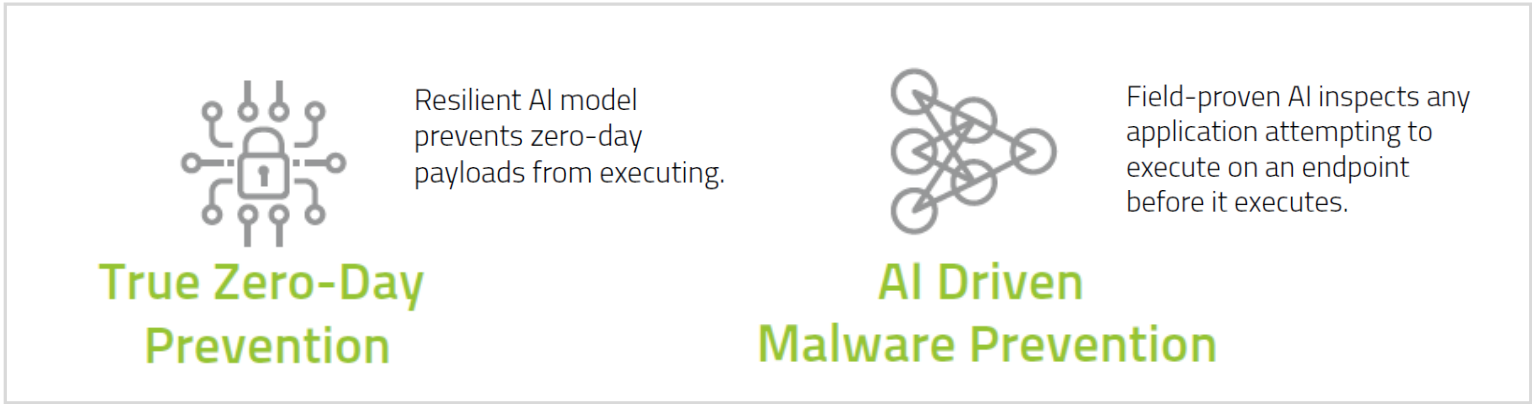
DARKTRACE

https://www.darktrace.com/

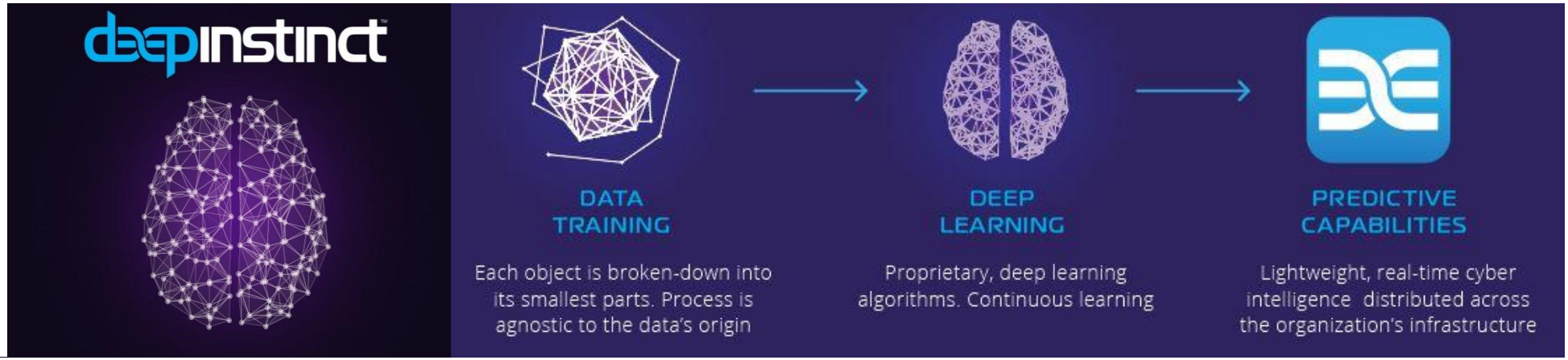


CYLANCE

https://www.cylance.com/



3.4 인공지능 기반 보안제품(2/2)



https://www.deepinstinct.com

https://www.ibm.com/security/cognitive

3.5 인공지능 보안이슈

※ 출처 : Explaining and Harnessing Adversarial Examples (ICLR 2015)

※ 출처 : Practical Black-Box Attacks against Deep Learning Systems using Adversarial Examples (2016)

3.6 제약사항 및 선결과제

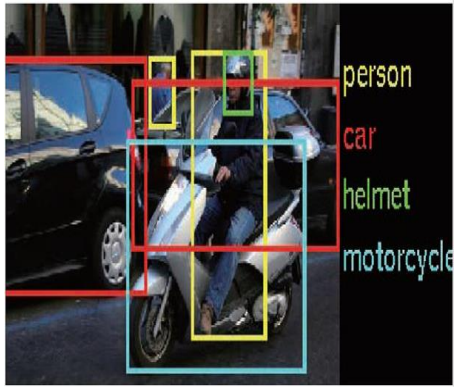
사이버보안 분야는 딥러닝 적용 초기단계로 인공지능 활용의 성패는 양질의 학습데이터와 알고리즘이 좌우

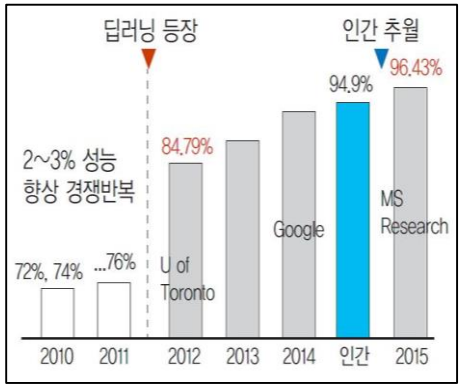
✓ **데이터 측면 : 학습을 위한 양질의 Data Set 확보에 애로**

※ 기밀, Privacy 특성으로 분석에 적합한 Labeled 데이터는 비공개

✓ **알고리즘 측면 : DNN for Security 알고리즘 연구는 기초 수준**

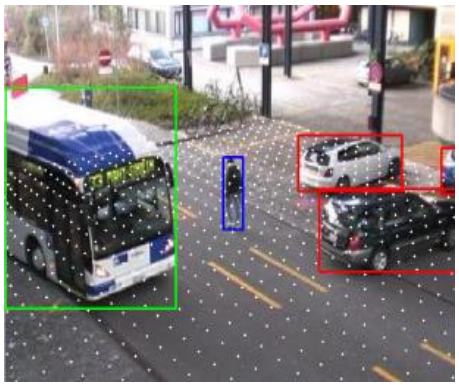
※ 현재 보안에 적용된 알고리즘은 시각/자연어처리 최적 모델





※ 시각분야 ImageNet 챌린지 대회를 통해 알고리즘 및 성능이 획기적 개선
('06년 딥러닝 알고리즘 발표후 제품화까지 5~10년 소요)

* 자율주행차



* 지능형 CCTV

참고 데이터 공유·확산 추진

악성코드 탐지
(1위팀 탐지율 88.66%)

악성 앱 탐지
(1위팀 탐지율 87.24%)

차량 이상징후 탐지
(1위팀 탐지율 91.44%)

악성코드 선제대응

(예선) '17.11.6~29 (본선) '17.12.8 (참여) 181개팀 372명

4 맺음말

- 現인공지능 수준은 다양한 **전문분야 업무지원 가능**
- 사이버보안 분야에 **인공지능(머신러닝) 적용시도 활발**

  ※ 인텔리전스 강화, 대응시간 단축 및 정확도 개선 등 기대
- 양질의 **학습데이터 확보** 및 보안특화 **알고리즘 개발** 중요
- **인공지능 오염 및 우회** 등 대응기술 부각 예상

감사합니다