AIS 2023

2023 사이버위협 대응 인공지능 보안 컨퍼런스

AI-Powered Customizable Anti-Malware Solution

CEO 정성균
AI Forensics Lab
Meta Forensics. Co., Ltd.
www.metaforensics.ai

About Me

정성균 | CEO
Meta Forensics Corp.
Email: sk.jung@metaforensics.ai

Digital Forensic Research Center (2016.03-2018.02)

- 고려대학교 정보보호대학원 정보보호학과 공학석사 졸업
- 다양한 파일 시스템, 데이터베이스, 파일 포맷 분석 전문성 획득
- 대검찰청 및 전국 검찰청 공식 수사 포렌식 도구 개발 및 납품 경험 보유

INetCop (2018.05 - 2020.01)

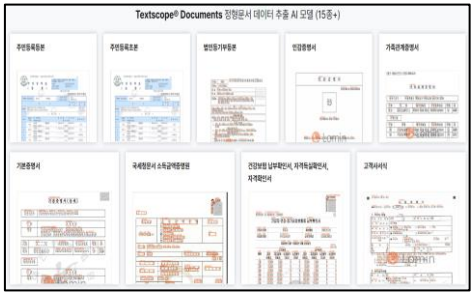
- 인공지능 기반 PE 악성코드 탐지 대회 2회 연속 우승
- 안드로이드(APK, ELF) 악성코드 탐지 엔진 개발 경험 보유
- 글로벌 주요 AV 테스팅 기관(US, UK, DE, CN) 대상 모두 만점 달성

Lomin (2020.02 - 2022.03)

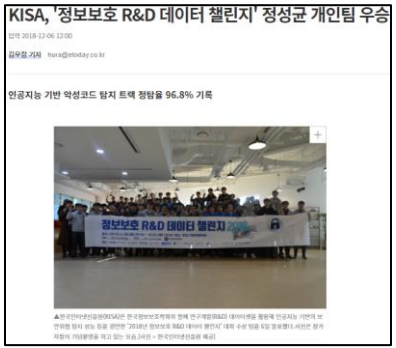
- 인공지능 스타트업 초창기 멤버 참여 경험 보유
- 자연어 처리 및 컴퓨터 비전 등 핵심 AI 기술에 기반한 제품 개발 경험 보유
- 업계 최고 모델 퍼포먼스 달성 및 스타트업 최초로 대형 금융/보험사 대상 AI BtoB 솔루션 납품 경험 보유

Meta Forensics (2022.04 - 현재)

- IITP '인공지능 기술 활용 디지털 증거 기법 개발' 위탁 과제 수행
- KISTI '소리 데이터 기반 사회 문제 해결을 위한 요소 기술 연구' 자문
- KISA '23년 침해사고 탐지정보 분석·대응 지원 사업' 자문
- 맞춤형 안티멀웨어 엔진 생성 플랫폼 'Code Semantics' 연구 개발



출처: https://lomin.ai/textscope-studio/

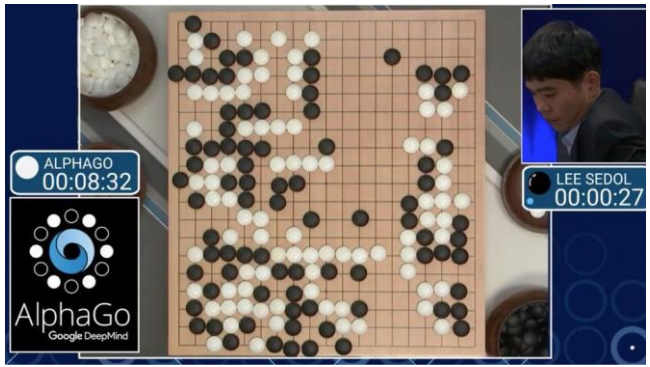


출처: https://www.etoday.co.kr/

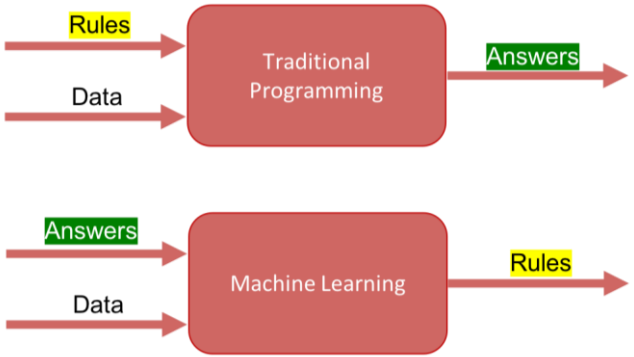
출처: https://www.aitimes.kr/

About artificial intelligence

- **도메인 지식의 한계 초월에 따른 발전 가능성**
  - 전통적인 프로그래밍(휴리스틱 등)은 전문가의 지식을 직접 코드에 반영해야 함
    - 하지만, 바둑과 같은 문제는 복잡성 때문에 모든 경우의 수를 계산하거나 개개인의 전략을 코드로 전환하는 것은 불가능에 가까움
  - 인공지능 기술은 특정 분야의 지식을 직접 가지고 있지 않아도 충분한 데이터와 학습 알고리즘이 있으면 다양한 도메인에서 높은 성능을 발휘할 수 있음
  - 바둑의 규칙 정도만 아는 개발자 조직이 세계 최고의 바둑 기사를 압도하는 자동 프로그램을 만들 수 있음
  - 즉, 인공지능을 통해 **특정 개인 및 조직이 가진 도메인 지식의 한계를 극복할 수 있고, 이는 기술 발전의 가능성이 무한함을 시사**



출처: https://www.bbc.com/news/technology-35785875



출처: https://datalya.com/blog/machine-learning/machine-learning-vs-traditional-programming-paradigm

About artificial intelligence

- **도메인 지식의 한계 초월에 따른 발전 가능성**

의료 AI 마침내 전문의 판독 능가…정확도 19% 더 높아

이인복 기자 | 발행날짜: 2022-02-14 12:13:45 업데이트: 2022-02-15 09:10:42

X레이 사진 2364개 대상 전문의 5명과 AI 비교 연구
머신 러닝 시스템 92% 정확도 기록…전문의 77.5%

[메디칼타임즈=이인복 기자] 머신 러닝을 통한 의료 진단 인공지능(AI)이 전문의 5명의 교차 진단보다 더욱 우수한 정확도로 질환을 진단하는데 성공했다.

엑스레이(X레이) 사진 2364개를 대상으로 골절 유무 진단을 맡긴 결과 전문의의 교차 진단 정확도는 77.5%에 그친 데 반해 AI는 92%로 무려 19%나 높게 나타난 것.

머신 러닝을 활용한 의료 인공지능이 전문의의 판독에 비해 19%나 정확도가 높다는 연구 결과가 나왔다.

출처: https://www.medicaltimes.com/

머스크 · 앤드류 양 등 공개서한…"AI 훈련 6개월 멈춰라"

A 김윤경 기자 | 승인 2023.03.30 08:59

"사회에 위험 초래"…GPT-4 개발 6개월 중단 요구

인공지능(AI)의 위험성을 경고하는 공개 서한이 공개됐다. 출처=게티이미지

일론 머스크 테슬라 최고경영자(CEO)를 포함한 많은 기술업계 인사들이 오픈AI의 최신 대형 인공지능(AI) 언어 모델 'GPT-4'에 대한 AI 시스템 훈련 중단을 촉구했다.

29일(현지시간) CNBC에 따르면, 머스크와 스티브 워즈니악 애플 공동 창업자 등 수백명은 공개 서한을 통해 "첨단 AI가 사회에 위험을 초래한다"며 모든 AI 연구소에 인간 수준의 지능과 경쟁할 수 있는 시스템 개발(훈련)을 6개월간 중단할 것을 촉구했다.

출처: https://www.smarttoday.co.kr/

메타, 생각을 이미지로 구현하는 AI 공개

뇌파 학습 과정 없이 이미지 학습만으로 두뇌 분석 구현

컴퓨팅 | 입력 :2023/10/19 10:58

남혁우 기자 | 기자 페이지 구독 | 기자의 다른기사 보기

메타가 실시간으로 두뇌 활동을 인식한 뒤 이미지로 재구성하는 인공지능(AI) 시스템을 공개했다.

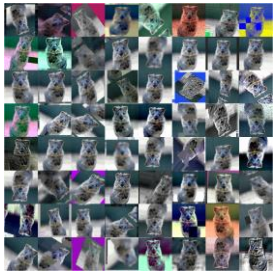
19일(현지시간) 실리콘앵글 등 외신에 따르면 메타는 비침습적 신경영상 기술인 뇌자기검사(MEG)를 사용해 뇌의 활동을 시각적으로 표현할 수 있는 AI시스템을 개발했다고 밝혔다.

뇌자기검사를 활용한 이 AI의 특징은 사람의 뇌파를 매번 학습시킬 필요가 없다는 점이다. 기존 이미지 데이터를 학습시키는 것 만으로 생각을 읽을 수 있고, 추후 정확도를 높일 수 있다고 메타가 강조했다.

출처: https://zdnet.co.kr/

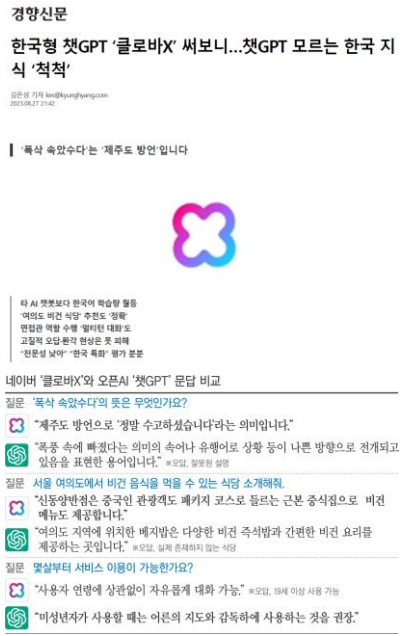
About artificial intelligence

- **인공지능 기술의 한계점**
  - **데이터가 없으면 AI도 없음**
    - 데이터가 부족하면 모델은 제대로 학습되지 않아 과소적합이 발생, 기본 패턴도 파악하지 못함
    - 잘못된 레이블이나 누락된 값이 포함된 불완전한 데이터는 모델의 예측 능력을 크게 떨어뜨릴 수 있음
  - **범용 솔루션 제작의 어려움**
    - 범용 인공지능 (Artificial General Intelligence, AGI)의 구현은 현대 과학과 기술의 가장 도전적인 문제
    - 아직까지는 특정 작업 내에서 일반화 능력을 극대화하는 것이 최선 (Artificial Narrow Intelligence, ANI)
  - **일반화 능력 고도화의 어려움**
    - 예) 이미지 학습 과정은 조명, 촬영 각도, 배경 등의 사소한 변화에도 매우 민감함
    - 실제 세계의 다양성을 모두 포괄하는 데이터 세트 구축은 현실적이지 않음



기술 경쟁력의 초점이 알고리즘에서 데이터로 옮겨가고 있음

출처: DALL·E 3



경향신문

한국형 챗GPT '클로바X' 써보니...챗GPT 모르는 한국 지식 '척척'

'폭삭 속았수다'는 '제주도 방언'입니다

타 AI 챗봇보다 한국어 학습량 월등
'여의도 비건 식당' 추천도 '정확'
편견관 역할 수행 '멀티턴 대화'도
고질적 오답·환각 현상은 못 피해
'전문성 낮아' '한국 특화' 평가 분분

네이버 '클로바X'와 오픈AI '챗GPT' 문답 비교

질문 '폭삭 속았수다'의 뜻은 무엇인가요?

"제주도 방언으로 '정말 수고하셨습니다'라는 의미입니다."

"폭풍 속에 빠졌다는 의미의 속어나 유행어로 상황 등이 나쁜 방향으로 전개되고 있음을 표현한 용어입니다." ※오답, 잘못된 설명

질문 서울 여의도에서 비건 음식을 먹을 수 있는 식당 소개해줘.

"신동양반점은 중국인 관광객도 패키지 코스로 들르는 근본 중식집으로 비건 메뉴도 제공합니다."

"여의도 지역에 위치한 베지밥은 다양한 비건 즉석밥과 간편한 비건 요리를 제공하는 곳입니다." ※오답, 실제 존재하지 않는 식당

질문 몇살부터 서비스 이용이 가능한가요?

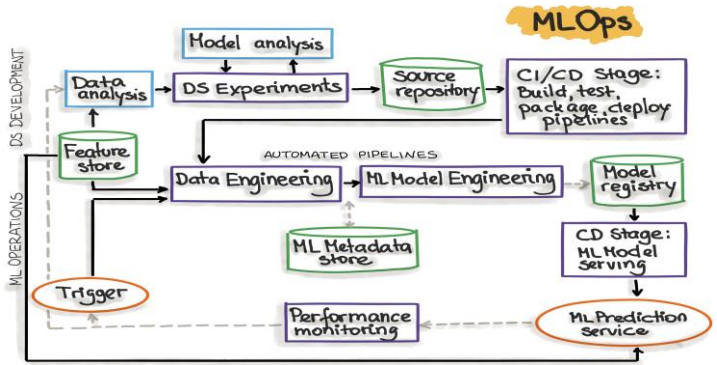
"사용자 연령에 상관없이 자유롭게 대화 가능." ※오답, 19세 이상 사용 가능

"미성년자가 사용할 때는 어른의 지도와 감독하에 사용하는 것을 권장."

출처: https://www.khan.co.kr/

About artificial intelligence

- **한계점 극복을 위한 AI 산업계의 동향**
  - **맞춤형 서비스의 수요**
    - 고객은 자신의 데이터에 기반한 맞춤형 서비스를 원함
      - 하지만 데이터의 민감성, 보안 및 개인정보보호 이슈로 인해 자신의 데이터를 직접 공유하려 하지 않음
    - 대부분의 AI 기업들은 고객의 데이터에 직접적인 접근을 하지 못하는 상황에서 작업해야 함
      - 개인화 서비스의 질을 높이기 어려움
  - **MLOps(ML + DevOps) 기반 플랫폼 서비스의 출현**



출처: https://ml-ops.org/

[이코노미스트] 입력 2023-03-30 18:20 수정 2023-03-30 20:35

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

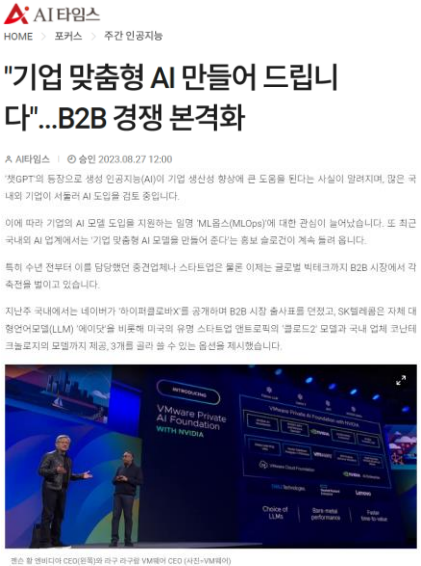
반도체 사업장 챗GPT 허용 20일, 정보 유출 사고 3건 발생
설비 계측수율 데이터, 미국 기업에 고스란히 전송...회수 불가
필요시 해당 임직원 징계...사내 전용 AI 서비스 구축 검토

삼성전자 평택캠퍼스 내 반도체 시설 모습. [사진 삼성전자]

[이코노미스트 정두용 기자] 우려가 현실이 됐다.

삼성전자가 디바이스솔루션(DS·반도체) 부문 사업장 내 챗GPT(ChatGPT) 사용을 허가하자마자 기업 정보가 유출되는 사고가 났다. 반도체 '설비 계측'과 '수율 불량' 등과 관련한 프로그램 내용이 고스란히 미국 기업의 학습 데이터로 입력됐다. 삼성전자는 이 같은 사고를 원천적으로 막고자 DS 내 조직인 혁신센터의 주관 아래 사내 전용 자체 인공지능(AI) 서비스 구축을 검토 중이다.

출처: https://economist.co.kr/



AI타임스

HOME > 포커스 > 주간 인공지능

"기업 맞춤형 AI 만들어 드립니다"...B2B 경쟁 본격화

AI타임스 | 승인 2023.08.27 12:00

'챗GPT'의 등장으로 생성 인공지능(AI)이 기업 생산성 향상에 큰 도움을 된다는 사실이 알려지며, 많은 국내외 기업이 서둘러 AI 도입을 검토 중입니다.

이에 따라 기업의 AI 모델 도입을 지원하는 일명 'ML옵스(MLOps)'에 대한 관심이 높아났습니다. 또 최근 국내외 AI 업계에서는 '기업 맞춤형 AI 모델을 만들어 준다'는 홍보 슬로건이 계속 들려 옵니다.

특히 수년 전부터 이름 담당했던 중견업체나 스타트업은 물론 이제는 글로벌 빅테크까지 B2B 시장에서 각축전을 벌이고 있습니다.

지난주 국내에서는 네이버가 '하이퍼클로바X'를 공개하며 B2B 시장 출사표를 던졌고, SK텔레콤은 자체 대형언어모델(LLM) '에이닷'을 비롯해 미국의 유명 스타트업 앤트로픽의 '클로드2' 모델과 국내 업체 코난테크놀로지의 모델까지 제공, 3개를 골라 쓸 수 있는 옵션을 제시했습니다.

젠슨 황 엔비디아 CEO(왼쪽)와 라구 라구람 VMWare CEO (사진=VMWare)

오픈AI와 엔비디아까지도 이 분야에 초점을 맞추고 있습니다. 오픈AI는 'GPT-3.5-터보'를 기업이 고쳐 쓸 수 있도록 미세조정 기능을 추가했고, 스케일 AI라는 ML옵스 전문 업체까지 파트너로 붙였습니다. 엔비디아는 VM웨어와 손잡고 아예 '기업용 생성 AI 플랫폼'을 구축한다고 발표했습니다.

사실 ML옵스는 알고리즘부터 데이터 라벨링, 최적화, 경량화, 서비스 배포, 재학습 등 소프트웨어와 하드웨어 여러 분야를 포함하고 있습니다. 이제까지는 각 분야에 맞춰 전문 기업이 시장을 구성하고 있었는데, 이 분야 덩치가 점점 커지는 분위기입니다.

또 빅테크의 경우 자사 LLM 채택을 늘리기 위해 관련 기술까지 제공한다는 의도가 있습니다. 궁극적으로는 이를 통해 클라우드 사업을 확장하겠다는 목표입니다. 본격적으로 AI로 돈을 벌어보겠다는 겁니다.

출처: https://www.aitimes.com/

AI in Cyber Security

- **보안 솔루션 내 인공지능 도입은 필수불가결**
  - **AI 기술의 수혜자는 방어자 뿐 아니라, 공격자도 포함됨**
    - AI는 목표 달성을 위해 인간의 이해 범주를 벗어난 방식으로 전개되는 기술
    - 공격에 악용된다면, 점차 인간이 패턴화 하기 어려운 사례가 증가할 것
  - **해킹의 자동화**
    - 생성형 AI 활용, 시스템의 취약점을 찾을 때까지 무수한 공격 시나리오를 자동 생성하고 시뮬레이션 수행
  - **맞춤형 피싱 공격 및 자동화**
    - 인공지능은 대량의 데이터를 분석하여 개인 또는 조직의 행동 패턴을 파악할 수 있음
    - 이를 활용하여 매우 정교한 피싱 이메일을 개발하고, 대상에게 전송하여 더 높은 확률로 정보 탈취 가능
    - 딥보이스 기반 보이스 피싱 자동화 등, 사람이 감지하기 어려운 공격 시도

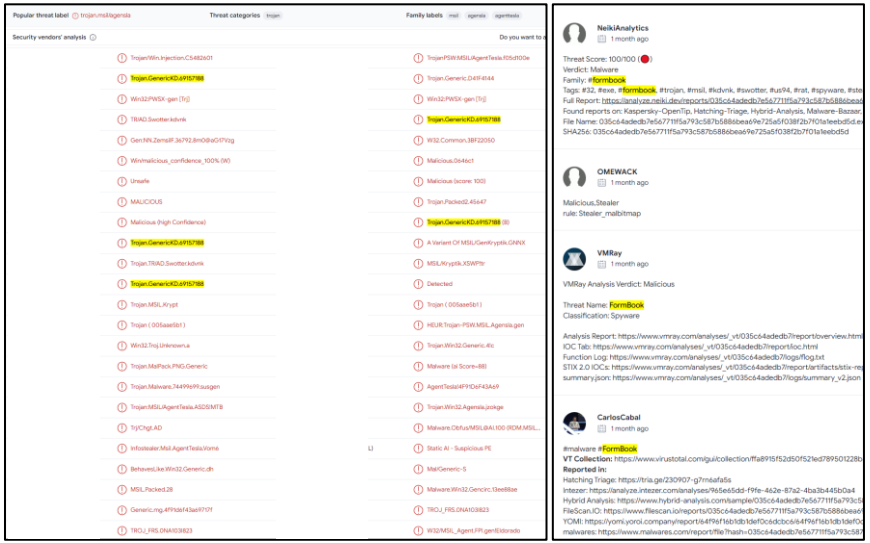
출처: DALL·E 3

Problem Definition

- **인공지능 솔루션 개발 방향?**

글로벌 업계에서 성능 경쟁이 가능한 고성능 모델 확보

- **바이러스 토탈로 보는 [주관적인] 시사점**
  - **중복된 탐지명 (가치 활용의 어려움)**
    - 일부 벤더 사의 경우, 상호간에 완전히 일치하는 탐지명을 출력하여 결과 참조에 주의를 요함
      - VT는 다양한 의견을 취합하는 데에 의의가 있기 때문에 동일한 엔진 결과를 여러 번 참조하는 것은 기피해야 함
  - **제한된 정보 및 일관되지 않은 탐지명 (정적 분석의 어려움)**
    - 다양한 안티 멀웨어 제품군이 존재하지만, 대부분 기본적인 탐지 기능만 제공하거나 각 벤더 사 별 상이한 결과를 출력해 실무에서의 응용이 어려움
      - 상호 간의 라벨링의 기준(방향성)이 다름으로 인한 결과일 수 있으나, 동적 분석 결과는 대부분 일치하는 모습을 보임
    - 정적 분석에 기반한 악성 유형 탐지는 아직 제한적인 모습을 보임
      - 실시간 처리, 분석 우회 기법, 리소스 증가 등의 제약 사항이 존재하는 동적 분석 또한 적절한 대체 수단이 될 수 없음



출처: https://www.virustotal.com/

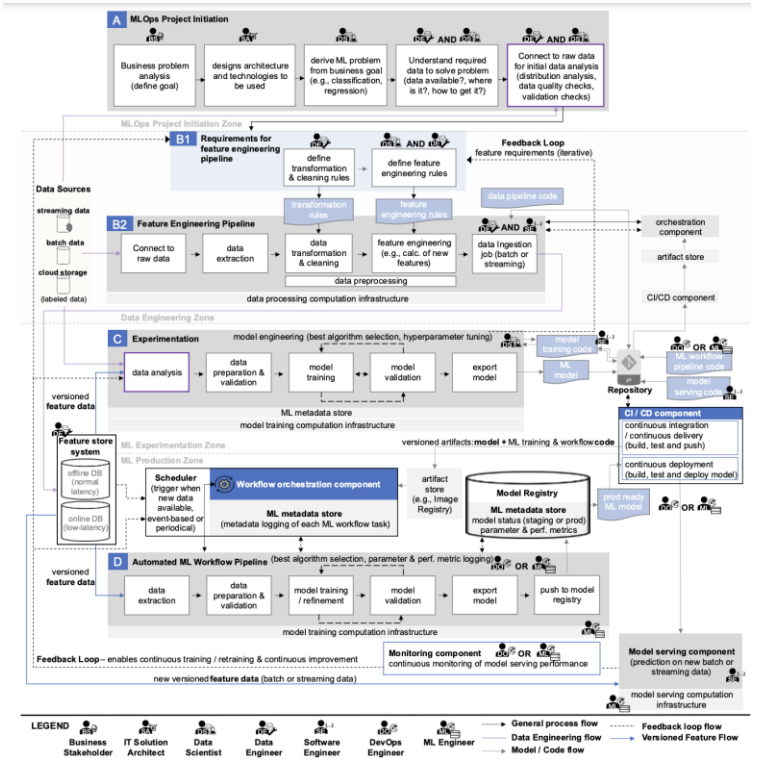
Problem Definition

- **타사의 검증된 엔진 도입의 한계**
  - **아이덴티티의 손실**
    - 동일한 엔진을 사용하는 다른 기업들과의 서비스 차별화를 달성하기 어려움
  - **주체적 서비스 개선 및 장애 대응**
    - 사용자의 피드백이나 시장 변화에 따른 서비스 개선이 제한적임
    - 특정 고객의 요구사항이나 환경에 맞추는 서비스의 커스터마이징이 제한됨
  - **데이터 자산 보호 및 활용**
    - 외부 엔진과의 데이터 교환 및 통신은 데이터 보안 유출의 위험을 증가시킬 수 있음
    - 아무리 양질의 데이터 자산을 보유하고 있더라도 이를 활용하는 데 제약이 생김

출처: https://medium.com/

Problem Definition

- **독자적 엔진 확보 및 운영의 어려움**
  - **광범위한 연구 개발 및 운영 분야 이해**
    - 보안과 AI는 각각 복잡하고 광범위한 분야이기 때문에, 두 분야의 지식을 통합하고 동시에 활용할 수 있는 전문성이 필요함
  - **매우 높은 성능 요구 사항**
    - 보안 분야에서는 작은 오차도 큰 위험을 초래할 수 있음
    - 정상/악성 분류는 거의 100%에 가까운 정확도를 보여야 하며, 이러한 성능 고도화에 실패한 경우 상용화가 어려움
  - **막대한 개발 비용**
    - 고급 인력에 기반한 연구 개발, 하드웨어 및 인프라, 개발 지연에 따른 기회 비용 등이 모두 손실로 돌아올 수 있으며 그로 인한 재개발 비용은 상당히 높게 측정됨

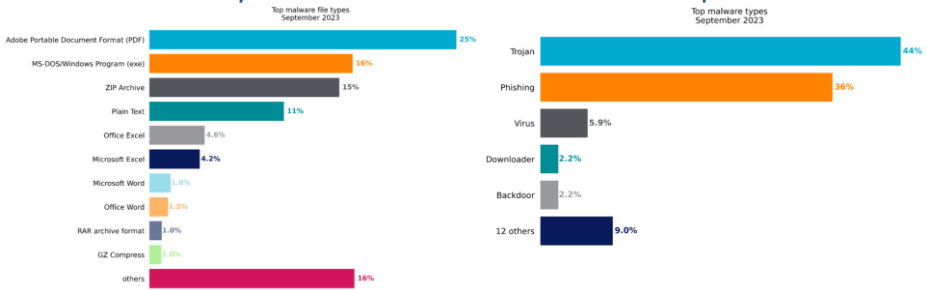


출처: https://hystax.com/

Our Solution

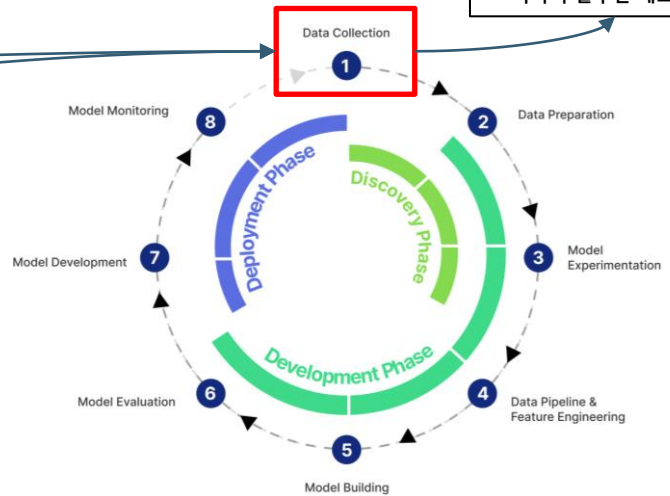
- **Customizable Anti-Malware 엔진 생성 Platform**
  - **Format-agnostic**: 악성코드가 될 수 있는 모든 파일 확장자 처리 능력
  - **Data-driven**: 데이터의 특징이나 변화에 따라 서비스의 동작을 최적화
  - **Fully-Automated Engine Deployment**: 사용자의 개입 없이, 엔진을 자동으로 배포

데이터 기반
요구사항 반영



출처: https://www.netskope.com/blog/netskope-threat-labs-stats-for-september-2023

데이터에 의해 이후 모든
모델링 프로세스가 결정,
독자적 솔루션 배포 가능



출처: https://markovate.com/blog/mlops-pipeline-integration/

Key Challenges for Customizable Anti-Malware Solution

- 안티 멀웨어 연구 동향

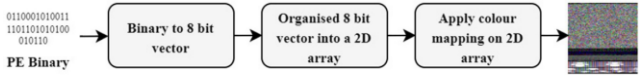


Fig. 1. An illustration of malware transformation into an image.

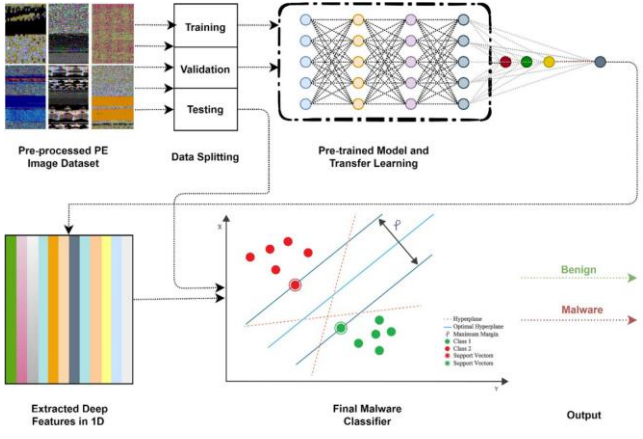
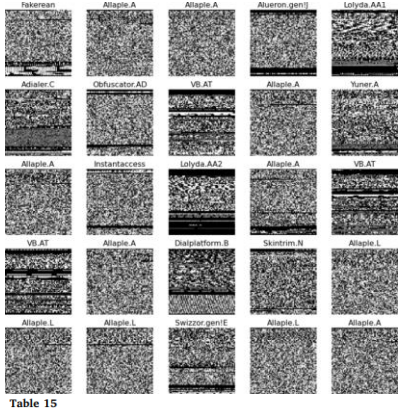


Fig. 2. The proposed malware detection framework.

출처: Shaukat, K., Luo, S., & Varadharajan, V. (2023). A novel deep learning-based approach for malware detection. Engineering Applications of Artificial Intelligence, 122, 106030.



Main Contributions

- Image-based PE Dataset Creation (RGB based)
- **Novel Hybrid Framework**
- Evaluation of Multiple Models
- **Outperforms other state-of-the-art techniques**

Table 15
A comparison of the proposed framework with the state-of-the-art.

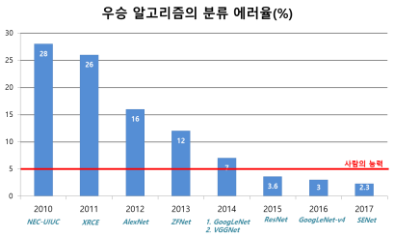
Sr. No.	year	Dataset	Reference#	Models	Accuracy	Precision	Recall	F-score
1	2011	Malimg	Nataraj et al. (2011)	K-NN	98.08%	–	–	–
2	2017	Malimg	Luo and Lo (2017)	CNN+LBP	93.72%	94.13%	92.54%	93.33%
3	2017	Malimg	Rezende et al. (2017)	ResNet50	97.48%	–	–	–
4	2017	Malimg	Rezende et al. (2017)	ResNet-50	98.62%	–	–	–
5	2017	Malimg	Makandar and Patrot (2017)	GIST+SVM	98.88%	–	–	–
6	2018	Malimg	Lo et al. (2019)	Xception	98.52%	–	–	–
7	2019	Malimg	Bhodia et al. (2019)	ResNet34	94.80%	–	–	–
8	2019	Malimg	Roseline et al. (2019)	Ensembling using RF	97.82%	98%	98%	98%
9	2019	Malimg	Agarap (2017)	CNN-SVM	77.22%	84%	77%	79%
10	2019	Malimg	Ben Abdel Ouahab et al. (2019)	K-NN	97%	–	–	–
11	2019	Malimg	Gibert et al. (2019)	CNN	97.18%	–	–	–
12	2019	Malimg	Vinayakumar et al. (2019)	CNN+LSTM	96.3%	96.3%	96.2%	96.2%
13	2019	Malimg	Vinayakumar et al. (2019)	RF	78.6%	–	–	–
14	2019	Malimg	Vinayakumar et al. (2019)	NB	80.5%	–	–	–
15	2019	Malimg	Vinayakumar et al. (2019)	KNN	41.8%	–	–	–
16	2019	Malimg	Vinayakumar et al. (2019)	DT	79.5%	–	–	–
17	2019	Malimg	Singh et al. (2019)	ResNet50	96.08%	95.76%	96.16%	95.96%
18	2020	Malimg	Vasan et al. (2020b)	VGG16,	97.59%	–	–	–
19	2020	Malimg	Vasan et al. (2020b)	ResNet50	95.94%	–	–	–
20	2021	Malimg	Hemalatha et al. (2021)	DenseNet	98.23%	97.78%	97.92%	97.85%
21	**2022**	**Malimg**	**Proposed framework**		**99.06%**	**98.47%**	**98.52%**	**98.49%**

Key Challenges for Customizable Anti-Malware Solution

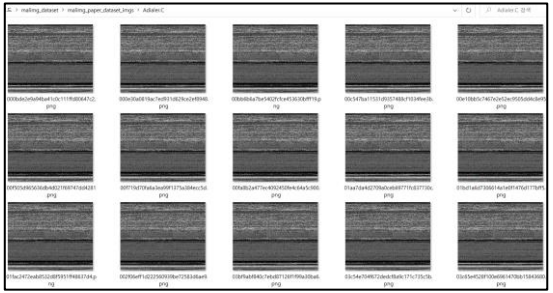
- 딥러닝 활용이 최선(?)

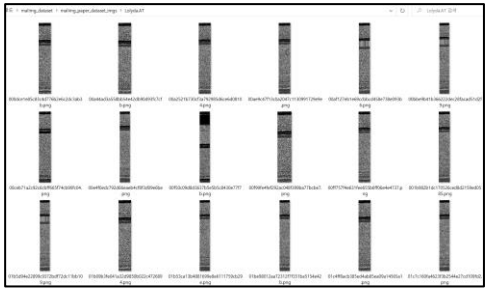
CNN 기반 이미지 분류 성능

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC, 2010-2017)
- 1,000개 class 대상
- Train : 약 1,200,000개
- Validation : 50,000개
- Test : 100,000개

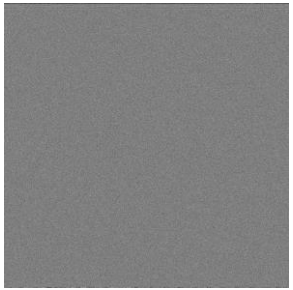


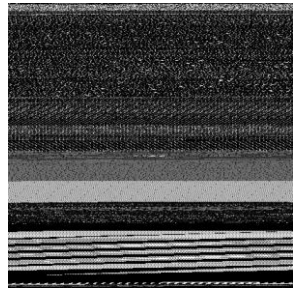
Experimental Dataset (Malimg)

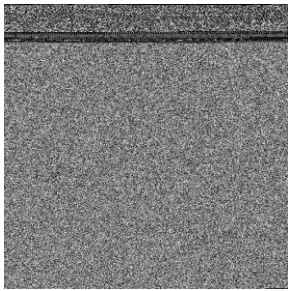


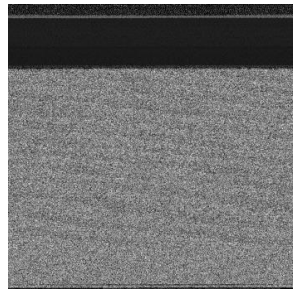


Real world (?)







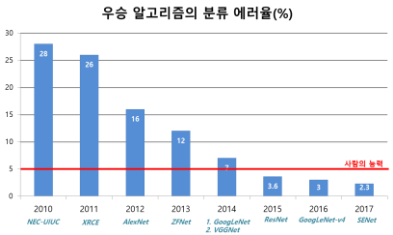


Key Challenges for Customizable Anti-Malware Solution

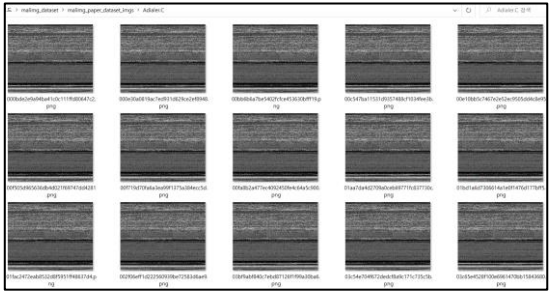
- 딥러닝 활용이 최선(?)

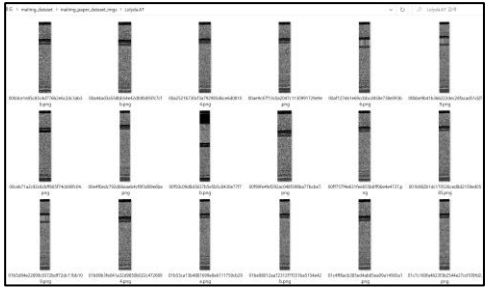
CNN 기반 이미지 분류 성능

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC, 2010-2017)
- 1,000개 class 대상
- Train : 약 1,200,000개
- Validation : 50,000개
- Test : 100,000개

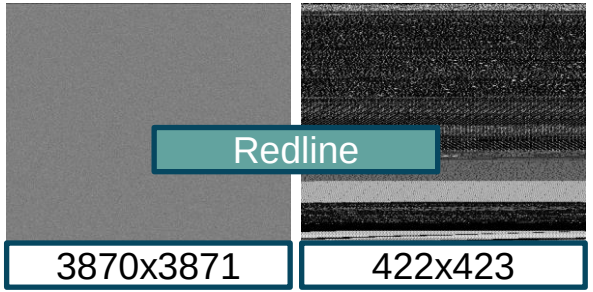


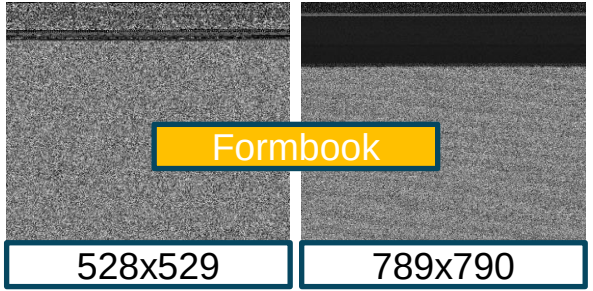
Experimental Dataset (Malimg)





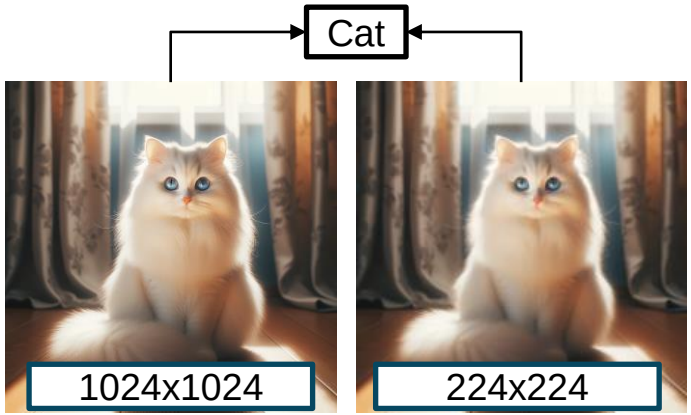
Real world (?)

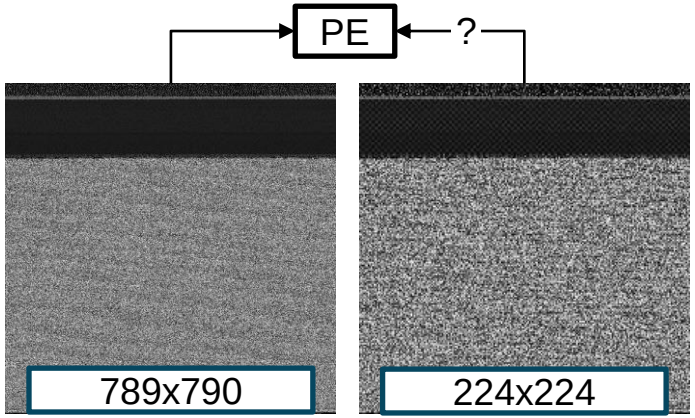




Key Challenges for Customizable Anti-Malware Solution

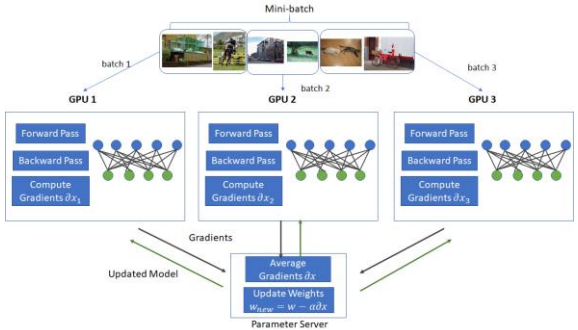
- **딥러닝 기반 접근 방식의 한계**
  - **딥러닝의 특성 및 장점**
    - 원본 데이터를 그대로 입력하여 사람의 선입견 없이, 주요 특성을 자동으로 선택하는 데 특화됨
    - 이를 통해 사람의 인식 방식과는 다른 방법으로 문제를 해결할 수 있으며, 때로는 사람의 인지능력을 초월함
  - **리소스 제한에 따른 데이터 손실**
    - 그러나 리소스 제한으로 실제로 원본 데이터를 그대로 입력하는 경우는 드뭄
    - 대신, **원본 데이터의 특성을 손상시키지 않는 선에서** 지정된 크기로 입력 크기를 조절



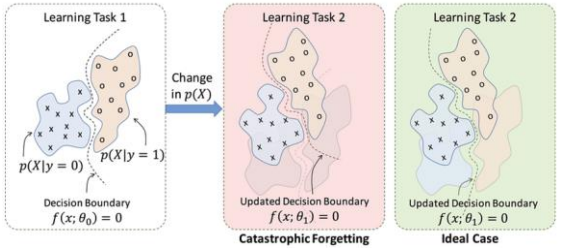


Key Challenges for Customizable Anti-Malware Solution

- **딥러닝 기반 접근 방식의 한계**
  - **Batch Processing 기반 최적화의 어려움**
    - 리소스의 제약 때문에 학습 과정에서 한 번에 전체 데이터 세트를 참조하지 못하고 일부의 정상/악성 사례만을 참조해야 함
    - 이런 방식은 전체 데이터 세트의 최적화를 위해 그 규모와 비례하여 상당한 시간이 소요
  - **모델 업데이트에서의 치명적 한계**
    - 사이버 보안 분야에서는 실시간으로 발생하는 새로운 위협에 대응하기 위한 **빠른 모델 업데이트가 그 어떠한 분야보다 중요**
    - 그러나 새로운 데이터에만 집중하여 학습을 수행하면, 과거에 학습했던 데이터에 대한 최적화 능력이 손실될 위험이 따름
    - 따라서 다시 전체 데이터 세트에 대한 Batch Processing 기반 학습을 진행해야 하며, 학습의 최적화 난도와 소요 시간도 함께 증가함



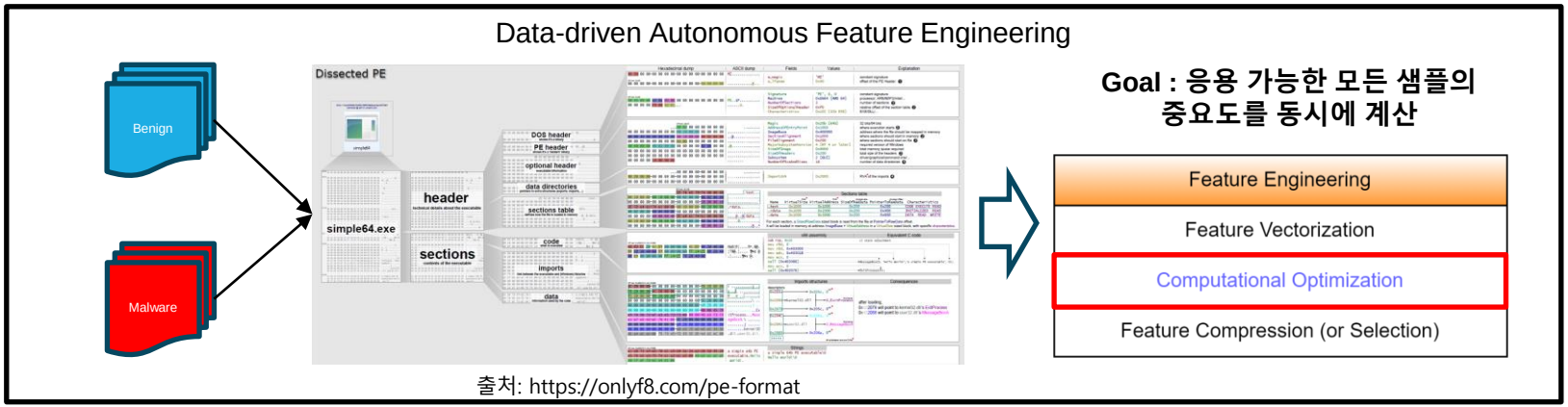
출처: https://www.telesens.co/2017/12/25/understanding-data-parallelism-in-machine-learning/



출처: Kolouri, S., Ketz, N., Zou, X., Krichmar, J., & Pilly, P. (2019). Attention-based selective plasticity. arXiv preprint arXiv:1903.06070.

Key Challenges for Customizable Anti-Malware Solution

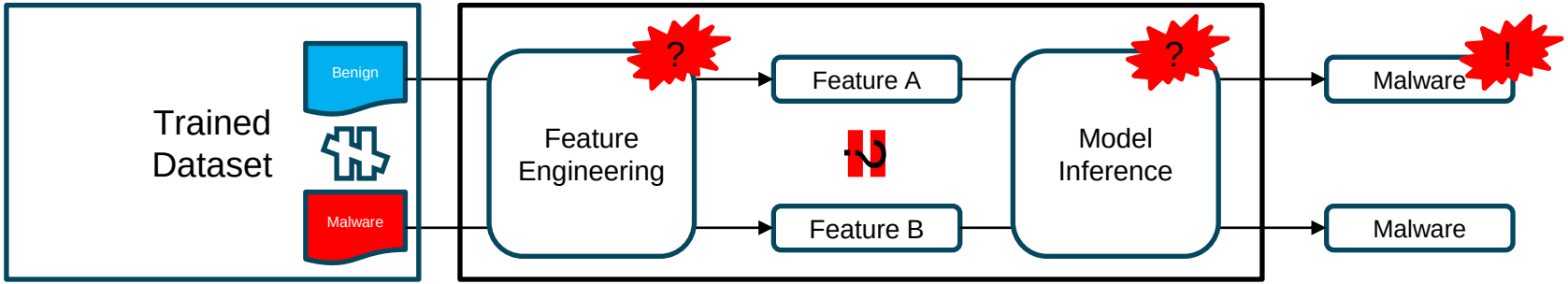
- **모델링 방안에 대한 고찰**
  - **우리는 이미 정답을 알고있다**
    - 딥 러닝을 사용하는 이유는 사람이 데이터의 **semantic을 모두 포괄하기 위한 이상적인 parser(feature extractor)**를 만들 수 없기 때문
      - 만약 object detection 문제에서 이미지 내 객체가 반드시 오른쪽에만 등장한다는 가정이라면, 왼쪽 절반을 자르고 분석하는 것이 합리적
    - 반면, 정상/악성 파일은 **정해진 포맷에 따라 semantic 정보를 담고 있음**
      - 어느 위치 및 주소에 어떤 의미를 가지는 값이 등장하는지 정확히 알 수 있음
      - 단, **분류 작업**에 구체적으로 어떤 값이 중요한지는 알 수 없으니, **학습을 통해 파악해야 함**



출처: https://onlyf8.com/pe-format

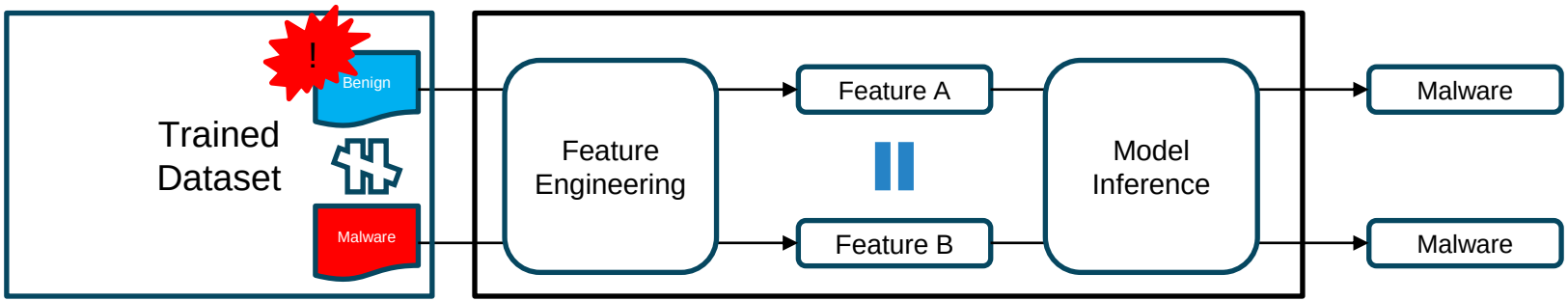
Key Challenges for Customizable Anti-Malware Solution

- **모델링 방안에 대한 고찰**
  - **모델링 방법에 정답은 없다**
    - 지속적으로 변화하는 위협에 대응하기 위해 최신 연구 동향을 주시해야 함
      - 단, 어떤 연구 결과나 기술도 실제 환경에서의 테스트와 검증 없이는 그 가치를 확신할 수 없음
      - 특히, **인공지능 기술은 실환경 실험에 기반한 의사 판단이 중요**
  - **모델링 방식이 합리적인지 판단하는 [주관적] 방법**
    - 사이버 보안 분야의 특성 상, **오류를 허용하는 validation & test는 지양해야 함**
      - 학습되지 않은 유형은 오류 발생이 당연함 (기존 데이터와의 연관성 부재)
        - **악성으로 학습하지 않은 유형을 악성이라고 하는 것은 해당 모델의 오탐 가능성이 높은 것을 의미**
      - 단, 확보한 데이터 세트 내에서는 **정탐률 100%를 보증해야 함 -> 오류를 전부 학습해서 정탐하도록 만들어야 함**
        - **만약, 불가능하다면 전통적인 시그니쳐 방식을 대체할 수 없음**



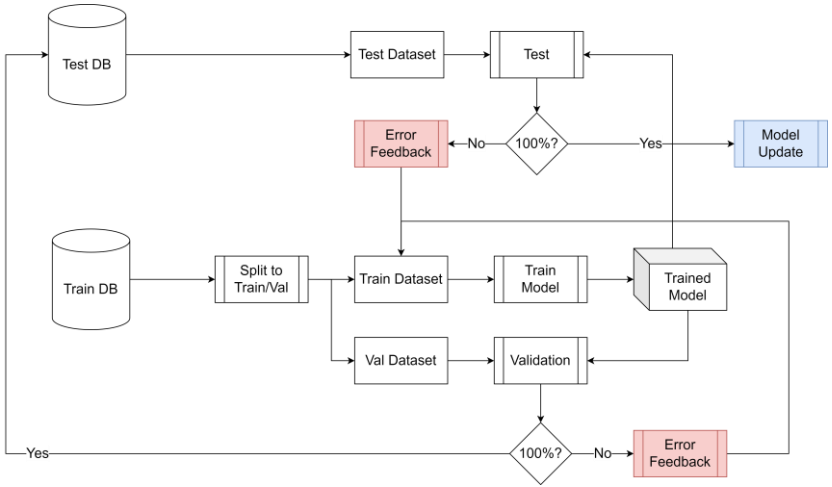
Key Challenges for Customizable Anti-Malware Solution

- **모델링 방안에 대한 고찰**
  - **모델링 방법에 정답은 없다**
    - 지속적으로 변화하는 위협에 대응하기 위해 최신 연구 동향을 주시해야 함
      - 단, 어떤 연구 결과나 기술도 실제 환경에서의 테스트와 검증 없이는 그 가치를 확신할 수 없음
      - 특히, **인공지능 기술은 실환경 실험에 기반한 의사 판단이 중요**
  - **모델링 방식이 합리적인지 판단하는 [주관적] 방법**
    - 사이버 보안 분야의 특성 상, **오류를 허용하는 validation & test는 지양해야 함**
      - 학습되지 않은 유형은 오류 발생이 당연함 (기존 데이터와의 연관성 부재)
        - **악성으로 학습하지 않은 유형을 악성이라고 하는 것은 해당 모델의 오탐 가능성이 높은 것을 의미**
      - 단, 확보한 데이터 세트 내에서는 **정탐률 100%를 보증해야 함 -> 오류를 전부 학습해서 정탐하도록 만들어야 함**
        - **만약, 불가능하다면 전통적인 시그니쳐 방식을 대체할 수 없음**



Key Challenges for Customizable Anti-Malware Solution

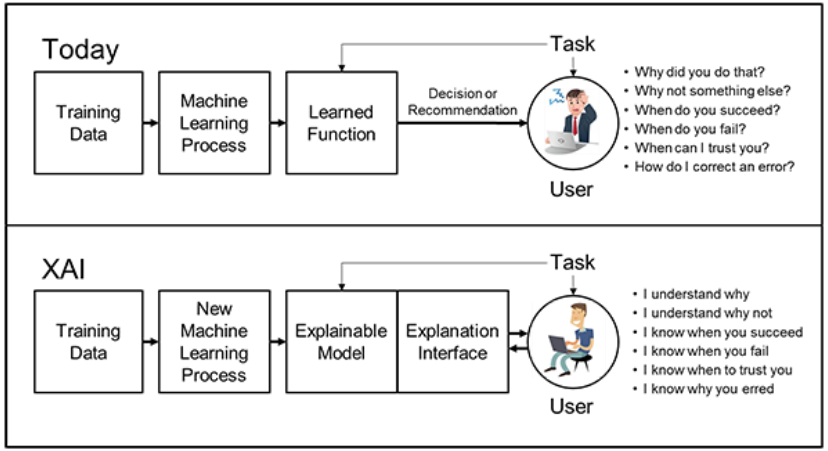
- **모델링 방안에 대한 고찰**
  - **단순 정탐률 100%는 성능을 보증하지 못함**
    - 일반화(Generalization) 검증이 필수
      - 얼마나 적은 학습 데이터로 전체 데이터 세트에의 최적화를 달성하는가?
    - PE 파일 포맷, 사내 데이터 기반 검증 결과
      - 패킹, 난독화, 암호화 등의 기법 적용 여부 구분없이 데이터를 수집함
      - **약 15,000개의 학습 데이터만으로 100만개 이상 규모의 데이터 세트 장악 (정상/악성 분류 작업 기준)**
  - **일반화 검증을 통한 데이터 최소화 실현**
    - 필요 시, **전체 데이터를 대표하는 학습 데이터**와 유사도(Feature Embedding Similarity)가 높은 데이터를 우선적으로 삭제
      - 리소스 최적화 및 빠른 학습 속도
      - 효율적인 데이터 관리 및 응용
      - 저장 공간 및 비용 절감



Enable Dataset Compression

Key Challenges for Customizable Anti-Malware Solution

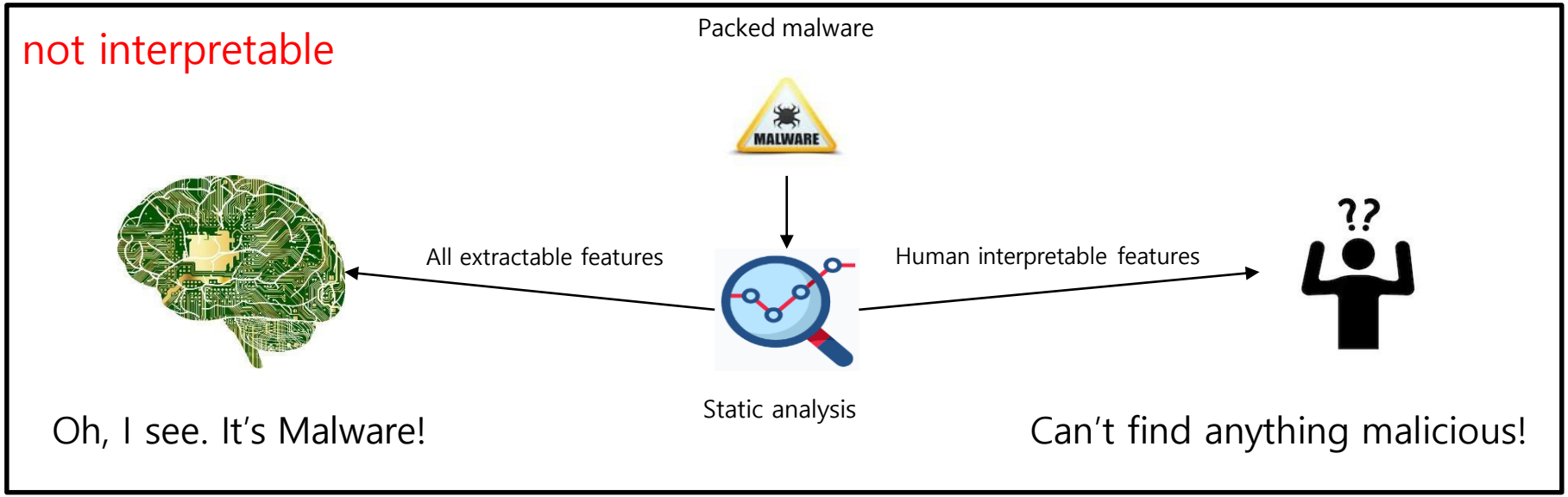
- **Explainable AI에 대한 고찰**
  - **설명 가능한 인공지능을 지향해야 하는가?**



출처: DARPA, Defense Advanced Research Projects Agency

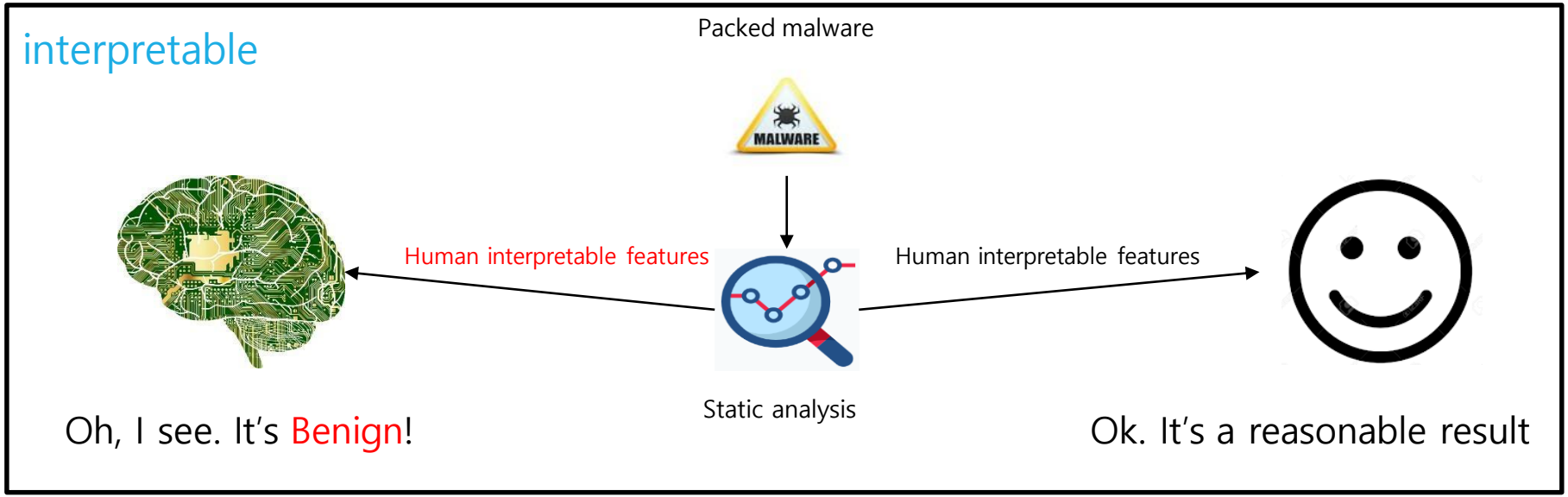
Key Challenges for Customizable Anti-Malware Solution

- **Explainable AI에 대한 고찰**
  - **Tradeoff between model complexity and model interpretability**



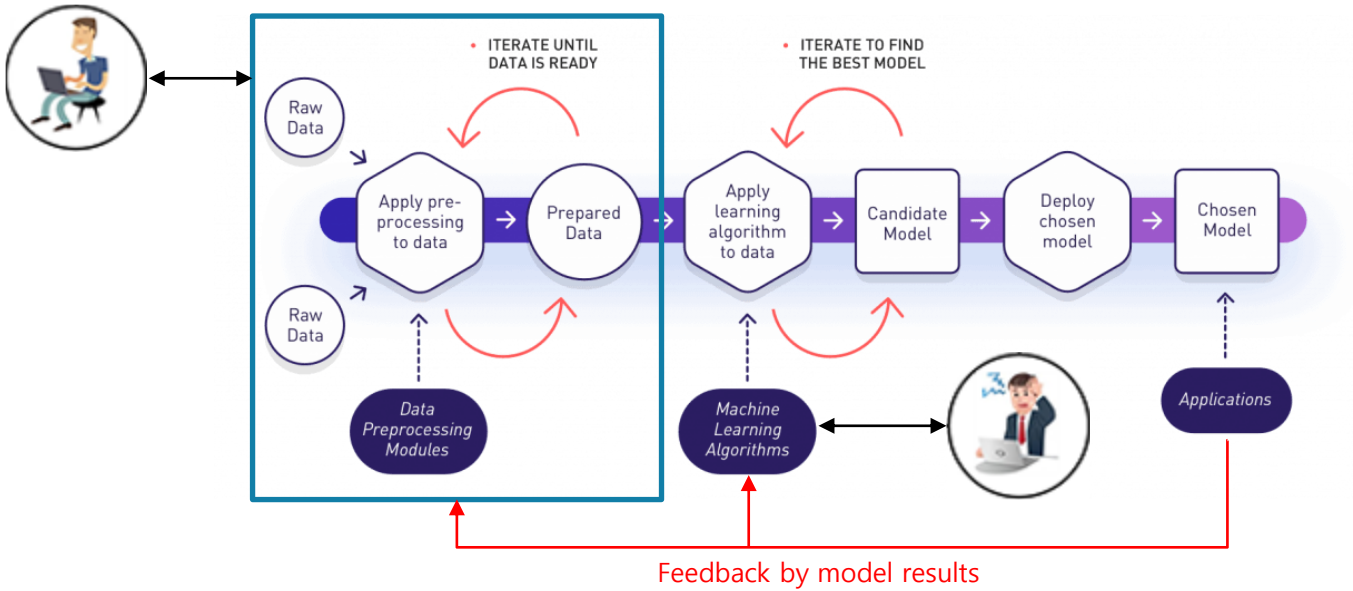
Key Challenges for Customizable Anti-Malware Solution

- **Explainable AI에 대한 고찰**
  - **Tradeoff between model complexity and model interpretability**



Key Challenges for Customizable Anti-Malware Solution

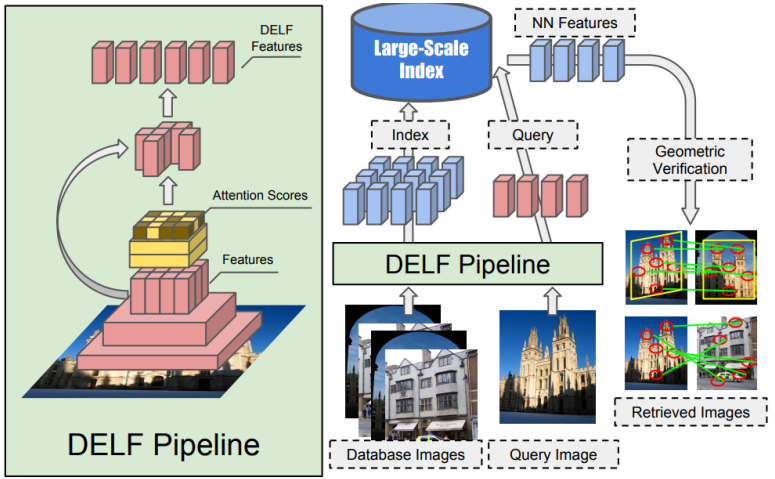
- **Explainable AI에 대한 고찰**
  - **Data preparation 과정을 통해 learning outcomes를 이해하고 제어할 수 있음**



출처: https://www.datasciencecentral.com/

Key Challenges for Customizable Anti-Malware Solution

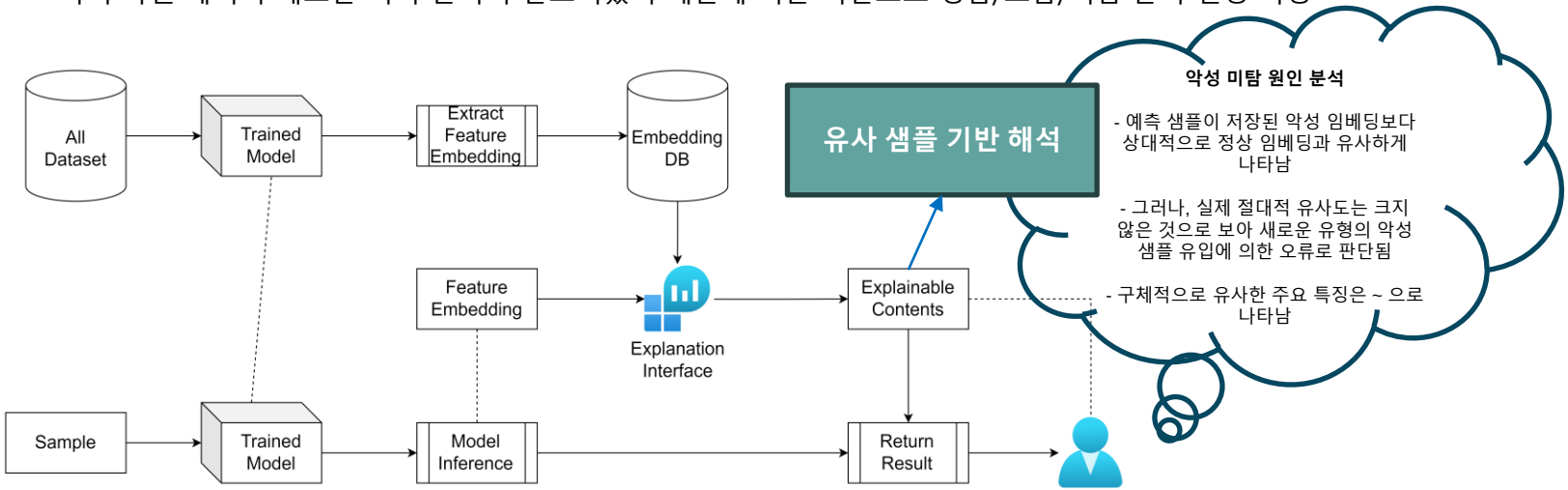
- **Explainable AI에 대한 고찰**
  - **Feature Embedding에 기반한 유사 사례 검색**
    - Embedding DB는 압축된 feature 정보만을 저장하기 때문에 저장 공간 요구량이 크게 감소함
      - **대량의 샘플을 대상으로 처리 및 분석을 실시간으로 수행할 수 있음**
    - 기 구축된 데이터 세트는 이미 분석이 완료되었기 때문에 이를 기반으로 정탐/오탐/미탐 근거 설명 가능



출처: Large-scale image retrieval with attentive deep local features

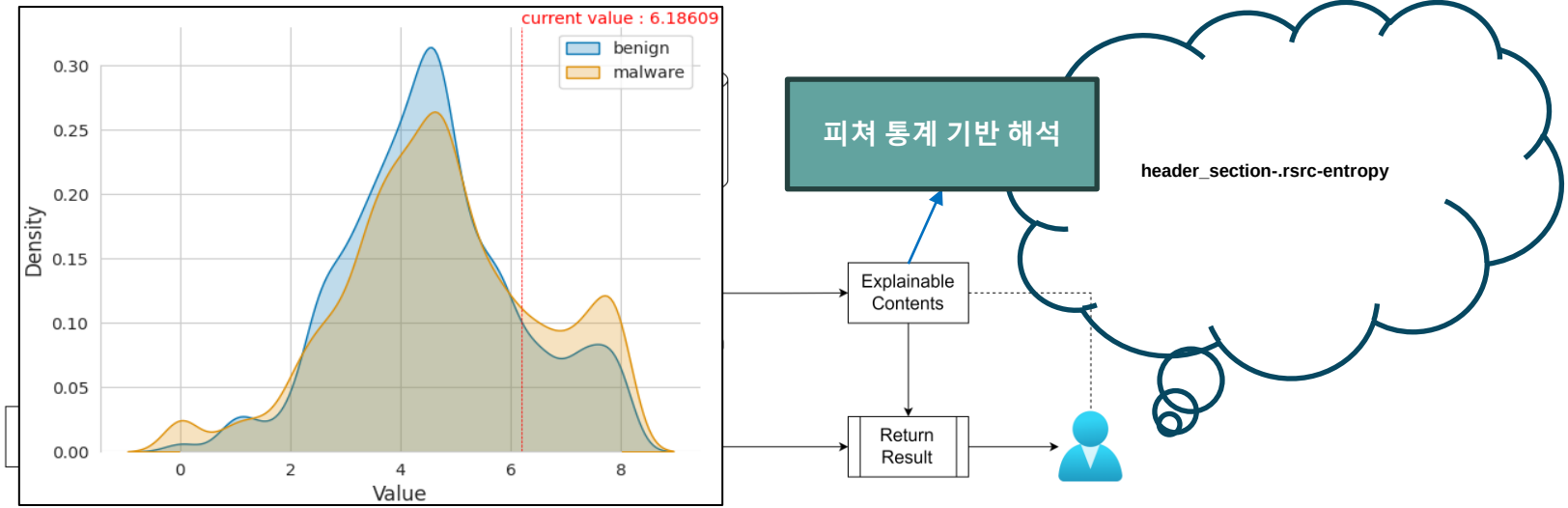
Key Challenges for Customizable Anti-Malware Solution

- **Explainable AI에 대한 고찰**
  - **Feature Embedding에 기반한 유사 사례 검색**
    - Embedding DB는 압축된 feature 정보만을 저장하기 때문에 저장 공간 요구량이 크게 감소함
      - **대량의 샘플을 대상으로 처리 및 분석을 실시간으로 수행할 수 있음**
    - 기 구축된 데이터 세트는 이미 분석이 완료되었기 때문에 이를 기반으로 정탐/오탐/미탐 근거 설명 가능



Key Challenges for Customizable Anti-Malware Solution

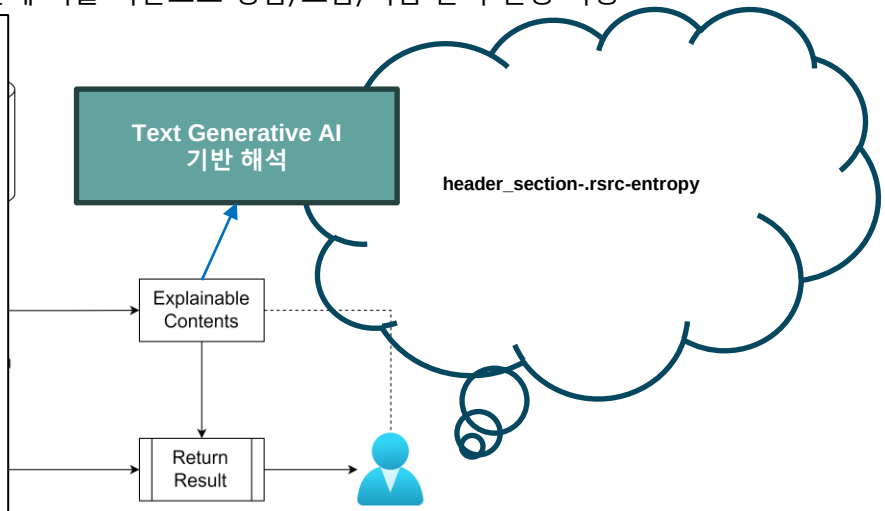
- **Explainable AI에 대한 고찰**
  - **Feature Embedding에 기반한 유사 사례 검색**
    - Embedding DB는 압축된 feature 정보만을 저장하기 때문에 저장 공간 요구량이 크게 감소함
      - **대량의 샘플을 대상으로 처리 및 분석을 실시간으로 수행할 수 있음**
    - 기 구축된 데이터 세트는 이미 분석이 완료되었기 때문에 이를 기반으로 정탐/오탐/미탐 근거 설명 가능



Key Challenges for Customizable Anti-Malware Solution

- **Explainable AI에 대한 고찰**
  - **Feature Embedding에 기반한 유사 사례 검색**
    - Embedding DB는 압축된 feature 정보만을 저장하기 때문에 저장 공간 요구량이 크게 감소함
      - **대량의 샘플을 대상으로 처리 및 분석을 실시간으로 수행할 수 있음**
    - 기 구축된 데이터 세트는 이미 분석이 완료되었기 때문에 이를 기반으로 정탐/오탐/미탐 근거 설명 가능

PE(포터블 실행파일) 파일 구조의 'header_section'는 다양한 정보를 포함하는데, 그 중 **'.rsrc' 필드는 리소스 섹션을 의미**합니다. 이곳에는 **실행파일이 사용하는 아이콘, 메뉴, 문자열 등과 같은 리소스 데이터**가 들어있습니다. 그리고 **'entropy'는 이러한 리소스 섹션이 얼마나 복잡한 패턴(랜덤)을 가지는지를 나타내는 통계적 측정항목으로, 정보 이론에서 파생된 개념입니다**. 값이 높을 수록 데이터가 더 복잡하거나 무작위적이며, 더 낮으면 더 단순하거나 예측 가능한 패턴을 가집니다. 일반적으로 하나의 참조 개체의 불확실성을 나타내며, 이 값이 크면 클수록 불확실성이 큰 것으로 해석할 수 있습니다. **악성코드는 행동을 숨기거나 탐지를 피하기 위해 자주 데이터를 암호화하거나 압축하는 등의 복잡한 패턴을 사용합니다. 그래서 .rsrc의 entropy이 높은 값(예를 들어, entropy > 7)을 가질 가능성이 높습니다. 반면에, 정상적인 실행 파일은 대게 낮은 엔트로피 값을 가질 가능성이 높습니다.** 따라서 .rsrc 섹션의 entropy 값은 파일이 악성인지 아닌지를 구별하는데 유용한 지표가 될 수 있고, 특히 머신러닝 알고리즘 등을 사용하여 악성코드를 탐지하는데 매우 중요한 속성이 될 수 있습니다.



Code Semantics

AI-Powered Customizable Anti-Malware Solution

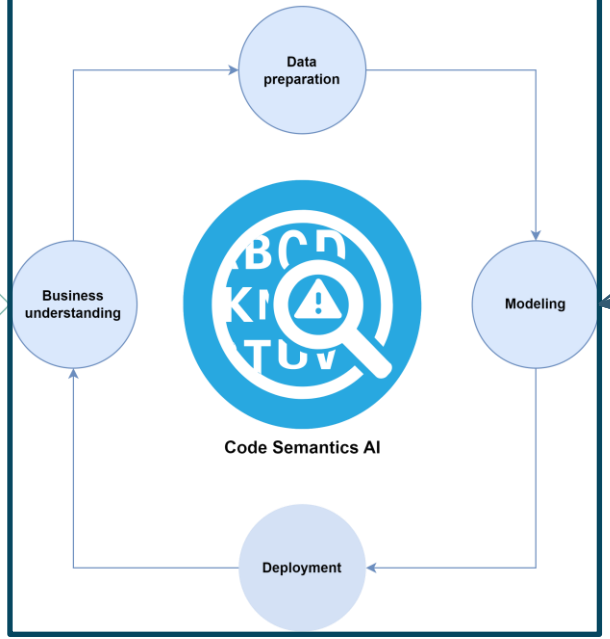
고객 요구사항 반영 및 기술지원

CUSTOMER INTERACTIONS

PRIVATE DATA

CURRENT BUSINESS PROTOCOLS

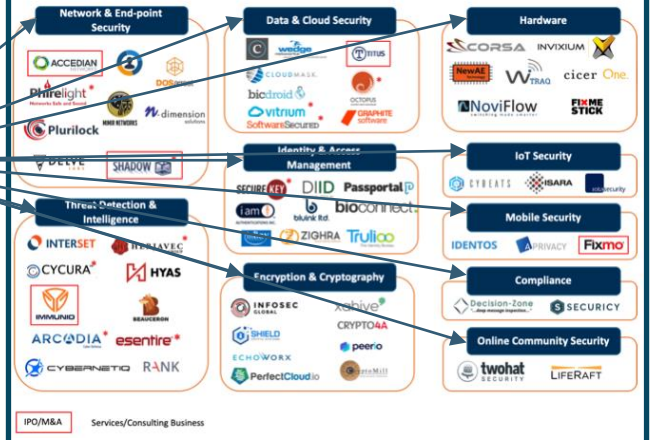
데이터 기반 모델 생성 및 배포 자동화



자체 엔진 생성 기반 비즈니스 차별화/다각화 지원

업계 최고 수준의 성능 경쟁력 확보 지원 (글로벌 백신 테스트 인증 등)

데이터 최소화, XAI 등 솔루션 운영 단계에서의 차별화 기능 지원



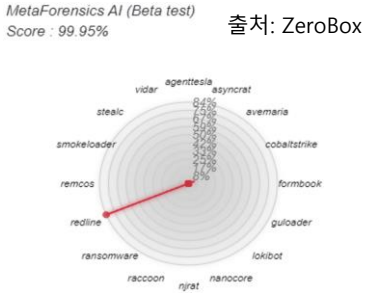
출처: https://medium.com/

Code Semantics

AI-Powered Customizable Anti-Malware Solution

- **실 적용 사례**
  - 고객의 데이터(요구사항) 기반 모델링 자동화 및 테스트

PE 기반 악성코드 유형 분류 커스터마이징



출처: ZeroBox

악성코드 유형	Num of Correct (Acc)		Num of Test	Num of Train
	Multi-Class	Multi-Label		
guloader	**12 (1.0)**	**12 (1.0)**	**12**	**44**
asyncrat	**12 (1.0)**	**12 (1.0)**	**12**	**51**
redline	24 (0.923)	23 (0.8846)	26	100
nanocore	7 (0.7)	7 (0.7)	10	40
ransomware	12 (0.75)	11 (0.6875)	16	64
avemaria	12 (0.8)	10 (0.6667)	15	57
smokeloader	16 (0.8889)	12 (0.6667)	18	72
lokibot	14 (0.7778)	12 (0.6667)	18	68
njrat	9 (0.75)	8 (0.6667)	12	48
vidar	13 (0.7647)	11 (0.6471)	17	67
formbook	21 (0.6774)	15 (0.4839)	31	123
remcos	12 (0.7059)	8 (0.4706)	17	68
cobaltstrike	5 (0.5)	4 (0.4)	10	40
agenttesla	15 (0.6521)	6 (0.2609)	23	92
raccoon	6 (0.375)	5 (0.3125)	16	64
stealc	4 (0.5)	1 (0.125)	8	28
총계	**194 (0.7433)**	**157 (0.6015)**	261	1,023

HTML 기반 유해 도메인 탐지 기능 커스터마이징

도메인 유형	Num of Correct (Acc)	Num of Detection	Num of Train
Unknown	-	49,990	-
정상	-	78,887	3,046
저작권 침해	7 (1.0)	7	26
변조	10 (0.7692)	13	183
도박	189 (1.0)	189	70
포르노	209 (1.0)	209	419
불법 업소 홍보	9 (0.8182)	11	82
총계	424 (0.9883)	129,306	3,826

THANKS

Meta Forensics
Revolutionizing Digital Investigation

Do you have any questions?

contact@metaforensics.ai