

0. 목차

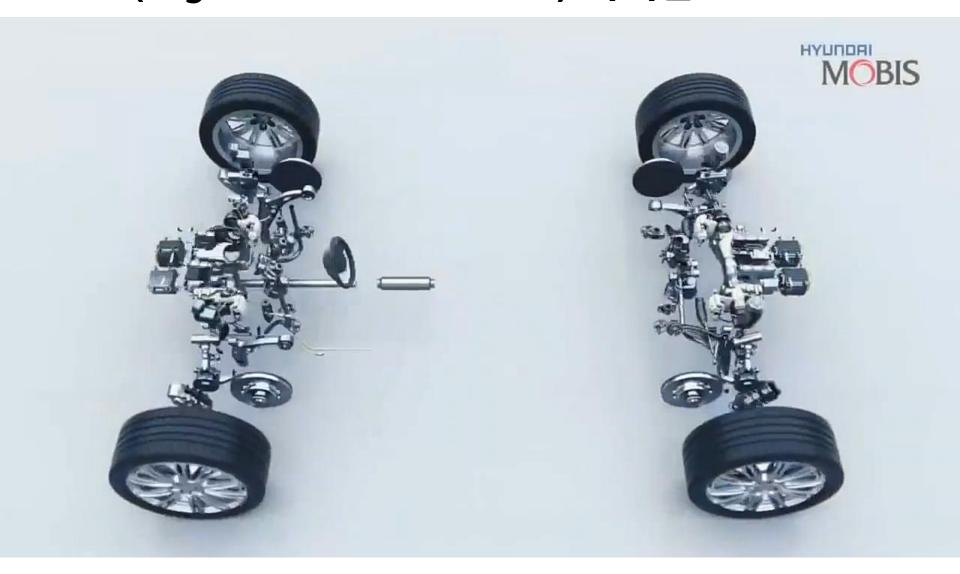


- 1. DT 시대란?
- 2. ABCD 기술 변화
- 3. 빅데이터 처리 프로세스
- 4. 디지털 전환



1. DT(Digital Transformation) 시대란? © magicecole





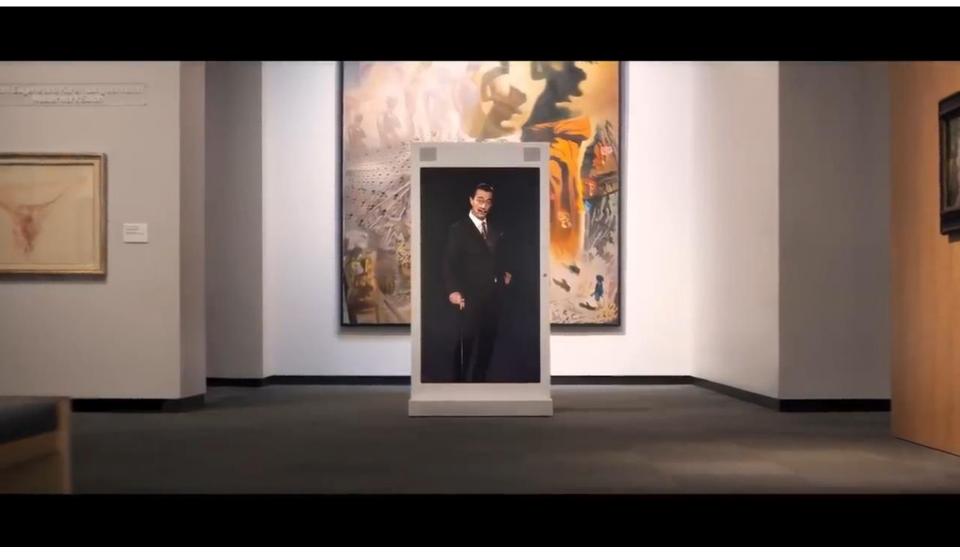
1. DT(Digital Transformation) 시대란? © magicecole



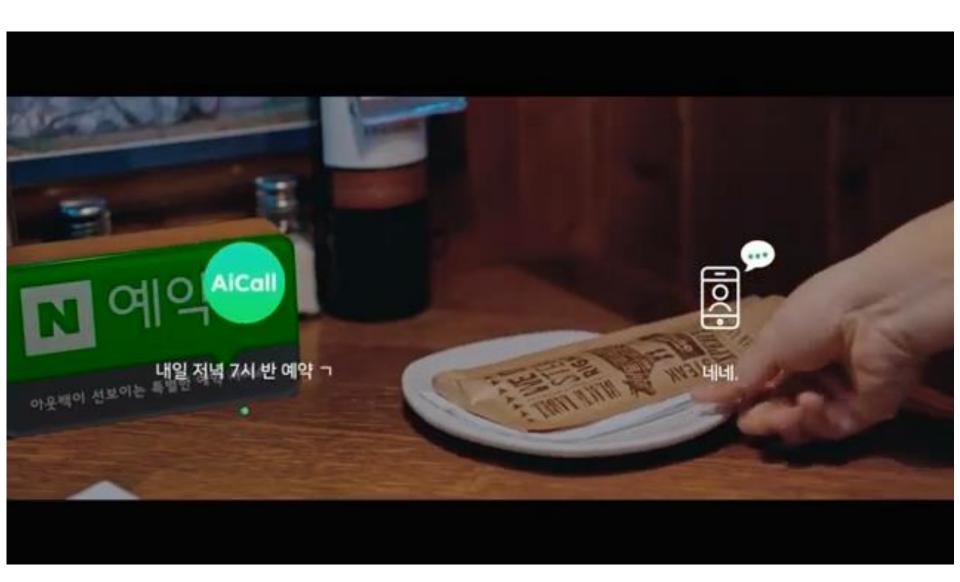


1. DT 시대란?



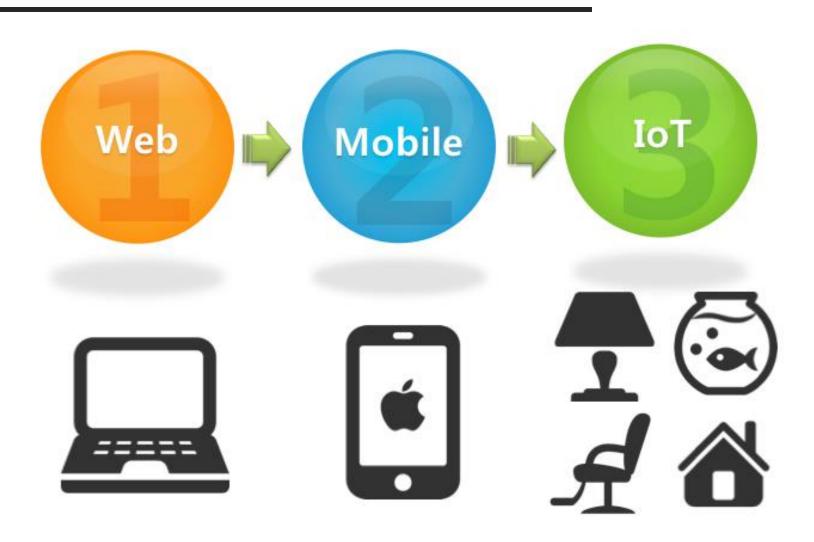








ICT 기술로 인한 일상 생활 속 근본적인 변화



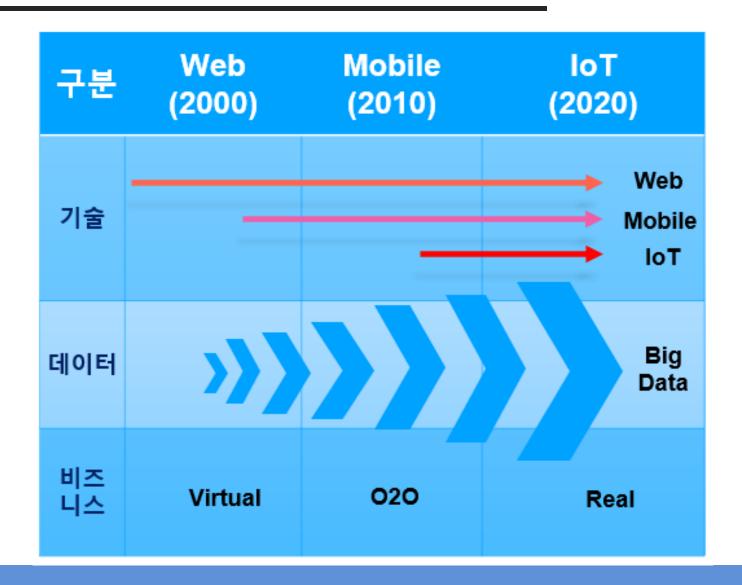


ICT 기술로 인한 일상 생활 속 근본적인 변화

구분	Web (2000)	Mobile (2010)	loT (2020)
기술	HTML, JS, CSS ASP, JSP, PHP Client/Server	iOS Android Cloud	Sensor Actuactor Edage
데이터	결제 데이터 회원 정보 검색 정보	SNS, 검색 비정형 데이터 	세상의 모든 데이터 Digital Twin
비즈 니스	쇼핑몰, 전자상거래, 	O2O 핀테크 	???



기술 발전과 폭발적인 데이터 생성, 그리고 확대되는 비즈니스





인터넷 기술의 등장과 성장 (1997 ~ 현재)



www.empas.com











고객층?

서비스?









소셜커머스 등장과 성장 (2009 ~ 현재)





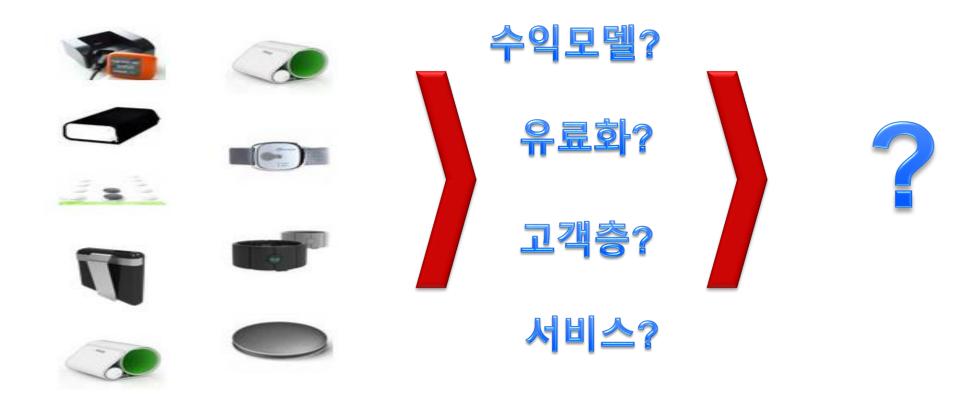








사물 인터넷 제품과 성장 (2013 ~ 현재)





기술과 서비스가 결합된 신 산업 기회





웹 – 모바일 – 사물인터넷 시대의 의미





데이터를 모으고 분석하고, 판단해 활동을 취하는 기기 및 시스템



AI 인프라

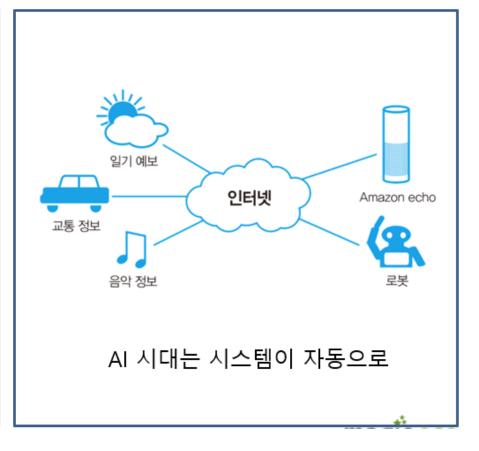


인터넷 시대와 AI 시대 서비스의 차이점

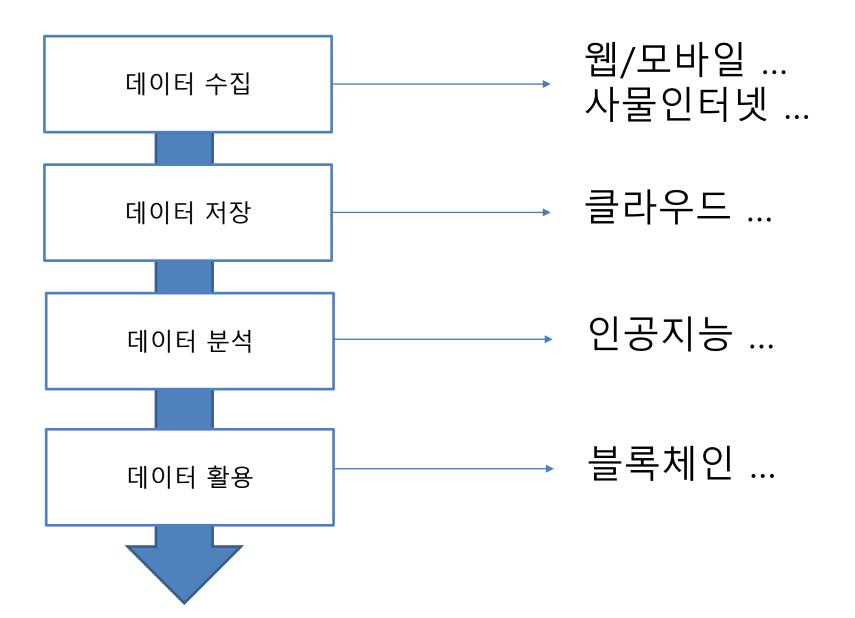
기존 인터넷

Google 일기 예보 검색 창 인터넷 교통 정보 음악 정보 검색 창 인터넷 시대는 내가 직접 개입

인공지능 시대









알파고(AlphaGo) 시리즈 성능 비교

이름	공개시점	전적	학습법	하드웨어
알파고 판	2015년 10월	판후이 2단에게 5-0 승리	딥러닝, 강화학습	GPU 176개, TPU 4개
알파고 리	2016년 3월	이세돌 9단에게 4-1 승리	딥러닝, 강화학습	GPU 176개, TPU 4개
알파고 마스터	2017년 5월	커제 9단에게 3-0 승리	딥러닝, 강화학습	TPU 47H
알파고 제로	2017년 10월	알파고 리에 100-0, 알파고 마스터에 89-11 승리	강화학습	TPU 4개

〈자료〉 동아일보, "기보없이 독학… 최강 알파고버전 등장", 2017. 10. 19.





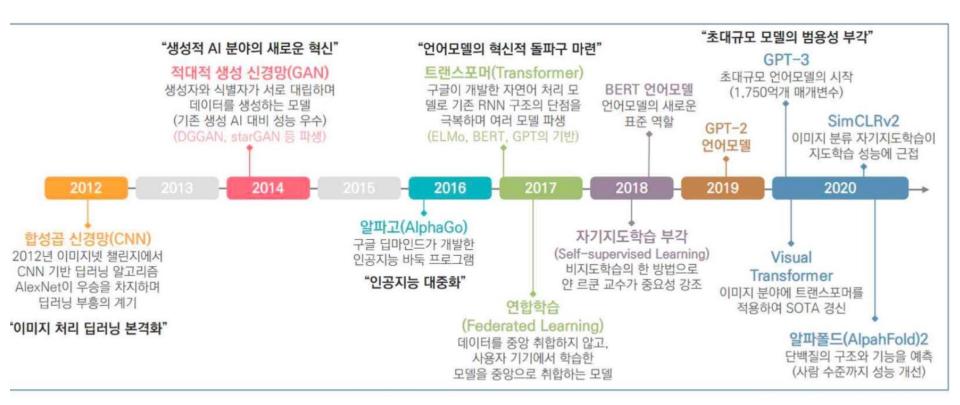




\$ 24.99



인공지능 기술의 주요 티핑 포인트



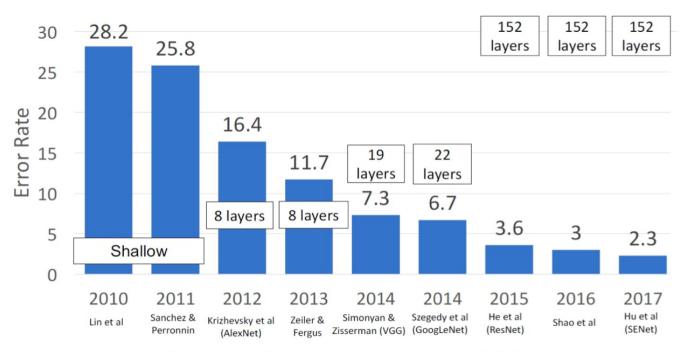
* 출처: NIA, 9대 핵심 기술 분석 및 주요 시사점, 2021.1.15



딥러닝...

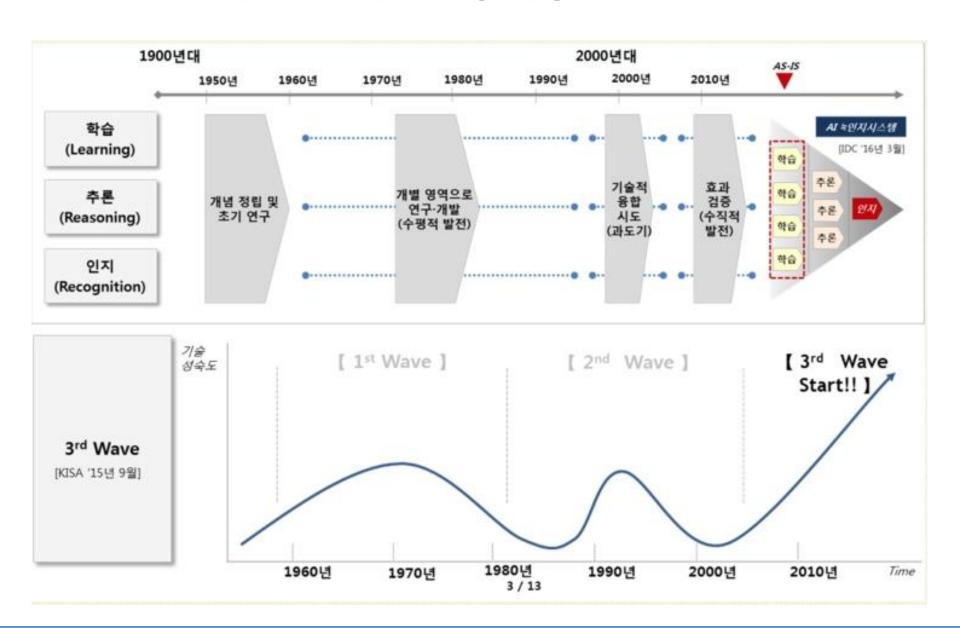
2012년 ILSVRC 이미지 인식대회 → 1,000여 카테고리 분류된 100만개 이미지 인식 2011년까지 인식률 75% 2012년 알렉스넷 인식률 84.7% 딥러닝 인식률 → 97% 육박 (인간 인식률 95%)

< 이미지넷 챌린지의 알고리즘 혁신>

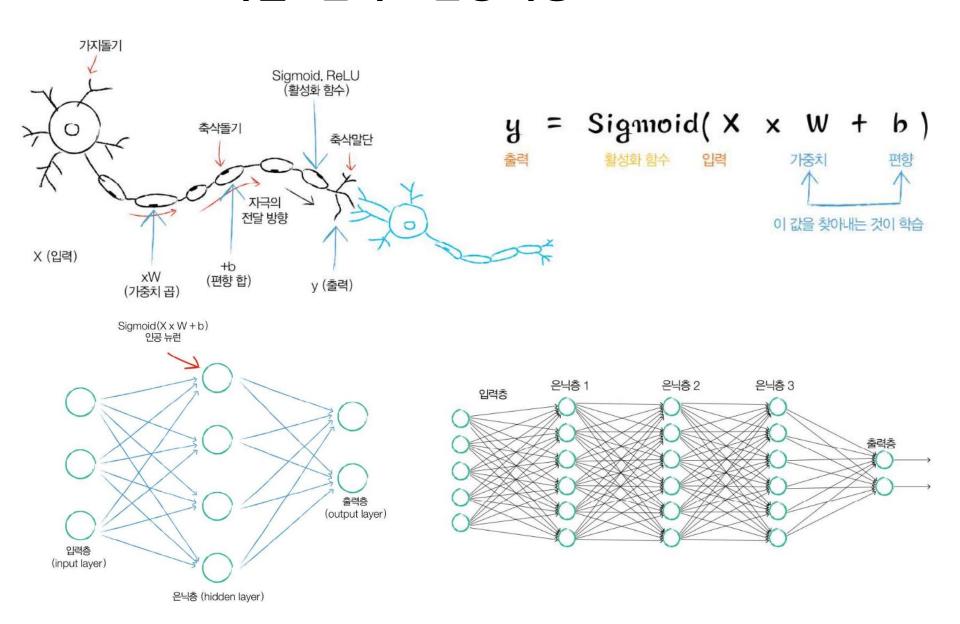


※ 자료: Convolution Neural Networks 강의자료 (Stanford Univ.)











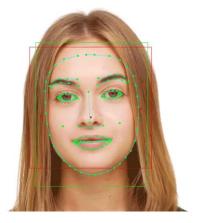
Tensorflow lite 기술







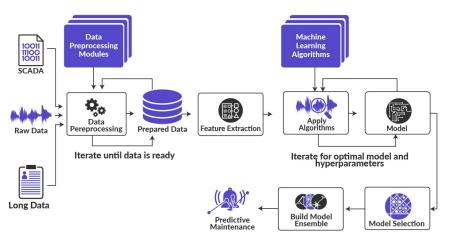
다양한 DeepLearning모델

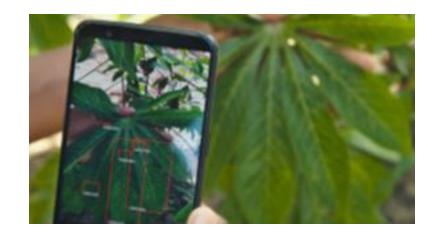




Al AutoML 기술

Presenso's AutoML and Deep Learning Engine



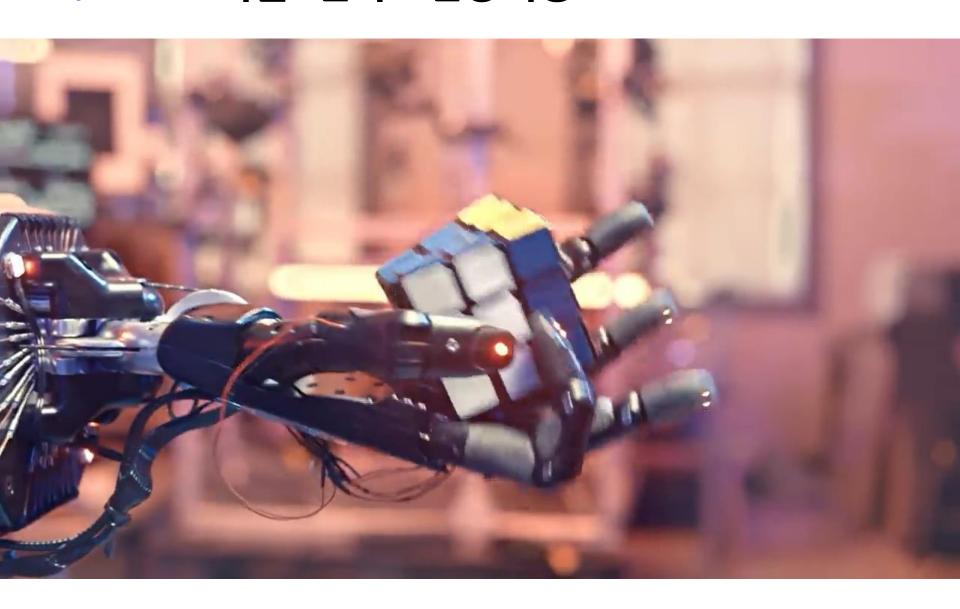














Open Al

https://openai.com/

GPT-3

(2020.5월) 1750억개 파라미터 4990억개 데이터셋중에서 가중치 샘플링 3000억(300B)

GPT-2

(2019.2월) 15억개 파라미터 Seq2Seq 알고리즘

GPT-1

(2018.6월)

구글

페이스북

ALBert

(2019.12월)

XLNet

(2019.6월)

Bert

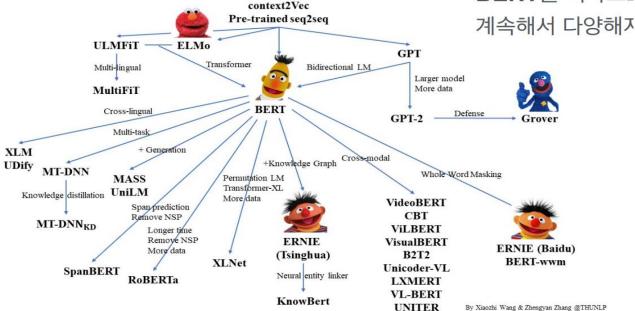
(2018.11월)

로버타(RoBERTa)

마이크로소프트

튜링NLG





Semi-supervised Sequence Learning



GPT3가 쓴 뉴스가 랭킹1위, 사람을 이겼다

② 입력 2020-08-19 11:32 │ ♀ 댓글 0

Y콤비네이터가 운영하는 美 해커뉴스에 인공지능 생성 기사 게재 버클리대 학생이 GPT3로 생성한 글... 알아차린 사람 거의 없어

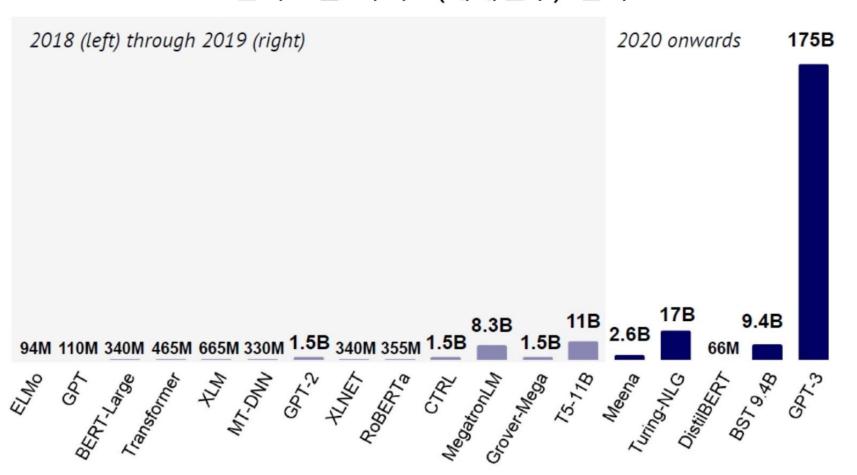
https://gptcrush.com/

https://github.com/elyase/awesome-gpt3

https://www.facebook.com/groups/TensorFlowKR/?post_id=1264627403878315



< 언어모델 사이즈(매개변수) 변화 >



※ 자료: State of AI(2020.10)



데이터 크기와 성능의 관계는?

인간 뇌 신경세포: 860억개 추정

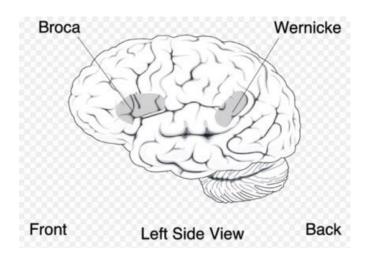
신경세포 연결 시냅스 100조개

GPT-3 1750억개 파라메터 → 자연어 처리

브로카(Broca): 언어 표현 영역

베르니케(Wernike): 언어 이해 영역

If. GPT-4가 GPT-3의 100배라면, 17조개 파라메터





DALL-E2 – OpenAI : 자연어를 이용한 그림 생성 AI





Github copilot : 자연어 기반 코딩 보조 AI

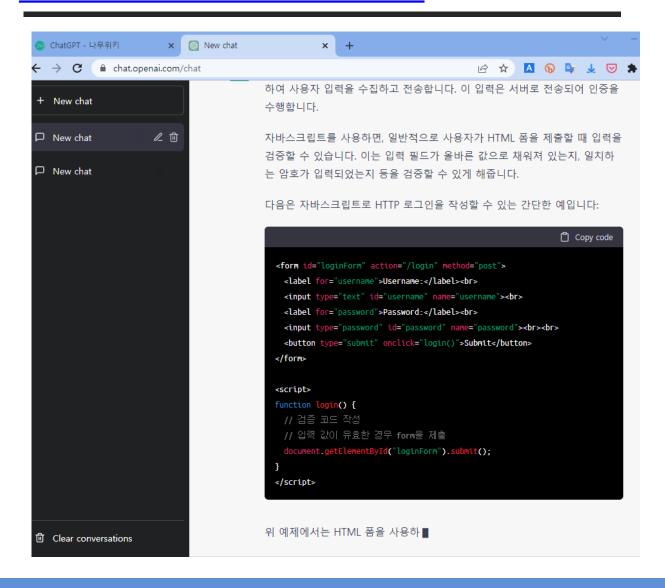
GitHub Copilot

```
s main.js 1
  import './style.css'

function intersectionOfArrays(a, b) {
  let result = []
  for (let i = 0; i < a.length; i++) {
    for (let j = 0; j < b.length; j++) {
        if (a[i] === b[j]) {
            result.push(a[i])
        }
    }
    return result</pre>
```



<u>Chat GPT : GPT 3.5 기반 대화형 인공지능</u>





인공지능 9대 핵심 기술

- 1 GPT-3
 - : GPT-3의 등장 및 언어모델의 스케일업
- 2 연합학습(Federated learning)
 - : 민감정보가 보호되는 분산 AI 학습
- 3 엣지 Al(Edge Al)
 - : 온디바이스 AI를 위한 모델 경량화
- 4 트랜스포머(Transformer)
 - : NLP의 트랜스포머가 컴퓨터 비전까지 확대
- 5 시스템2 AI(System2 AI)
 - : 단순 이해를 넘어 인과적 관계를 이해
- 6 자기지도학습(Self supervised learning)
 - : 데이터 라벨링의 한계를 극복
- 7 생성적 AI(Generative AI)
 - : 인식(판별)을 위한 AI를 넘어 창조하는 AI
- 8 전이학습(Transfer learning)
 - : 누구나 딥러닝을 할 수 있도록 지원
- 9 AutoML
 - : AI도 AI가 만들어 낸다

주요 시사점

1 초대규모 모델 부상

. 트랜스포머 : 병렬처리 기술

자기지도학습: 라벨링 불필요

GPT-3: 초대규모 언어 모델

$oldsymbol{2}$ 응용서비스 지원

. 연합학습: 민감데이터 분산학습

엣지AI: 클라우드 활용 불필요

. 전이학습: 딥러닝 활용 지원

3 우수인재 필수

. 인과 관계 알고리즘 설계 역량 . 생성적 AI 연구 역량 필요

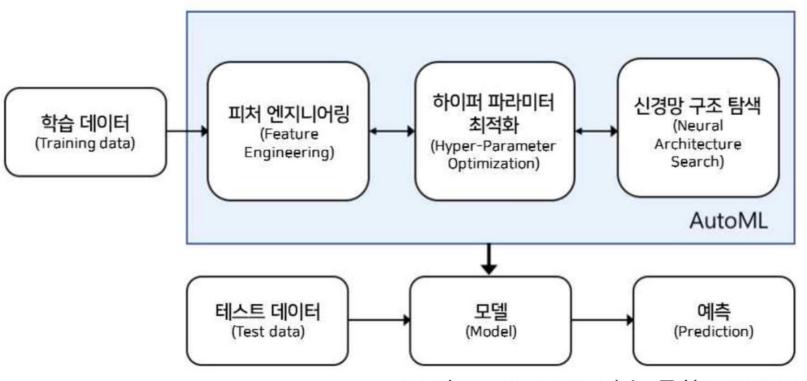
.....

알고리즘 자동 최적화 역량 필요



AI를 만드는 AI 기술의 발달...

<머신러닝 생애주기 중 AutoML 영역>



※ 자료: AutoML 기술 동향(ETRI,'19.8)



https://webrtclab.herokuapp.com/speech-recognition/

	peech Recognition API를 활용하면 더 재밌는 기능을 만들 수 있습니다. 이크에 이야기를 해보세요. (마이크와 가까울수록 인식률이 좋습니다)	
마이크 •	Text to speech	

예약어 (마이크를 켠 상태에서 아래 문자를 읽어보세요)

- 레드, 그린, 옐로우, 오렌지, 그레이, 골드, 블랙
- 알람
- 노래 켜, 노래 꺼, 음악 켜, 음악 꺼
- 볼륨업, 볼륨다운
- 스피치, 말해줘



https://teachablemachine.withgoogle.com

Teachable Machine

이미지, 사운드, 자세를 인식하도록 컴퓨 터를 학습시키세요.

사이트, 앱 등에 사용할 수 있는 머신러닝 모델을 쉽고 빠르게 만들어 보세요. 전문지식이나 코딩 능력이 필요하지 않습니다.

시작하기





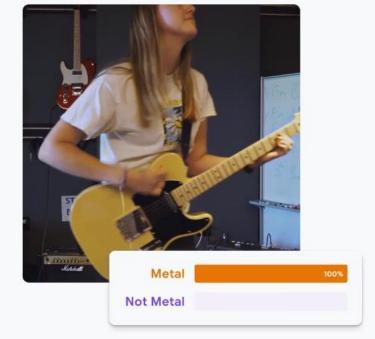




























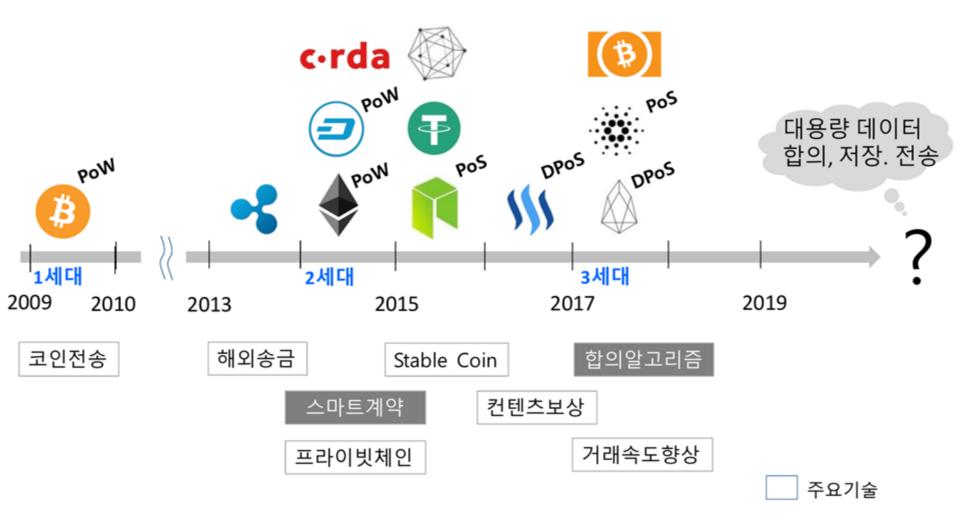


12년전 라스즐로 핸예츠는 1만 비트코인으로 피자 2판 구매

- 2010년 1만 비트코인 == 약 50,000원
- 2021년 1만 비트코인 == 약 670,000,000,000원
- 2022년 1만 비트코인 == 약 270,000,000,000원







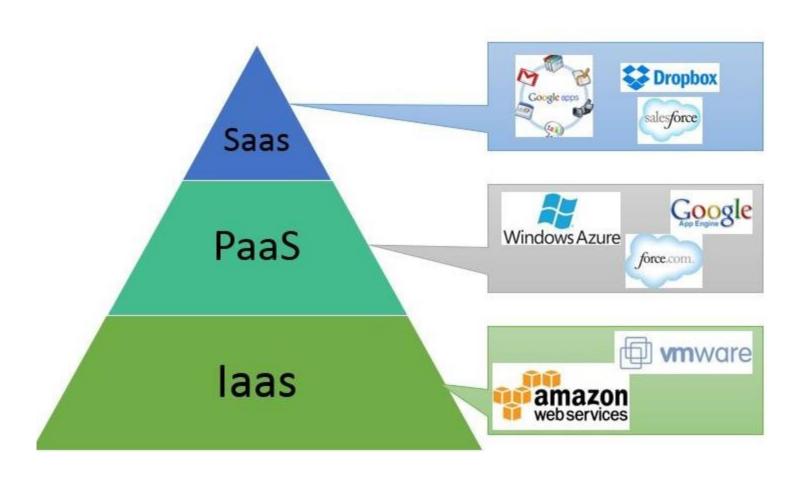














클라우드 절대 강자! → AWS

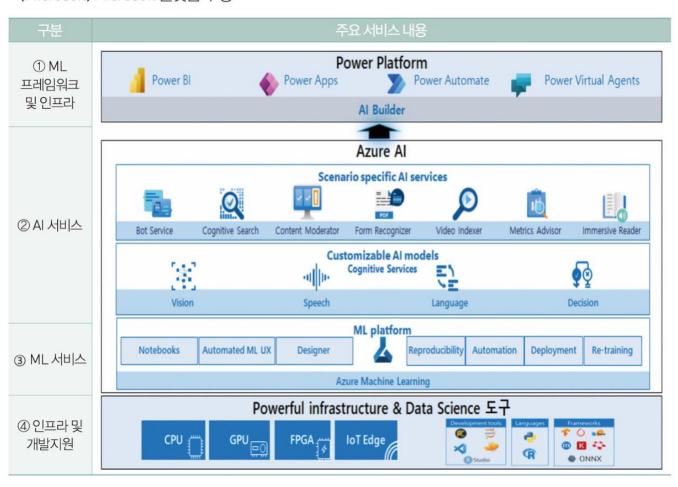
• (AWS) AWS 플랫폼 구성





개발생태계의 정점 → Microsoft

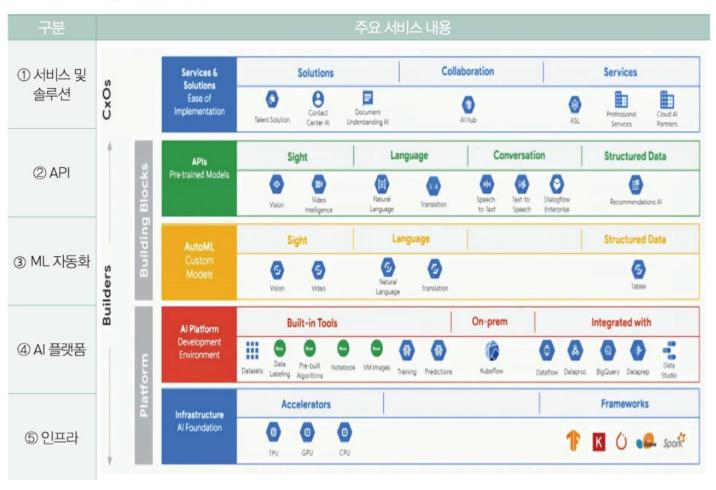
• (Microsoft) Microsoft 플랫폼 구성





검색을 넘어 새로운 세상으로 → Google

• (Google) Google 플랫폼 구성

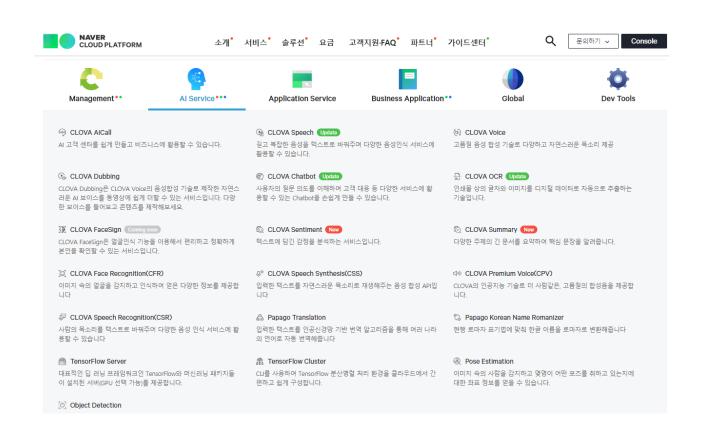




STT
Speech-to-Text

TTS
Text-to-Speech

챗봇





킬로바이트(KB)	1,000	10³	1,000
메카바이트(MB)	100만(萬)	10 ⁶	1,000,000
기가비이트(GB)	10억(億)	10°	1,000,000,000
테라바이트(TB)	1조(兆)	10 ¹²	1,000,000,000,000
페타바이트(PB)	1,000조	1015	1,000,000,000,000,000
엑사바이트(EB)	100경(京)	10 ¹⁸	1,000,000,000,000,000,000
제타바이트(ZB)	10部(垓)	1021	1,000,000,000,000,000,000
요타바이트(YB)	1자(秭)	1024	1,000,000,000,000,000,000,000



1메기비0匝

=100만 바이트 한 스푼 정도의 모래



17 7世间里

= 10억 바이트 생수꽁 잘반을 제품 정도의 모래



THIS HIS IN

= 1조 바이트 국민주예65 제곱미터) 아과트에 10센터미터 깊이로 모래 쌓기



1端町町0厘

=1,000테라바이트 태운대 박사장의 모래



1액사바이트

=1,000페타이트 미국 메인 주에서 노스캐 플라이나 주까지 메인선의 모래(한반도 모든 메사장 모래의 함)



1제타바이트

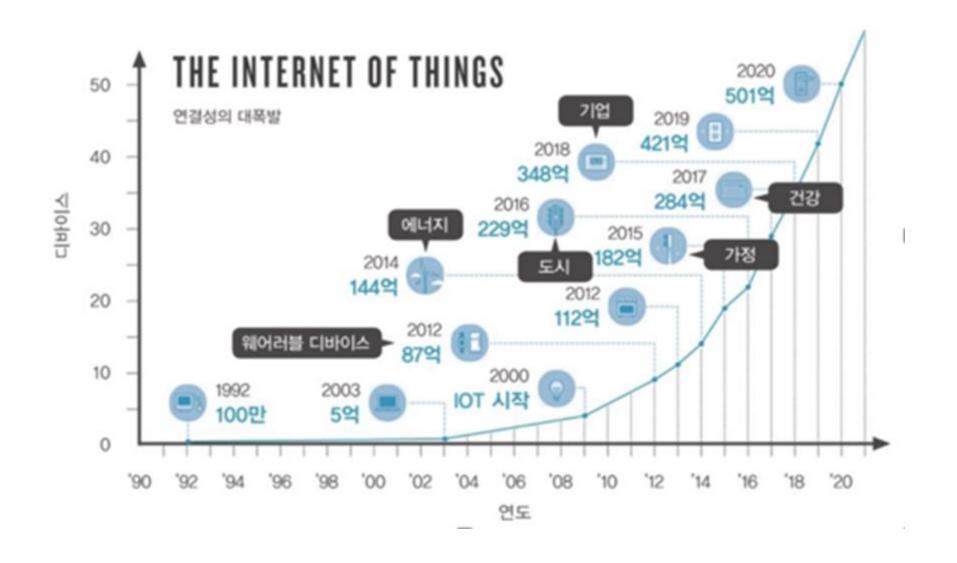
=1,000역사바이트 미국 전체 해안선의 모래 (한반도 모든 박사장 모래 함의 1,000배)



1分配性の無

=1,000제타바이트 미국 전체를 90미터 같이로 덮어버린 모래의 양

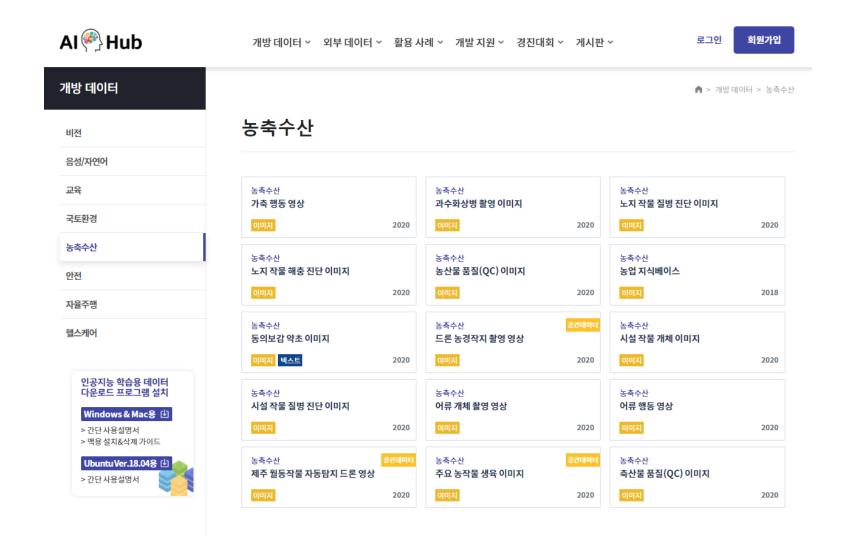










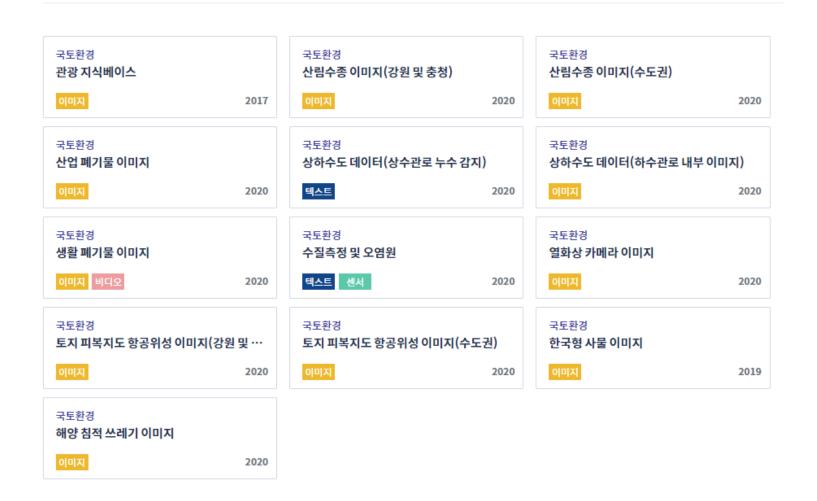




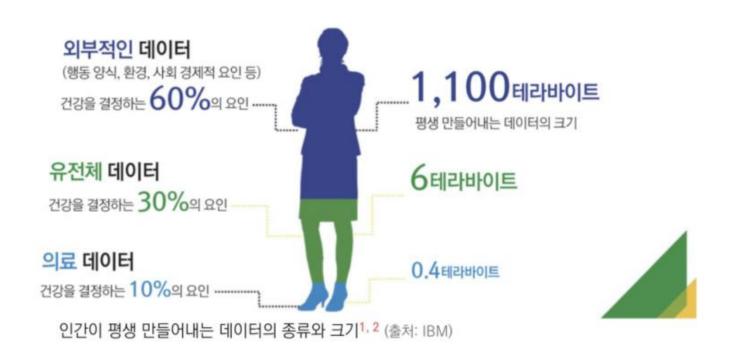




국토환경







측정 → 저장 → 통합 → 분석

근거 중심 의료

의료진(주체), 데이터가 측정되고 저장되는 공간(병원)

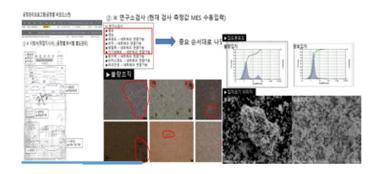
맞춤 의료 : 예방의료 + 예측의료

2003년 게놈 프로젝트 완료, 27억 달러 비용과 13년 기간 소요 2014년 기준, 게놈 분석 1,000달러 비용과 하루 소요

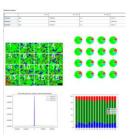








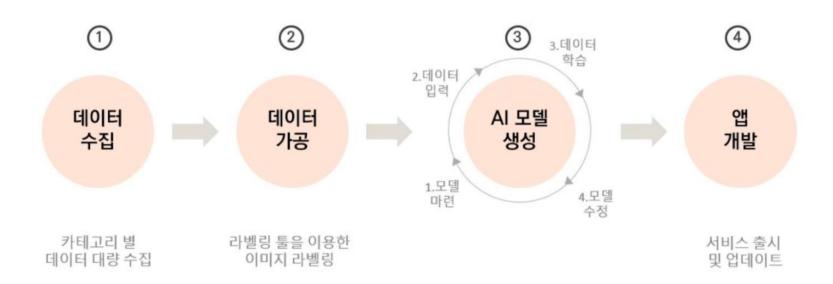




완성치수		
높이/내경 27.5	가로/의경 29.5	세로/길이 42
제품 구분		
	당 T:각형 v 재종 DR70S(S) v	
설계 및 공정		

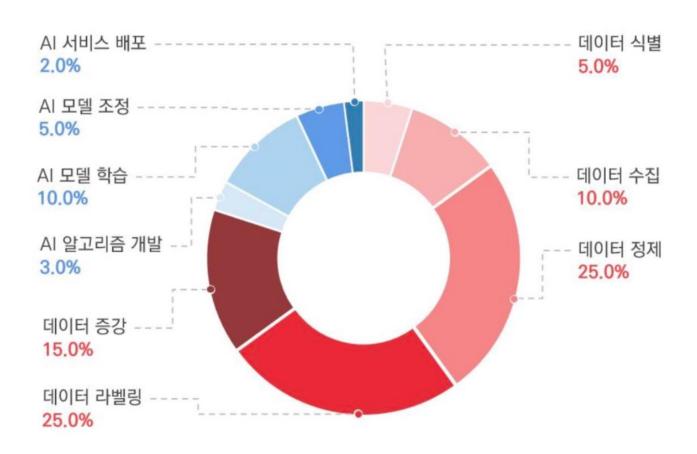


데이터 수집에서 서비스 개발까지...





AI 프로젝트에 소요되는 시간 비율



※ 자료: The Ultimate Guide to Data Labeling for ML, Cloudfactory, 재구성



AI 학습용 데이터 종류별 소요 비용

데이터 종류	수집	정제 및 라벨링	검수	모델검증	기타
의료영상	5%	60%	25%	10%	-
음성 데이터	50%	30%	20%	-	-
번역 엔진	30%	40%	10%	20%	
대화 데이터	40%	30%	10%	20%	-
CCTV 영상	30%	30%	10%	30%	-
헬스케어		30%	30%	20%	10%

※ 자료 : 국내 AI 학습용 데이터 구축 기업 FGI 인터뷰 자료 요약, 한국정보화진흥원(NIA) 작성, 2020



빅데이터 처리를 위한 필요 인력



데이터 사이언티스트, 데이터 아키텍트

• 총체적인 전략 설계 · 분석 · 담당

데이터 엔지니어, DB 관리자, 실무 개발자·관리자

• 데이터 산업 전반의 허리 역할

데이터 라벨러, 데이터 수집가 등

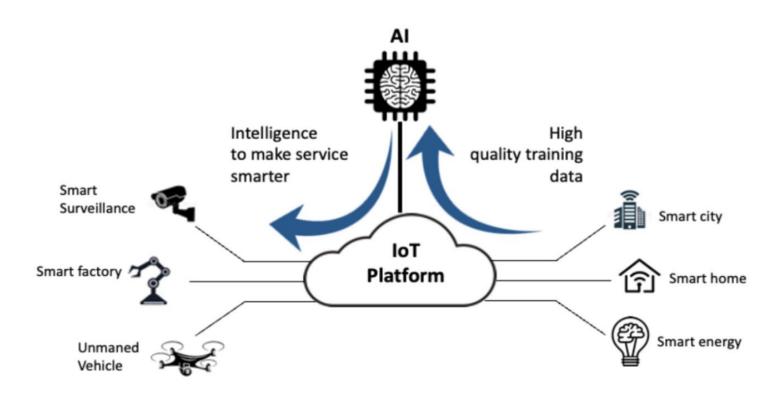
 정보를 수집 가공해 AI가 학습할 수 있는 형태로 가공하는 기초 작업

※ 자료 : 국내 AI 데이터 기업 FGI 인터뷰 자료 및 AI가 배울 자료 만드는 '뉴잡', 매일경제('20.6.21)로 재구성



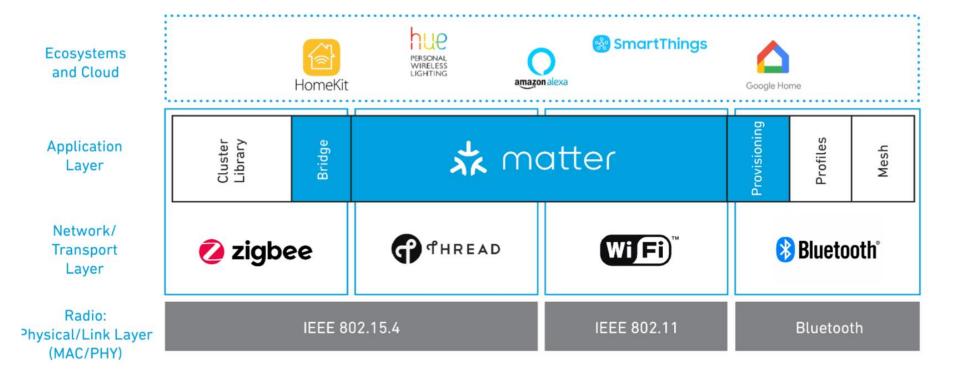






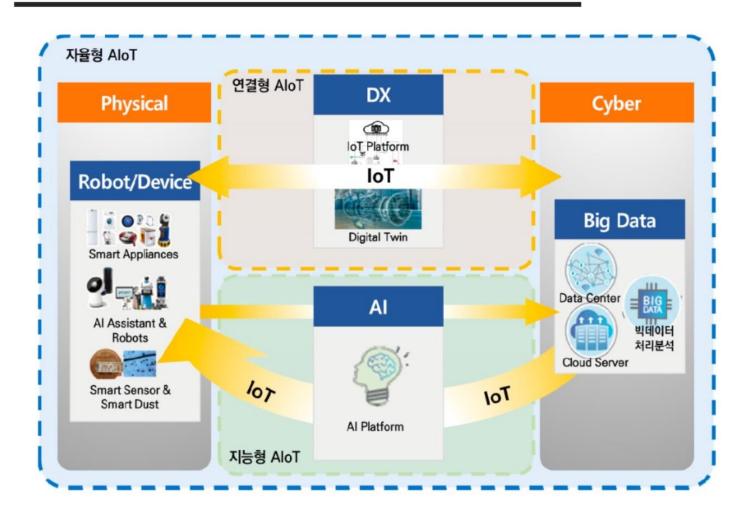


사물인터넷 연결의 새로운 표준 – matter



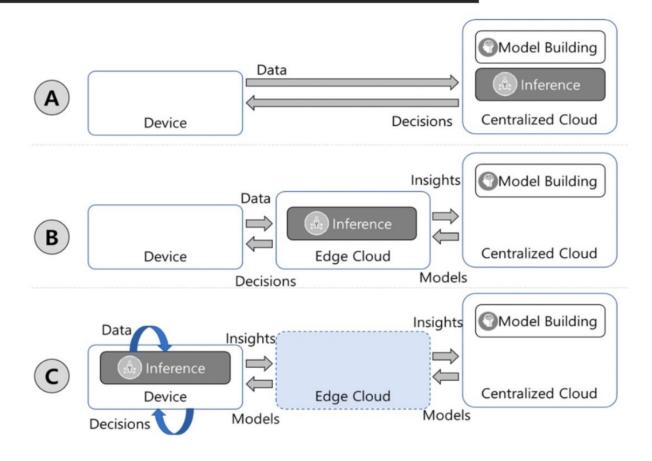


자율형 AloT(Al + loT)





자율형 AloT 기술 적용



〈자료〉 Seraphin B. Calo, Maroun Touna, Dinesh C. Verma, Alan Cullen, "Edge computing architecture for applying Al to IoT," IEEE International Conference on Big Data, 2017, pp.3012-3016.



Web 3.0 = AI + 메타버스 + 블록체인 + Web 2.0

WEB 3.0

XR(&메타버스)as the New OS, 탈중앙화 웹, 보안·개인정보 보호, 익명성, 크리에이터 경제, 로컬 경험과 커머스 등

WEB 2.0

모바일, OS(앱 경제), 구독 경제, 플렛폼 지배, 스트리밍, 공유 경제

WEB 1.0

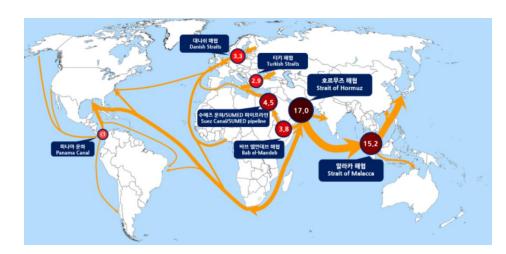
데스크탑PC, 브라우저, 배너광고, e커머스 시작, 사람들의 적응



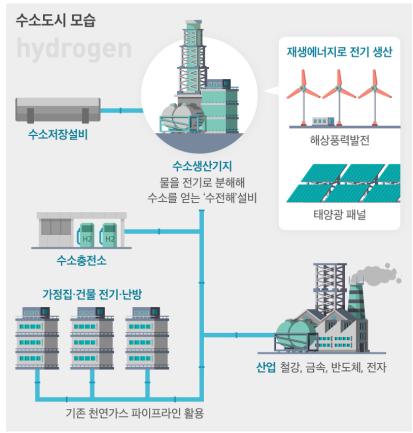
Web 3.0 = AI + 메타버스 + 블록체인 + Web 2.0

구분 특징	WEB 1.0	WEB 2.0	WEB 3.0
상호작동 (Interaction)	데스크탑 브라우저	모바일 터치 스크린	XR, 음성, IoT기기
	(클릭 & 타이핑)	(터치 & 스와이프)	(보여주거나 말하기)
컴퓨팅 위치 (Computation)	특정 장소 속 서버	클라우드 컴퓨팅	분산형 컴퓨팅
	(유선 망을 통해)	(3G ~ 4G)	(AI, 5G와 엣지컴퓨팅)
정보 저장 형태	구조형(Structured)	비구조화(Unstructured)	분산 원장 기술 (Distributed ledger tech)
(Information)	(SQL)	(빅데이터)	(블록체인)









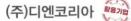














부동산 중개 서비스에서 시세 예측 서비스로 고도화

OVERVIEW

활용내용 및 성과	디엔코리아는 자체 부동산 선호 데이터와 지인플러스의 단지 특성에 따른 아파트 시세 데이터를 결합하여 고객이 현재 아파트 매도 호가가 적절한 수준인지 판단 할 수 있도록 부동산 시세 예측 서비스 제공
활용 데이터	전국 지역별 아파트 시세 데이터
데이터 제공기업	유통소비 빅데이터 플랫폼(지인플러스)

기업소개

데이터 기반의 과학적 행정 추진 및 정책 실효성 제고



디엔코리아는 부동산 중개 서비스 '동네'를 구축한 스타트업으로서, 사용자가 선호하는 맞춤 추천 서비스를 통해 고객이 원하는 집을 빠르고 정확 하게 연결해줍니다. 또한 3D VR 기반의 오픈 하우스 투어 서비스를 제공하여 고객이 부동산 매물을 현장감 있게 살펴볼 수 있도록 하고 있습니다.

[3D VR 투어]

데이터 활용사례

아파트 시세 데이터를 활용하여 실거래가 반영 '단지 추천 서비스' 제공

현재 국토교통부에서 제공하고 있는 아파트 실거래가 데이터는 부동산 중개서비스에 꼭 필요한 데이터입니다. 하지만 1개월의 데이터 갱신 주기로 인해, 아파트 실거래가 반영에 어려움이 있었습니다.

이에 비해 유통소비 빅데이터 플랫폼의 전국 지역별 아파트 시세 데이터는 아파트 단지 정보와 결합된 아파트 실거래가 데이터를 실시간으로 제공한다는 점에서 큰 차이가 있었습니다.

여기에 자체 부동산 선호 데이터를 결합하여 현재 아파트 매도호가가 적절한 수준인지 판단할 수 있는 모델을 개발함으로써 고객 성향별 부동산 추천 순위를 산정할 수 있었습니다.



['동네' 앱 서비스]

54 2021 빅데이터 플랫폼 활용 우수사례집

출처: 2021 활용 우수사례집 빅데이터플랫폼. NIA







주택/다가구 부동산 시세정보 제공을 통한 부동산 권리조사 효율 증대

OVERVIEW

활용내용 및 성과	부동산 권리조사 회사인 리파인은 수작업으로 진행되던 부동산 매물의 권리조사를 빅밸류의 부동산 정보 AP 호출을 통해 효율화하여 하루 350건 가량의 대출심사 처리 가능
제공 데이터	연립다세대 시세 산정 사례 데이터 연립다세대 예측 시세 정보, 주택 배치도 연립다세대 단독다가구 정제 정보 등
활용 기업	리파인

기업소개

공간 빅데이터 · AI 전문기업

리파인에 데이터를 제공한 빅밸류는 빅데이터와 인공지능을 통한 '부동산 금융의 혁신'을 목표로 설립한 스타트업입니다. 빅데이터와 인공지능 기술을 활용해 부동산 시세를 자동으로 산정하는 서비스를 개발하여. 주요 시중은행과 금융회사에 공급 중입니다. 빅밸류의 서비스는 기준시세가 없어 은행 대출이나 자산평가 시 애로를 겪었던 연립/다세대 주택소유자들의 고충을 해소했다는 평가를 받고 있습니다. 앞으로도 꾸준한 연구와 서비스 개발을 통해 공간정보에 전문화된 빅데이터와 인공지능 기술을 선도하는 기업이 되고자 합니다.



[빅밸류에서 제공하는 서비스]

에이터 활용사례

인근 유사 부동산 랭킹 알고리즘을 적용한 정확한 부동산 데이터 시세 산정

빅밸류는 주택실거래가, 건축물 대장, 토지대장 등 다양한 데이터를 기반으로 개발한 자체 유사도 랭킹 알고리즘에 인근 유사 부동산 정보(면적, 준공시기)를 반영하여 시세 정확도를 높인 부동산 시가추정 서비스를 개발하였습니다. 이 서비스는 매달 업데이트되는 공공정보를 수집, 저장하고, 재학습하여 최신 시세를 1개월 단위로 정기 제공하고 있습니다.



[시세산정 프로세스]

70 2021 박데이터 플랫폼 활용 우수사레집

출처: 2021 활용 우수사례집 빅데이터플랫폼, NIA

69

magicecole

스마트치안 빅데이터 플랫폼



주택도시보증공사



전국의 경찰관서 위치정보를 반영한 입지평가 고도화

OVERVIEW

활용내용 및 성과	주택도시보증공사는 경찰관서 위치정보 데이터를 활용하여 기존 관공서 위치를 포함한 주거환경의 평가점수를 자동으로 산출하고, 인근 사업장과 자동으로 조회 (비교)할 수 있도록 시스템 고도화
활용 데이터	지역별 파출소 위치정보
데이터 제공기관	스마트치안 빅데이터 플랫폼(경찰대학)

기관소개

주택도시 개발 보증리스크에 대비한 분양가 심사제도운영

주택도시보증공사는 주거복지 증진과 도시재생활성화를 지원하기 위한 각종 보증업무 및 정책사업을 수행합니다. 주택분양보증 심사업무의 일환으로 고분양가 심사제도를 운영하고 있으며, 보증심사는 입지평가, 사업안정성 등 다양한 항목을 평가합니다.

■ 데이터 활용사례

전국 경찰관서(치안센터) 위치 정보를 반영한 입지평가 고도화

주택도시보증공사는 보증심사의 일환으로 입지를 평가하며, 입지평가 항목에 관공서 위치 정보를 포함하고자 검토 중에 있었습니다. 그러나, 전국 경찰관서 위치정보 데이터는 여기저기 분산되어 있어 개별적으로 수집할 경우 많은 시간이 소요됩니다.

그러던 중 스마트치안 빅데이터 플랫폼을 통해 치안 관련 공공시설(경찰관서)의 정확한 위치 데이터를 얻을 수 있었고, 이를 입지평가에 반영하였습니다. 전국 18개 지방청 및 255개 경찰서의 위치와 건물경계 정보를 활용하여 관공서의 위치를 포함한 입지평가 요소를 반영하였습니다. 이를 바탕으로 각 항목을 평가한 최종 입지점수를 자동으로 산출하고, 인근 비교사업장을 조회할 수 있도록 시스템을 개선하였습니다.



[시각화 화면 예시]

출처: 2021 활용 우수사례집 빅데이터플랫폼, NIA



[보건의료 인공지능 플랫폼 개념도]

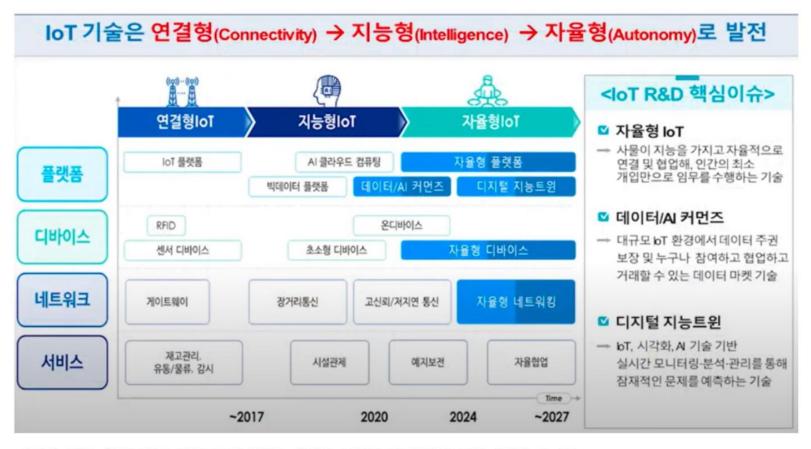






〈자료〉 테크월드, "현실 속 디지털 트윈 적용 사례", 2019. 6. 19.



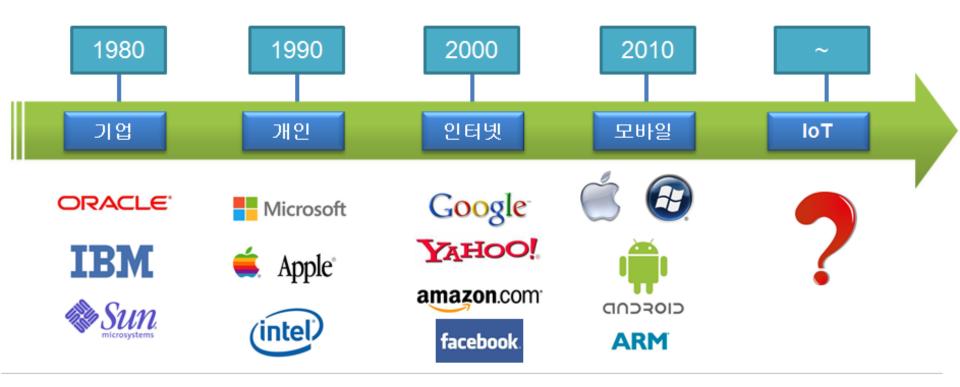


〈자료〉IITP, 「ICT R&D 기술로드맵 2025」 온라인 공청회-사물인터넷 분야, 2020. 7. 27.











감사합니다.

magic@magice.co