

빅데이터_기초

류영표

강사 소개

류영표



■ 근무 경력

- 동국대학교 수학과 졸 / 응용수학 석사수료
- 한국파스퇴르연구소 Image Mining 인턴(Deep learning)
- 전, (주)셈웨어(수학컨텐츠, 데이터 분석 개발 및 연구인턴)
- 현, Upstage AI 부스트캠프 멘토

■ 강의 경력

- 현대자동차 연구원 강의(인공지능/머신러닝/딥러닝/강화학습)
- 딥러닝 집중 교육과정 강사
- (재) 월튼블록체인 6일 과정(파이썬기초, 크롤링, 머신러닝)
- 서울특별시X AI 양재허브 X 모두의연구소 (중급 NLP과정) 보조강사
- SK아카데미_HLP(임원) 1차/2차 보조강사
- (주) 모두의연구소 Aiffel 1기 퍼실리테이터(인공지능 교육)

■ 주요 프로젝트 및 기타사항

- 제 4,5회 산업수학 스터디 그룹(산업문제 : 피부암, 유방암 분류)
- 제 1회 산업 수학 스터디 그룹(산업문제 : 질병에 영향을 미치는 유전자 정보 분석)
- 인공지능(AI)기반 데이터사이언티스트 전문가 양성과정 1기 수료
- 제1회 인공지능(AI)기반 데이터사이언티스트 전문가 양성과정 최우수상 수상(Q&A 챗봇)
- 빅데이터 여름학교 참석(문제 : 혼잡도를 최소화하는 새로운 노선 건설 위치의 최적화 문제)

빅데이터란?

빅데이터란 무엇인가?

빅데이터 정의

인공지능

빅데이터 기초

빅데이터란?

- 빅 데이터는 미래 경쟁력을 좌우하는 21세기의 원유이다
 - Gartner
- 빅데이터는 혁신, 경쟁과 생산성에 있어서 차세대 첨단 주자다
 - McKinsey
- 빅데이터는 화폐나 금처럼 새로운 자산이 될 것이다.
 - Davos Forum



빅데이터란?

BIG DATA ≠ 대용량 자료

- 조직의 내외부에 존재하는 다양한 형태의 데이터를 수집, 처리, 저장
- 목적에 맞게 분석함으로써 해당 분야의 필요 지식을 추출하고
- 이를 조직의 전략적 의사결정에 활용하거나 시스템화하여 상시적으로 생산성 향상에 활용하거나 새로운 비즈니스 모델의 창출에 활용하고자 하는 패러다임.

지식을 탐색하는 방법들

연역 :

이미 알고 있는 일반적인 지식,
법칙, 원리로부터 논리적인 규칙,
즉 합리적 추론에 따라 필연적
결론을 이끌어 내는 것

합리적인 방법

귀납 :

개별 사례들에 대한 관찰을 통
해 일반적인 결론을 이끌어
내는 접근방법
즉, 과거의 사례나 축적된 데
이터를 분석하여 현재의 문제
를 해결하는 접근법

경험주의 방법

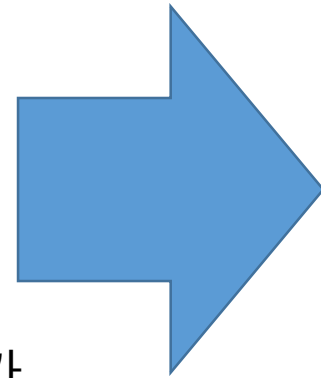


빅데이터는 전형적인 귀납적 지식 탐색 접근법에 해당

빅데이터의 출현 배경?

정보 환경의 변화

- 정보 기술의 발전에 따른 컴퓨터 활용 확대
- 소셜 미디어의 급격한 확산
- IOT 확산에 따른 센서 데이터 증대
- 멀티미디어 콘텐츠와 콘텐츠 사용에 관한 정보 증가
- 데이터 저장 및 처리기술의 발전



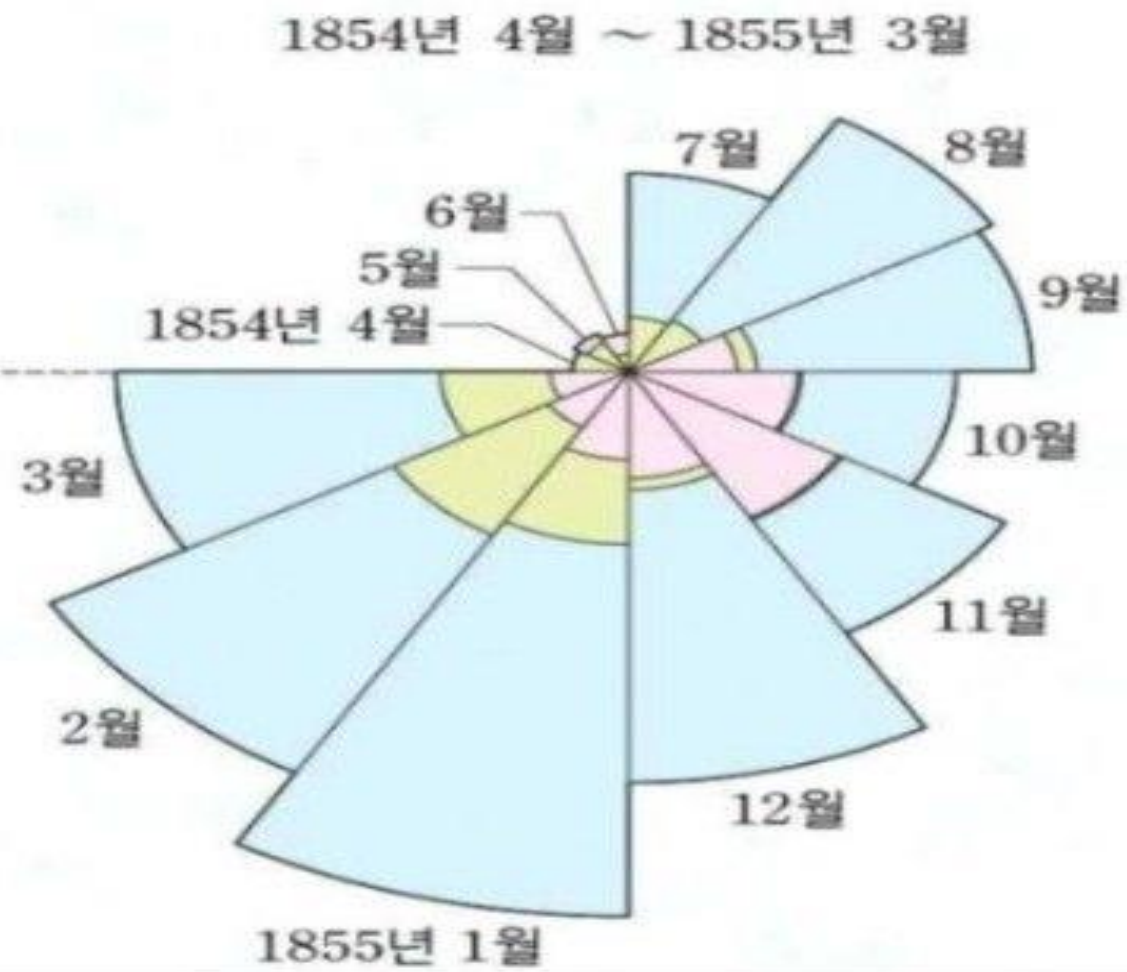
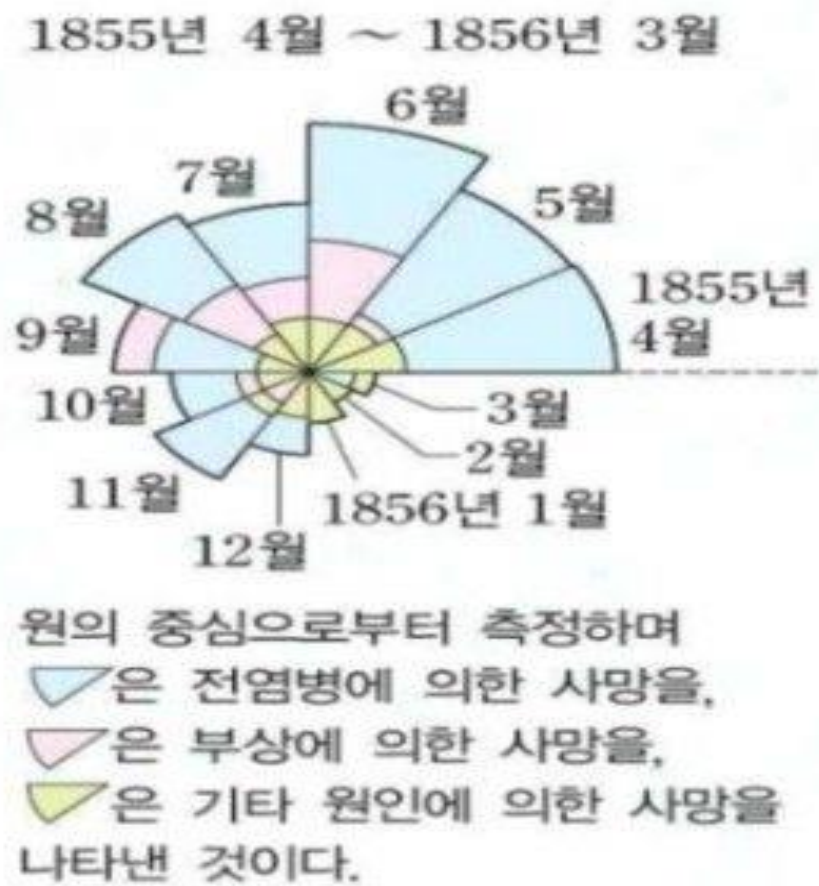
데이터 증가와
수집 비용의 감소

오래된 역사(빅데이터)

나이팅 게일도 사용한 데이터 분석

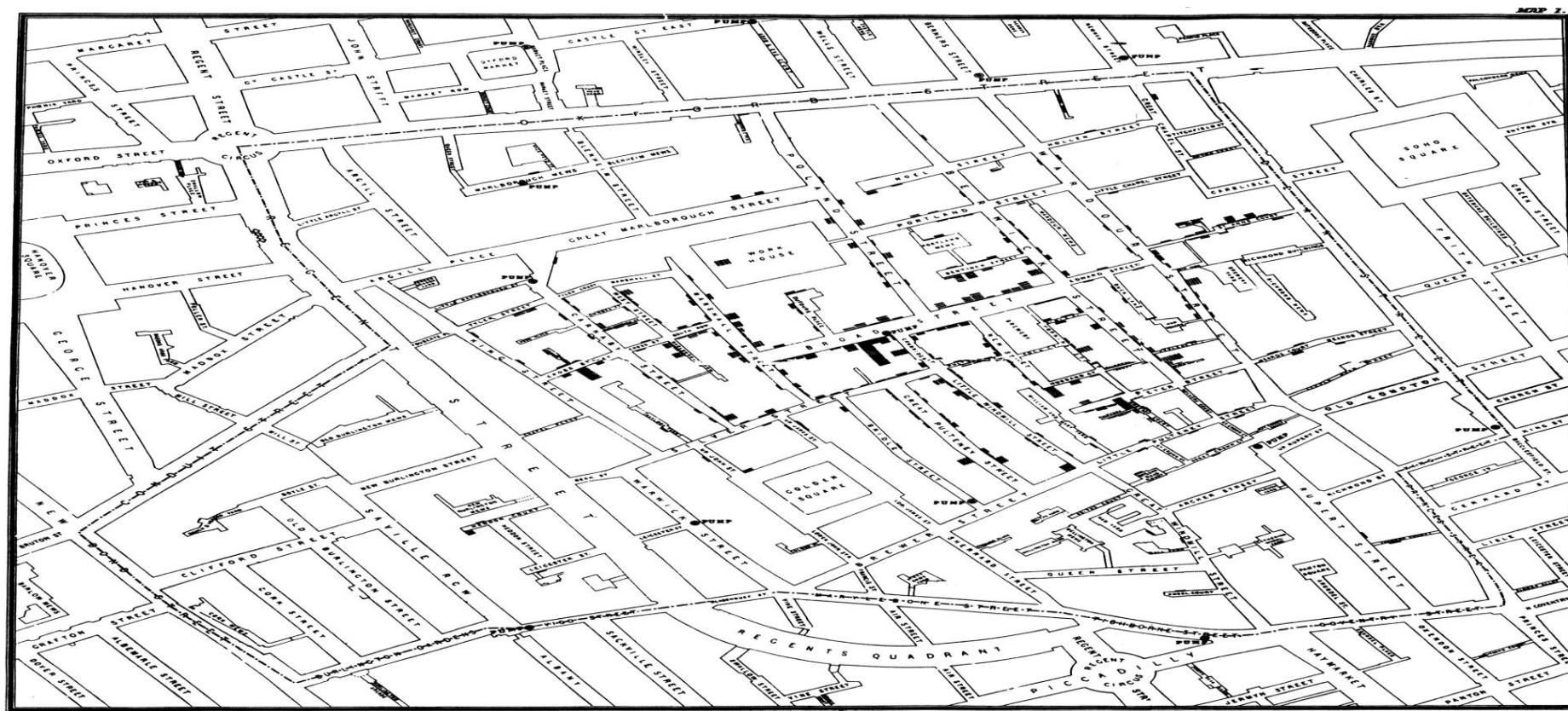
- 백의의 천사로 알려진 나이팅게일은 통계를 적극적으로 활용
- 1854년 러시아와 연합국 간에 '크림 전쟁' 발발
- 당시 런던에서 간호사로 일하고 있던 나이팅 게일은 병원 내 주 사망 원인을 데이터에 기반으로 분석하고, 전투로 인한 사망보다 전염병으로 인한 사망자 수가 많다는 사실을 밝혀 냄
- 이 정보는 병원 위생상태 개선이 매우 급선무임을 인식하게 하여 많은 생명을 구할 수 있었음.

오래된 역사(빅데이터)



오래된 역사(빅데이터)

전염병 시초였던 흑사병



데이터 분석과 비즈니스 인텔리전스

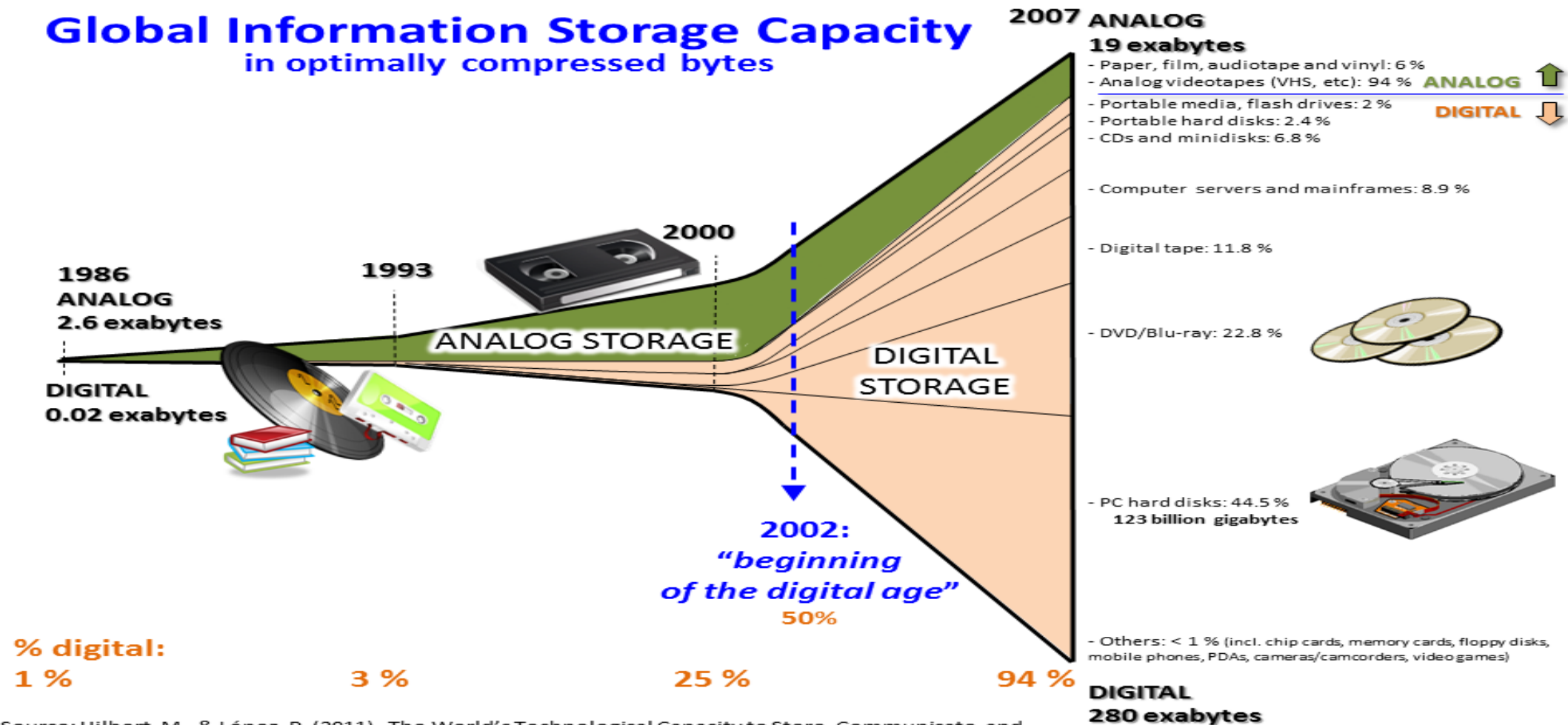
- 기업이 보유하고 있는 대량의 데이터를 수집, 정리, 분석해 이를 통해 도출된 정보를 기업의 의사결정에 활용하는 일련의 프로세스
- 기업의 사용자가 더 좋은 의사결정을 하도록 데이터를 수집, 저장, 분석, 접근을 지원하는 기술이자 응용시스템
- Data Mining 과 같은 데이터 분석 기술.

전통적인 데이터 분석과 빅데이터 분석의 차이



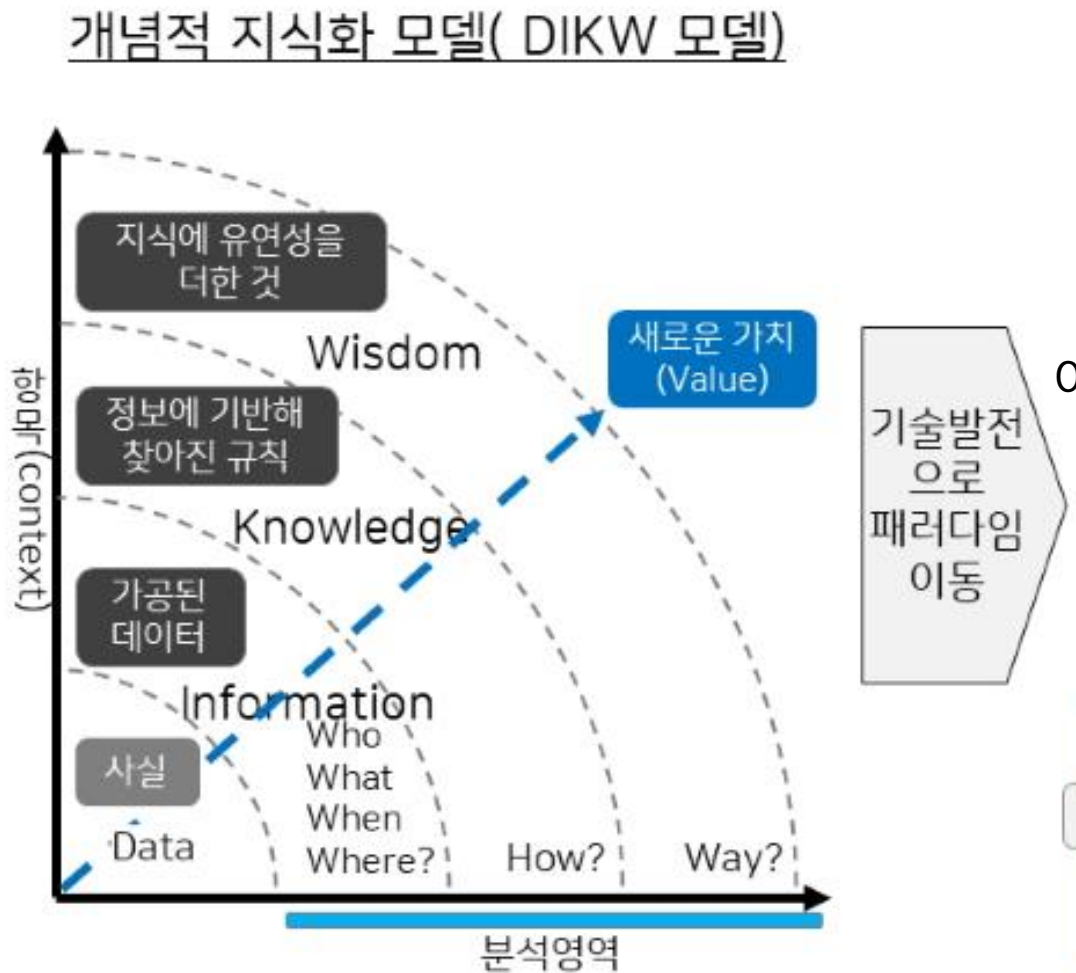
출처: 삼성경제연구소 '빅데이터: 산업 지각변동의 진원'

전통적인 데이터 분석과 빅데이터 분석의 차이

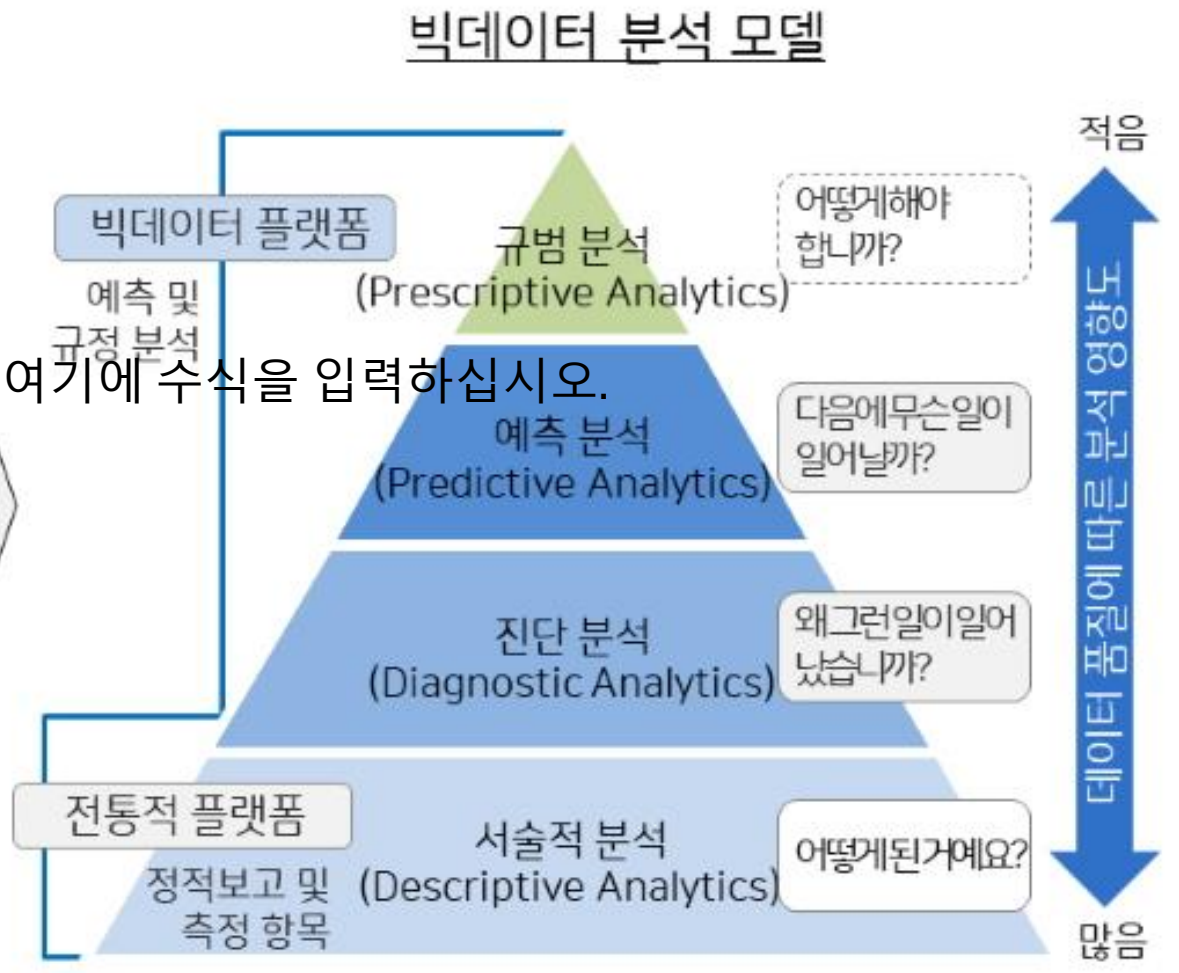


https://upload.wikimedia.org/wikipedia/commons/7/7c/Hilbert_InfoGrowth.png

전통적인 데이터 분석과 빅 데이터 분석의 차이

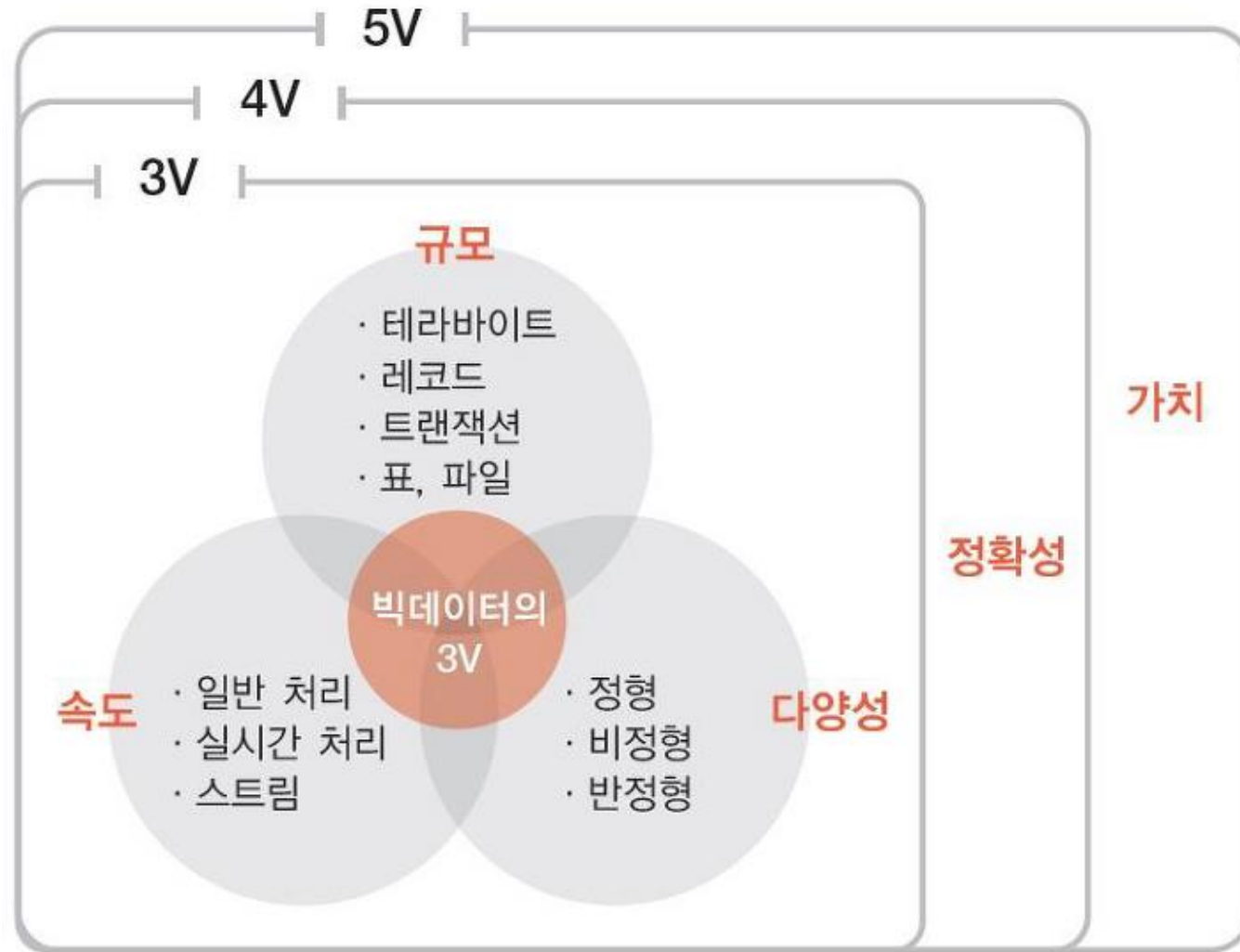


출처 : 로리 제니퍼 (2007). "지혜 계층 : DIKW 계층의 표현 " 자료 재구성

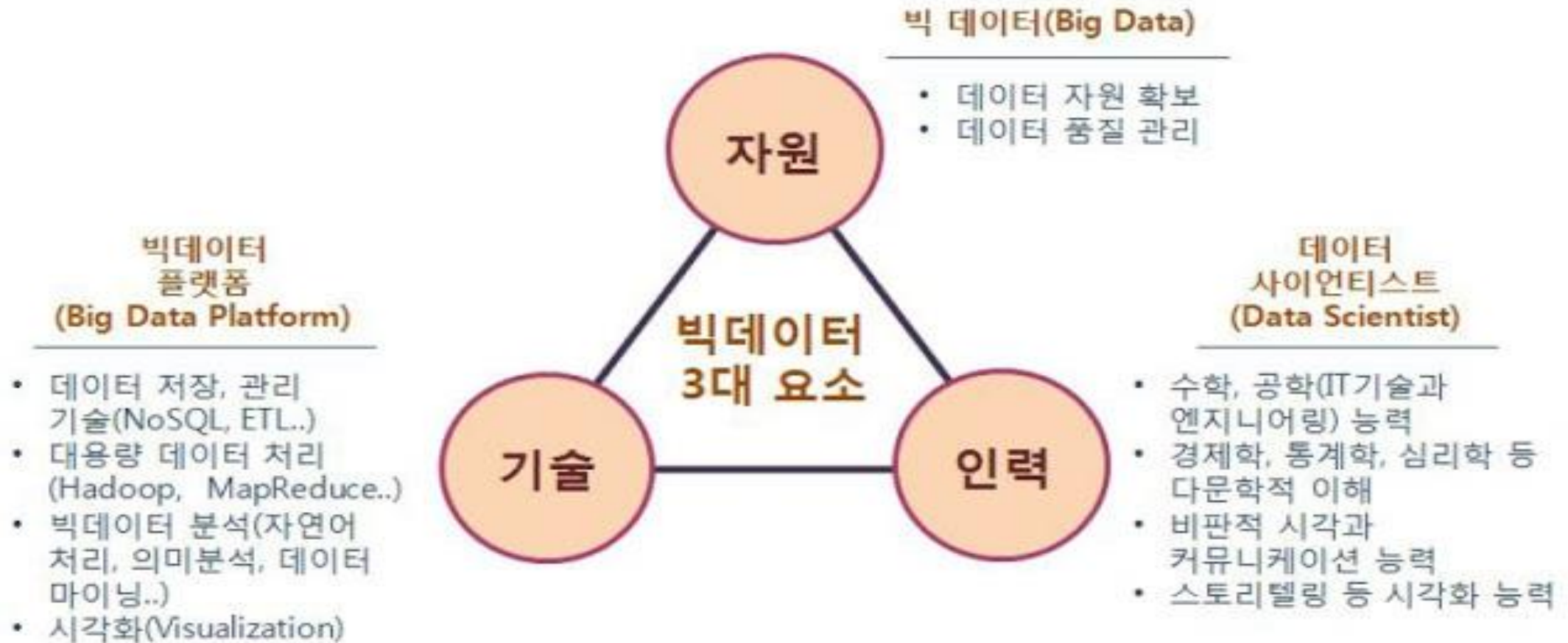


출처 : Gartner 자료 재구성

전통적인 데이터 분석과 빅 데이터 분석의 차이



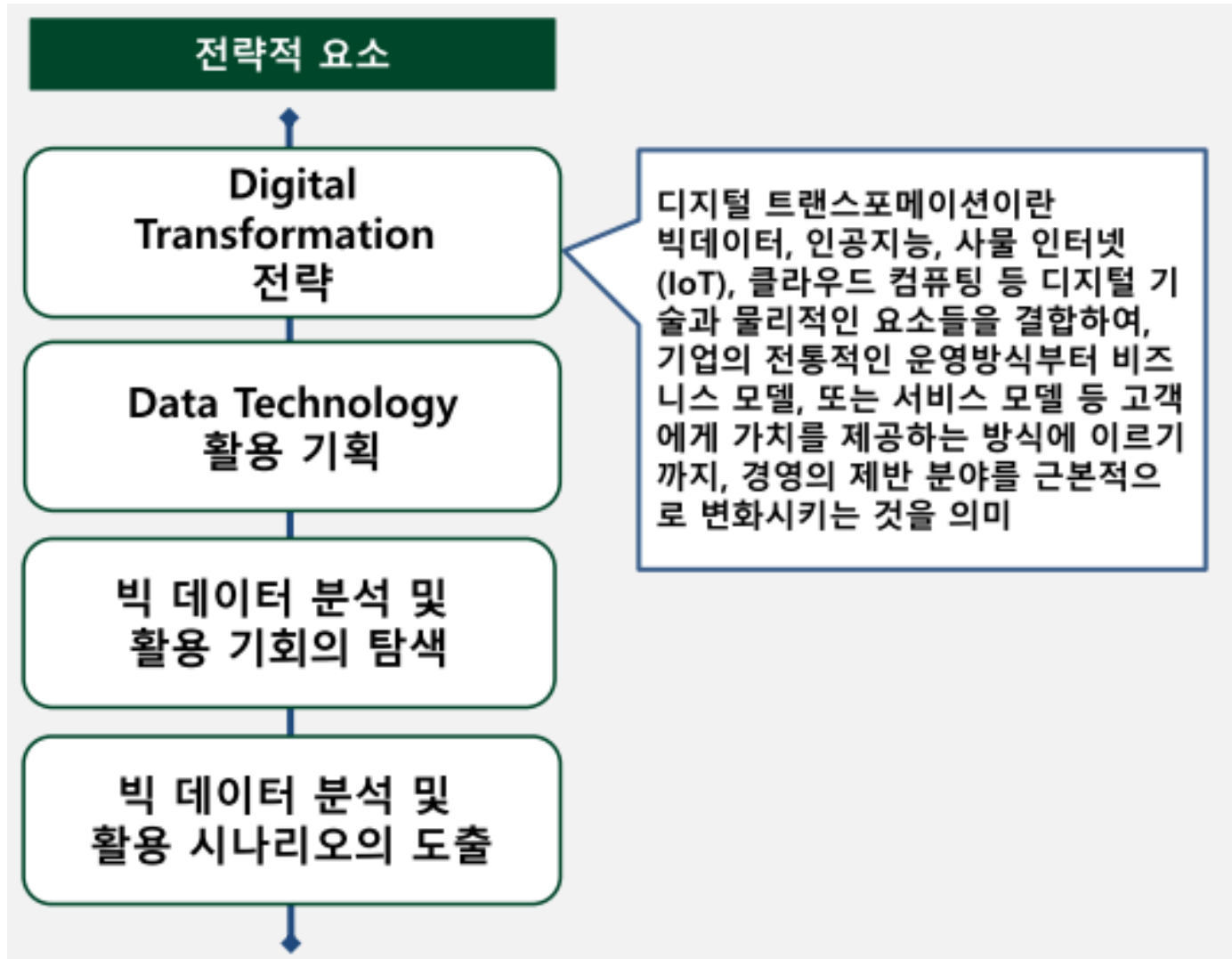
빅데이터의 주요 요소 3가지



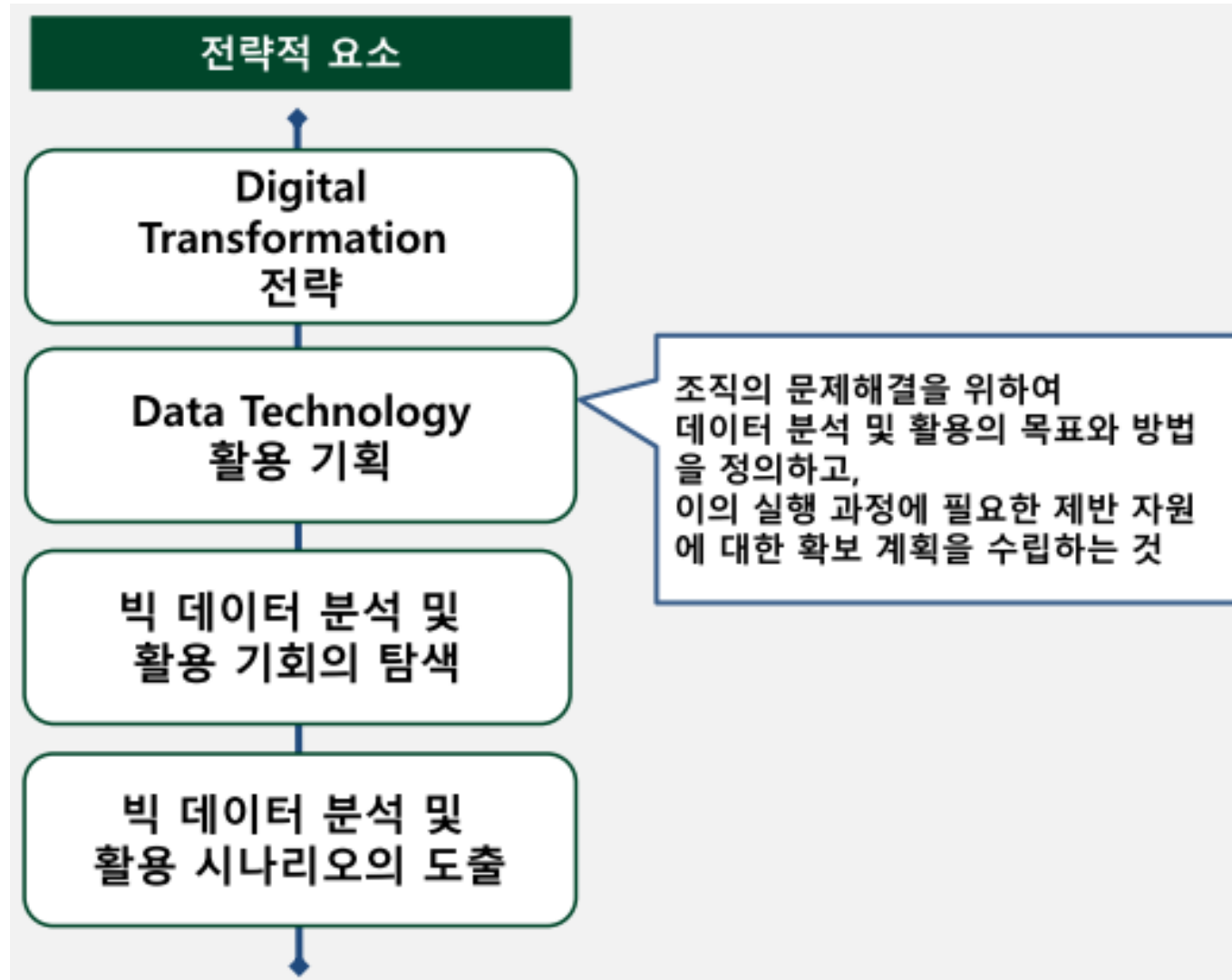
빅데이터의 주요 요소



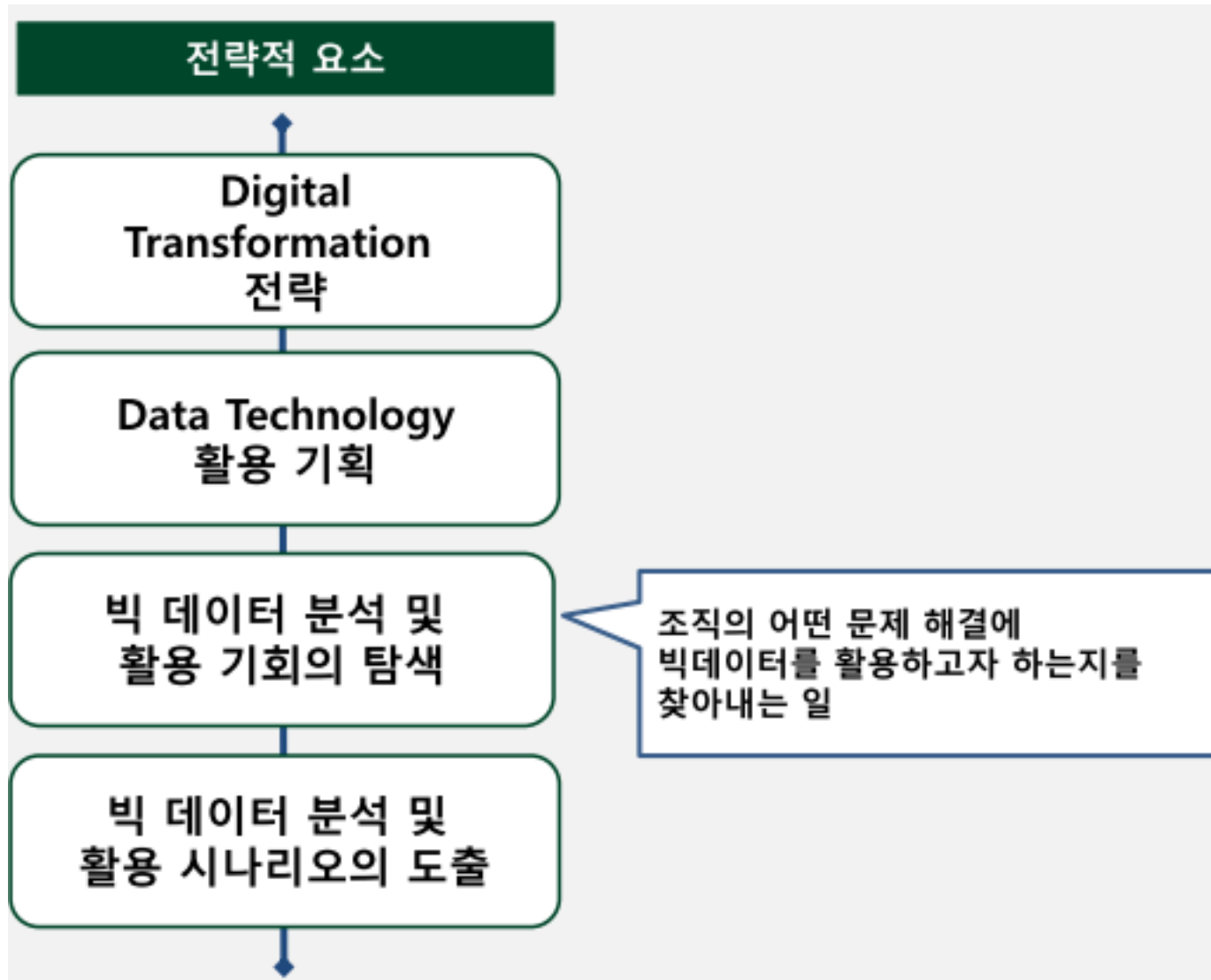
빅데이터의 주요 요소



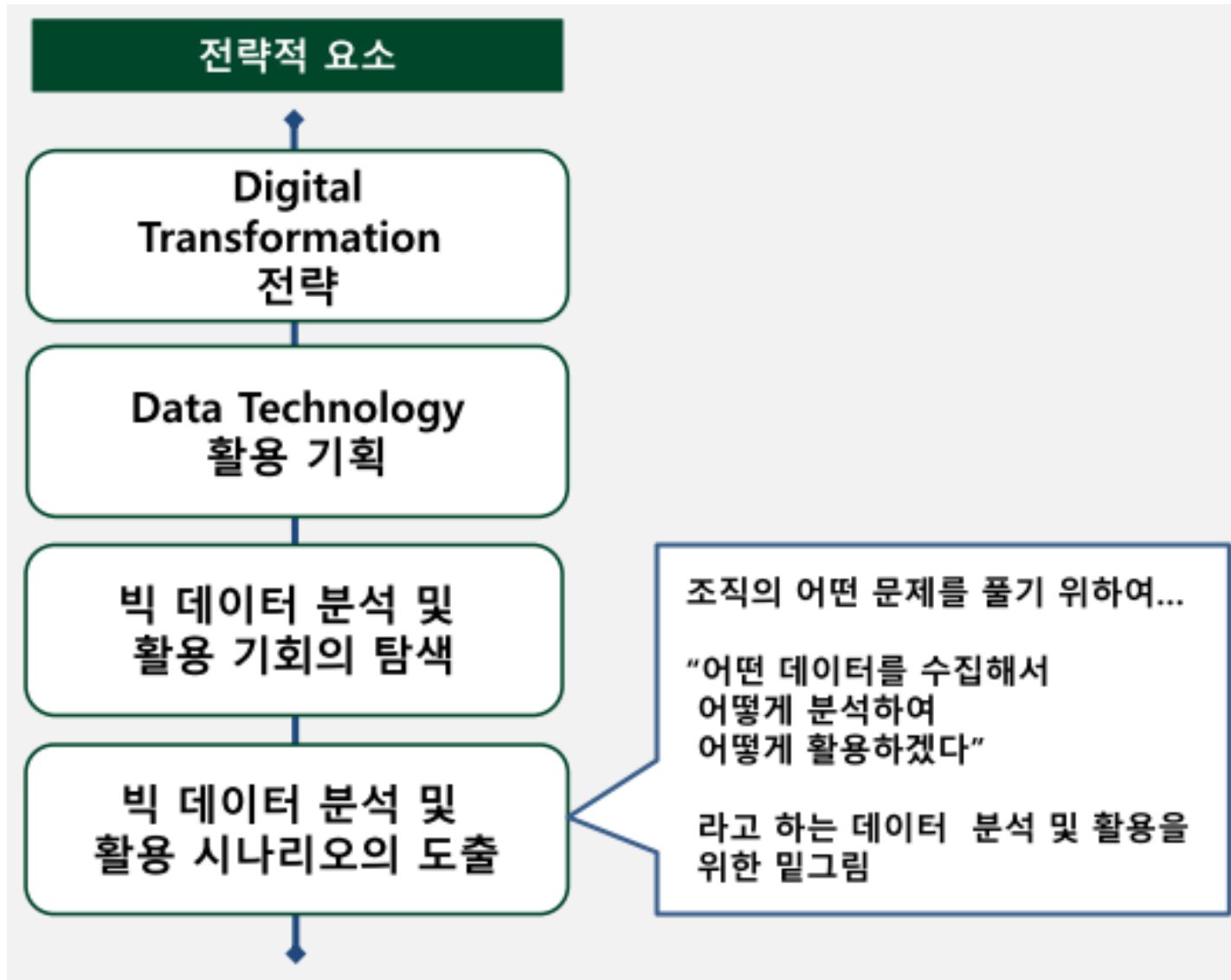
빅데이터의 주요 요소



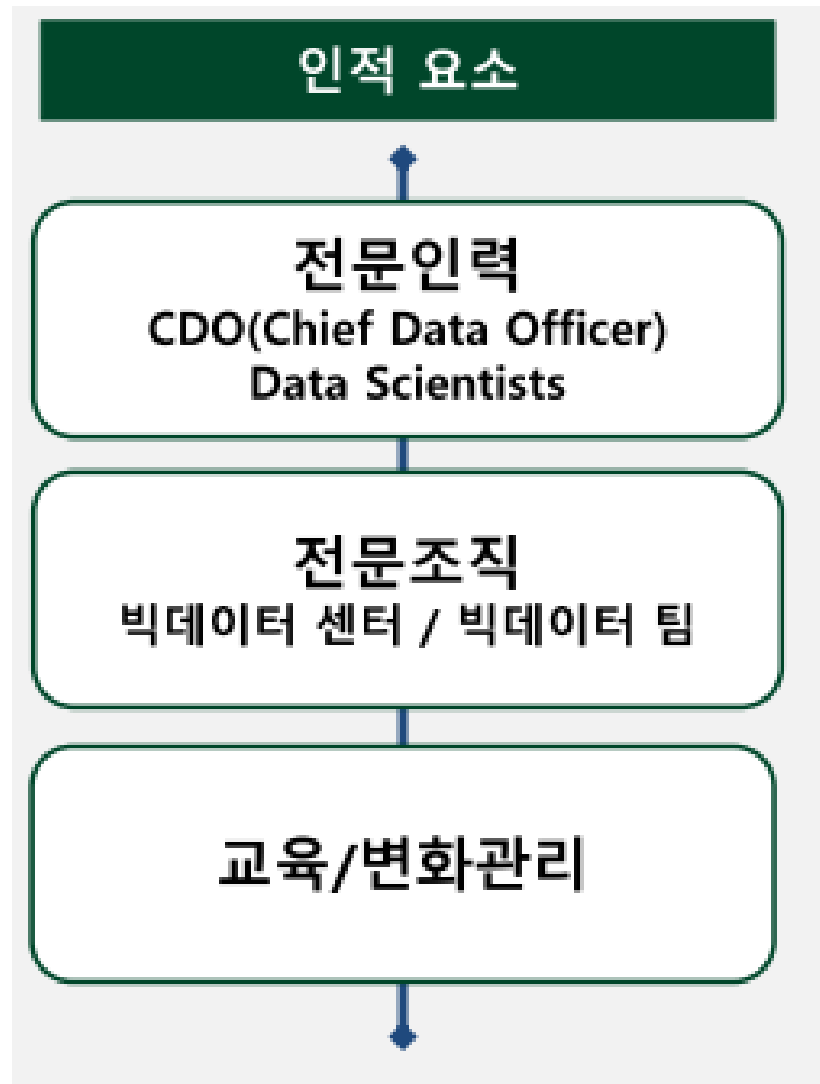
빅데이터의 주요 요소



빅데이터의 주요 요소



빅데이터의 주요 요소

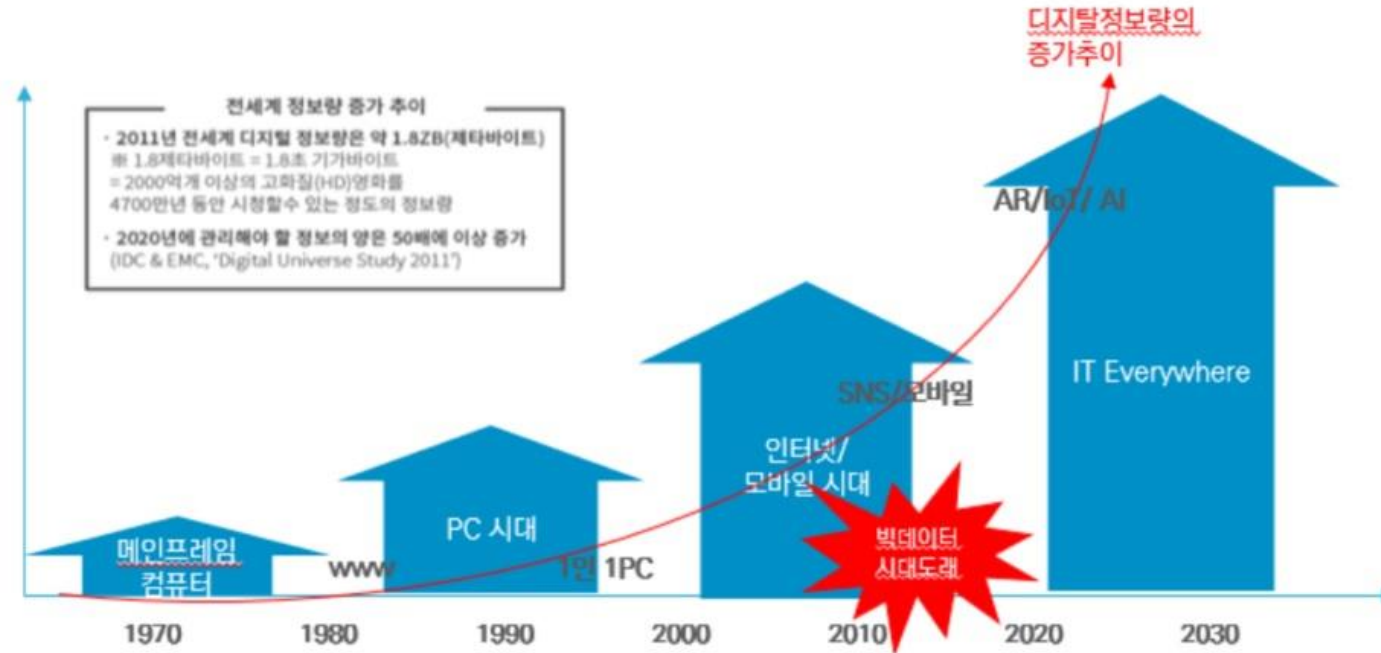


데이터 과학이란?

데이터 과학이란



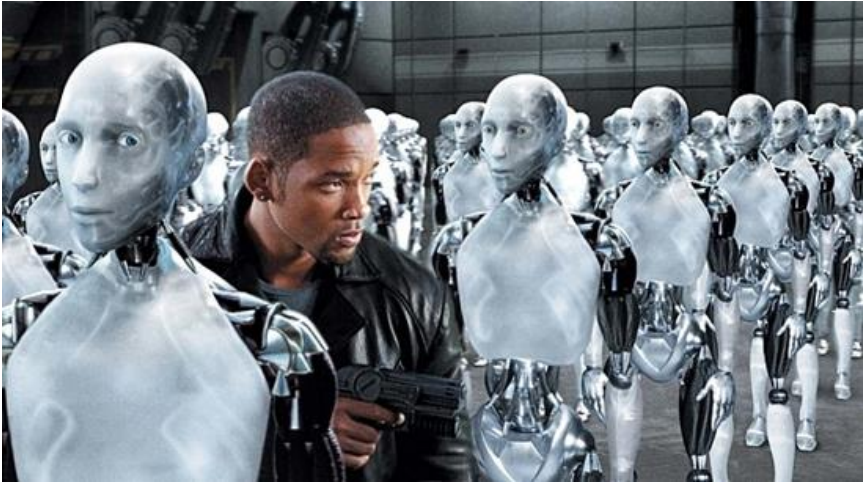
빅데이터 와 인공지능



데이터 중심 사회

데이터 규모	EB(Exa Byte) 90년대 말 100EB	ZB(Zeta Byte) 진입 2011년 1.8ZB	ZB(Zeta Byte) 본격화 2020년 은 2011년의 5배
데이터 유형	정형 데이터 (데이터베이스, 사무정보)	비정형 데이터 (이메일, 멀티미디어, SNS)	사물정보, 인지정보 (RFID, Sensor, IoT)
데이터 특징	구조화	다양성, 복합성, 소셜	현실성, 실시간성

인공지능



<https://www.hankyung.com/theopen/article/105648>



<http://sisa-n.com/View.aspx?No=85159>



<https://byline.network/2018/09/17-24/>



<http://dongascience.donga.com/news.php?idx=10919>

체스 로봇



인공지능

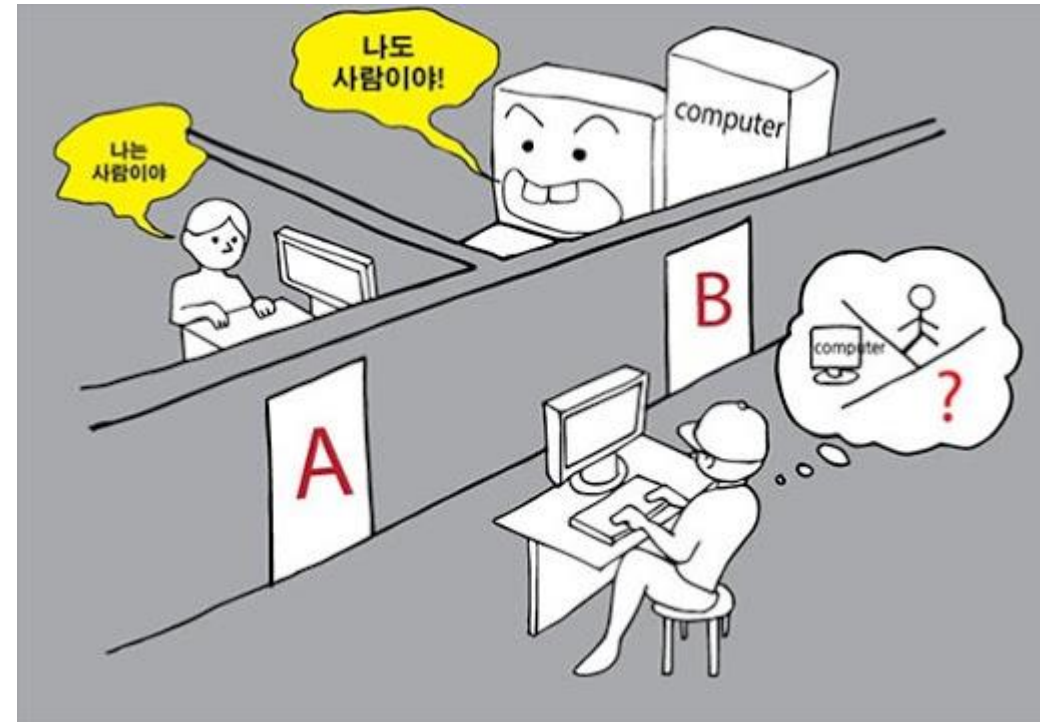
- 지능(Intelligence)

- 본능적으로 자동적으로 행동하는 대신에, 생각하여 이해하고 행동하는 능력

- 인공지능(Artificial Intelligence)

- 기계로부터 인공적으로 만든 지능
- 튜링테스트(Turing Test)

-> 조작적 정의 : 측정할 수 있는 조건으로 어떤 속성을 기술.



인공지능의 역사

- Dartmouth Conference(1956, Dartmouth University, USA)
 - John McCarthy(존 메카시) : <지능이 있는 기계를 만들기 위한 과학과 공학> 이라는 논문에 인공지능 처음 등장.

1956 Dartmouth AI Project



Five of the attendees of the 1956 Dartmouth Summer Research Project on AI reunited in 2006: Trenchard More, [John McCarthy](#), [Marvin Minsky](#), [Oliver Selfridge](#), and [Ray Solomonoff](#). Missing were: [Arthur Samuel](#), [Herbert Simon](#), [Allen Newell](#), [Nathaniel Rochester](#) and [Claude Shannon](#).

<https://slideplayer.com/slide/13045745/>

인공지능 정의

- 인공지능(Artificial Intelligence)

- 사람의 생각과 관련된 활동, 예를 들면 의사 결정, 문제 해결, 학습 등의 활동을 자동화 하는 것

(Bellman, 1978)

- 사람이 하면 더 잘 할 수 있는 일을 컴퓨터가 하도록 하는 방법을 찾는 학문(Rich & Knight, 1991)
- 지능적인 에이전트를 설계하는 학문(Poole et al. 1998)
- 인지하고, 추론하고, 행동할 수 있도록 하는 컴퓨팅에 관련된 학문(Wilson, 1992)
- 인공물이 지능적인 행위를 하도록 하는 것(Nissson,1990)

Strong AI VS Weak AI

- **강한 인공지능(Strong AI)**

- 사람과 같은 지능
- 마음을 가지고 사람처럼 느끼면서 지능적으로 행동하는 기계
- 추론, 문제해결, 판단, 계획, 의사소통, 자아 의식(Self-awareness), 감정(Sentiment), 지혜(sapience), 양심(conscience)
- 튜링 테스트

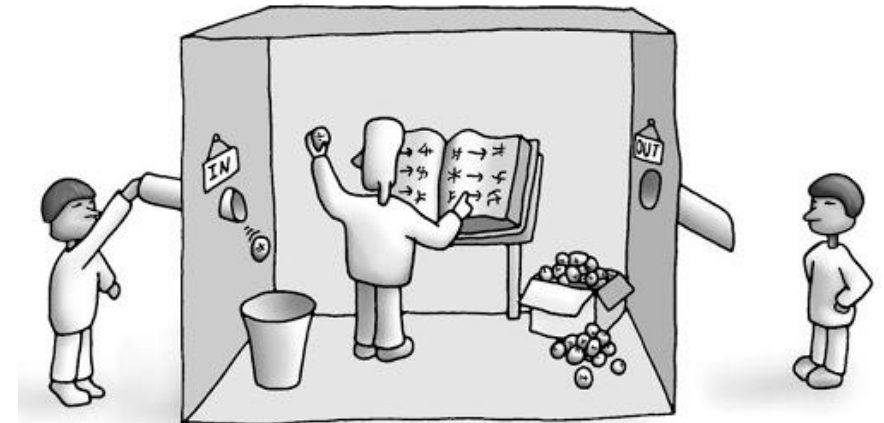
- **약한 인공지능(weak AI, narrow AI)**

- 특정 문제를 해결하는 지능적 행동
- 사람의 지능적 행동을 흉내 낼 수 있는 수준
- 대부분의 인공지능 접근 방향
- 중국인 방 사고 실험(Chinese room thought experiment)

Strong AI VS Weak AI

• 중국인 방 사고실험(The Chinese Room Thought Experiment)

- John Searle (1980) 제시
 - 문 밑으로 중국어로 쓴 질문지를 전달
 - 방 안에서 중국어를 모르는 사람이 글자모양에 따른 중국어 단어 조합 방법 매뉴얼을 참조하여 답변에 대한 단어 조합
 - 조합된 단어들을 문 밖으로 내보냄
 - 문 밖 사람은 중국어를 이해하는 사람이 방안에 있다고 생각
 - 단지 흉내만 내고 이해하는 것은 아님
- 이해하지 못하고 흉내 낼 수 있어도 지능적(intelligent) 행동.



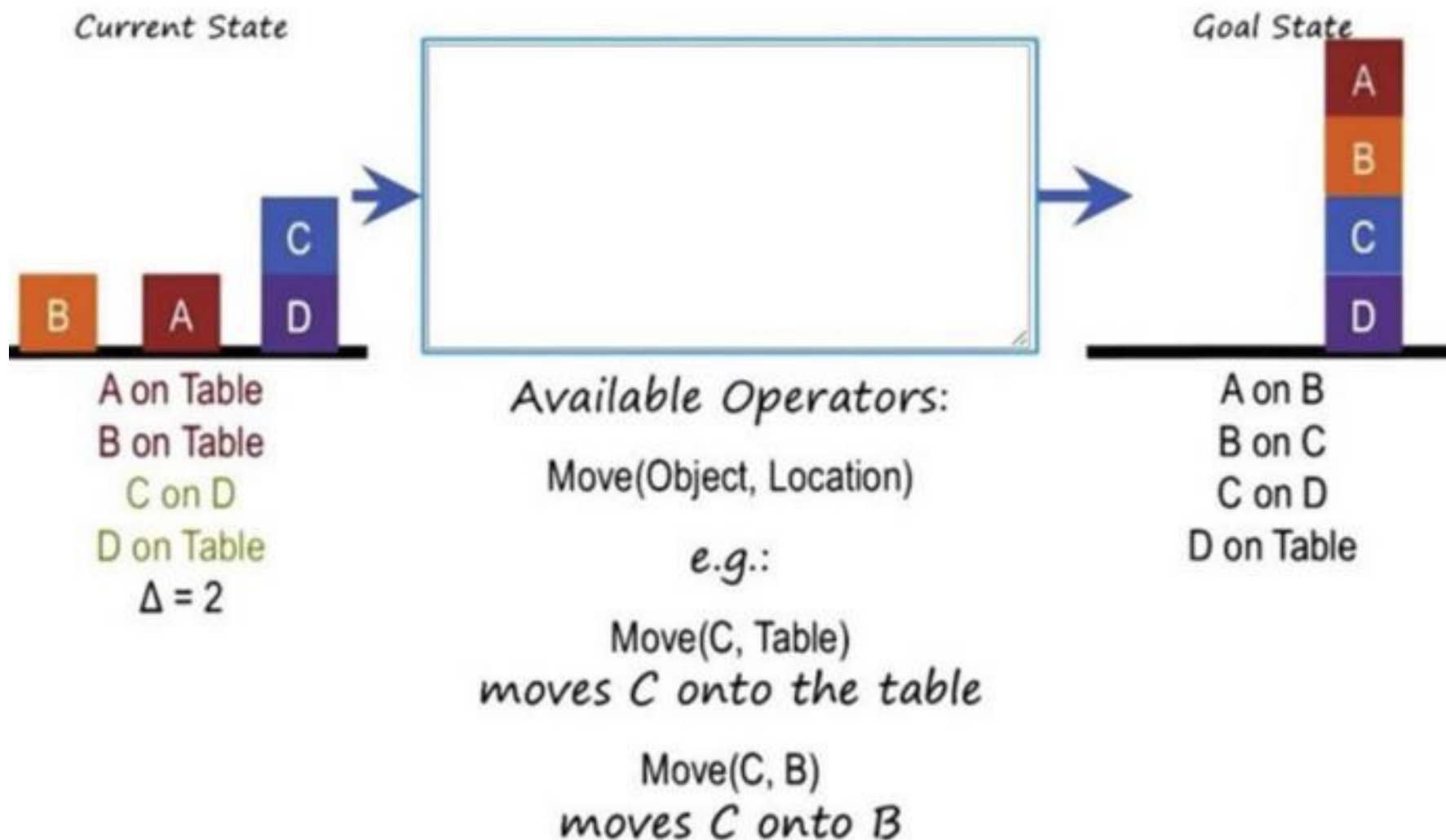
인공지능의 역사

•1960년대 이전

- 큰 기대와 여러 가지 시도, 매우 제한된 성공
- Advice Taker – 공리(axiom) 기반 지식표현 및 추론 (McCarthy)
- Perceptron – 신경망 모델 (Rosenblatt, 1958)
- 수단 – 목표 분석(means-ends analysis) 기법(Newell & Simon, 1958)
 - 범용 문제해결을 목표로 한 GPS(General Problem Solver) 개발
 - 해결해야 하는 문제를 상태(State)로 정의
 - 현재 상태와 목적 상태(goal state) 간의 차이 계산
 - 목적 상태로 도달하기 위한 조작자(Operator, 연산자) 선택 적용하는 과정 반복



블록이동 문제



AI 황금기(1956 ~ 1970년대)

- 다트머스 컨퍼런스 이후에, AI 라는 새로운 영역은 발전의 땅을 질주하기 시작
- 이 기간에 만들어진 프로그램 많은 사람은 '놀랍게(astonishing)' 만들었는데, 프로그램은 대수학 문제를 풀었고 기하학의 정리를 증명했으며 영어를 학습하였다.
- 몇몇 사람들은 이와 같은 기계의 '지능적' 행동을 보고 AI로 모든 것이 가능할 것이라 믿음.

AI 암흑기(1970년대 ~ 1980년대)

- 70년대 이르자, AI는 비판의 대상이 되었고, 재정적 위기가 올
- AI 연구자들은 그들의 눈앞에 있는 복잡한 문제를 해결하는데 실패함
- 연구자들의 엄청난 낙관론은 연구에 대한 기대를 매우 높여놓았고, 그들이 약속했던 결과를 보여주지 못하자, AI에 대한 자금 투자는 사라짐.

인공지능의 봄(1980년대)

- 분야 전문가의 '지식'을 if ~~~ then 의 형태로 추출하여, 컴퓨터에 저장하고, 추론 기능을 부여, 전문가의 지식체계 및 문제해결을 모사하는 시스템을 구축
- 1980년대에는 전 세계적으로 사용 된 '전문가 시스템'이라고 일컫는 인공지능 프로그램의 형태였고, 인공지능 검색에 초점이 맞춰짐
- 같은 시기에 일본 정부는 자신들의 5세대 컴퓨터 프로젝트와 인공지능에 적극적으로 투자
- 1980년대에 존 홉필드와 데이비드 루멜하트의 신경망이론의 등장.

모라벡의 역설(Moravec's Paradox)

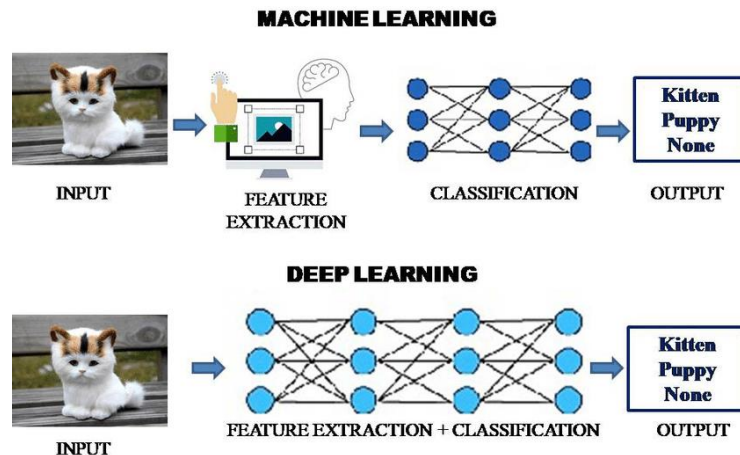
- 인간에게 쉬운 것은 컴퓨터에게 어렵고, 반대로 인간에게 어려운 것은 컴퓨터에게 쉽다는 역설
- 1970년대, 미국의 로봇 공학자인 한스 모라벡(Hans Moravec)이 '어려운 일을 쉽고, 쉬운 일은 어렵다' 표현으로 컴퓨터와 인간의 능력 차이를 역설적으로 표현
- 인간은 느끼기, 듣기, 보기, 의사소통 등의 일상적인 행위는 매우 쉽게 할 수 있는 반면, 복잡한 수식 계산 등을 하기 위해서는 많은 시간과 에너지를 소비
- 컴퓨터는 인간이 하는 일상적인 행위를 수행하기 매우 어렵지만 수학적 계산, 논리 분석 등은 쉽게 수행 가능

The Second AI Winter(1987 ~ 1993)

- AI와 비즈니스 커뮤니티의 매력은 상실했고 경제적 거품이라는 고전적 형태의 1980년대에 빠짐
- 1974년 전문가 시스템에 대한 열정이 통제할 수 없을 정도로 퍼져나가고 이에 대한 실망이 확실히 따라 올 것이라는 걱정이 있었고, 이 때 투자가 끊기고 살아 남은 연구들에 의해서 'AI Winter'이라는 단어가 생김

제3차 붐(2010년대)

- 빅데이터와 기계학습, 딥러닝 귀납에 의한 지식획득 가능성 극대화
- 비정형 데이터의 활용 가능성 극대화
- 심층학습(Deep learning)의 발전
- 상업적 성공 사례 다수 출현(자율주행차, 음성비서, 왓슨 등)



https://www.researchgate.net/figure/Machine-learning-vs-deep-learning_fig2_338372388

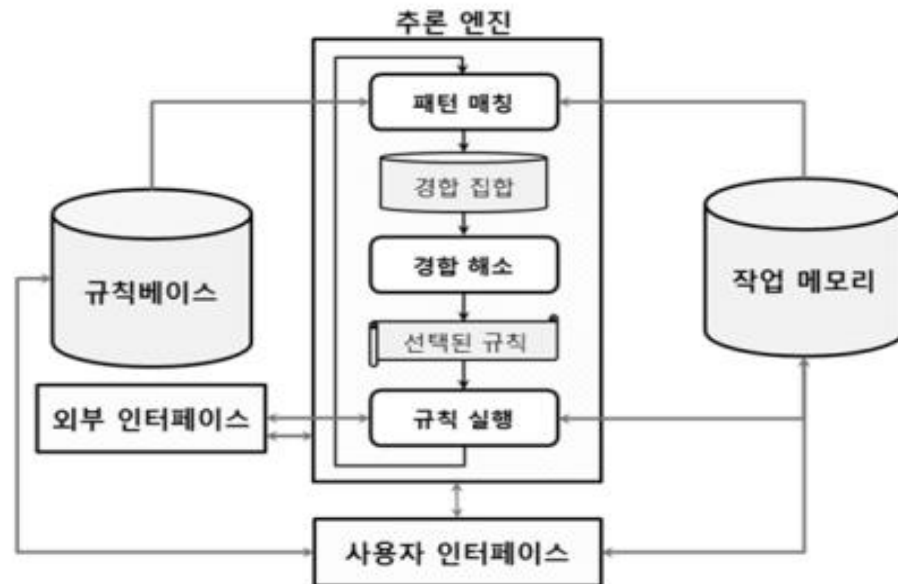
<https://www.yeongnam.com/web/view.php?key=20171208.010110752310001>

알파고 : 바둑 인공지능(2016.03)

응용 분야

• 지식 기반 시스템(Knowledge-based system)

- 지식을 축적하고, 이를 이용하여 서비스를 제공하는 시스템
- 전문가 시스템(Expert System)
- 특정 문제 영역에 대해 전문가 수준의 해법을 제공하는 것
- 간단한 제어시스템에서부터 복잡한 계산과 추론을 요구하는 의료진단, 고장진단, 추천 시스템 등



응용 분야

- 컴퓨터 비전(Compter Vision)

- 시각적 세계를 해석하고 이해하도록 컴퓨터를 학습
- 카메라, 동영상에서 디지털 이미지를 고전적인 컴퓨터 비전 알고리즘이나 딥러닝 모델을 사용하여 객체를 정확하게 식별하고 분류하는 학습.
- 안면인식 프로그램
- 자동차 보험 처리



응용 분야

- 컴퓨터 비전(Compter Vision) 의 최종적인 목표 -> 자율주행차

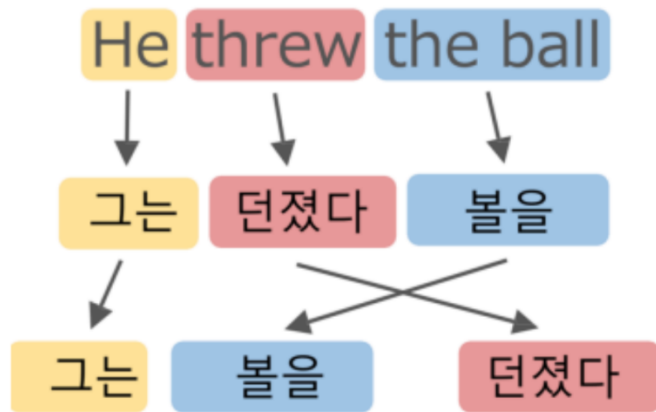


응용 분야

•자연어 처리(Natural language processing)

- 사람이 사용하는 일반언어로 작성된 문서를 처리하고 이해하는 분야
- 문서요역, 기계번역

문구기반 번역의 예



단어의 모호성



응용 분야

- 자연어 처리(Natural language processing) -> 글을 새롭게 창조.



응용 분야

- 음성 인식(Speech recognition)

- 사람이 음성 언어를 컴퓨터가 해석해 그 내용을 문자 데이터로 전환하는 처리(Text-to Speech)



응용 분야

- 음성 합성(speech synthesis)
 - 말소리의 음파를 기계가 자동으로 만들어 내는 기술



응용 분야

- **지능 로봇(Intelligent robots)**

- 로보틱스(robotics)
- 로봇에 관련된 기술 분야로써 기계 공학, 센서 공학, 마이크로 일렉트로닉스, 인공지능 기술 등을 종합적으로 활용.



인공지능의 한계점



일반적인 판다 사진(왼쪽)에 노이즈를 삽입하게 되면(오른쪽), 사람은 그대로 판다로 인식하지만, AI는 긴팔원숭이로 잘못 인식하기도 한다.



최근 동향

아무도 풀지못한 보이니치 필사본 해석한 인공지능

Aedi_ | 2018. 2. 1 00:29



보이니치 필사본 해석에 도전하는 인공지능

며칠전 우리나라 '승정원일기' 번역작업을 AI가 도전 앞으로 해석에 45년이 필요한 작업을 단축시키겠다는 기사가 보도됐습니다. 승정원일기는 조선시대 왕명의 출납을 관리하던 승정원에서 기록한 것으로 조선 최고의 기밀기록으로 알려져있습니다. 총 3,243권으로 이뤄진 이 기록물은 2억 4,000만자가 적혀있다고 하죠. 전문가들은 AI가 승정원일기를 모두 번역하는데 약 18년이 소요될 것으로 전망했습니다.

최근 동향

IT/정보

닭 울음소리 의미 분석해낸 인공지능 기술

Aedi_ | 2018. 2. 5. 00:03



과연 동물들은 의사소통을 할 수 있는 언어를 가지고 있을까요? 이 궁금증은 인류가 지금까지 궁금해 하기만 했을뿐 정확한 답을 찾을 수 없었던 분야입니다. 하지만 최근 인공지능 기술이 급속도로 발전하면서 동물의 언어를 이해하려는 시도가 하나둘 나타나기 시작하고 있습니다. 스웨덴의 한 연구단체는 4년 안에 돌고래의 언어를 완벽하게 해석할 계획을 밝히기도 했습니다. 돌고래야 유치원생 수준의 지능을 가진것으로 알려졌기 때문에 돌고래어(?)를 가지고 있을 확률이 높지만 다른 동물들은 어떨까요?

최근 동향

구글, 인공지능으로 항공기 지연 예측한다

작성자 AIT

작성일 2018-02-01 18:10

조회 851



항공편을 이용한 여행 중 가장 큰 스트레스는 항공기 지연 소식이다. 이유를 알 수 없는 일방적인 통보에 기분을 망치기 쉽다. 구글은 인공지능(AI)을 이용해 항공기 지연을 예측해서 알려주는 서비스를 시작한다.

최근 동향

심전도 앱과 부정맥 알림 기능, Apple Watch에 탑재



인공지능의 윤리

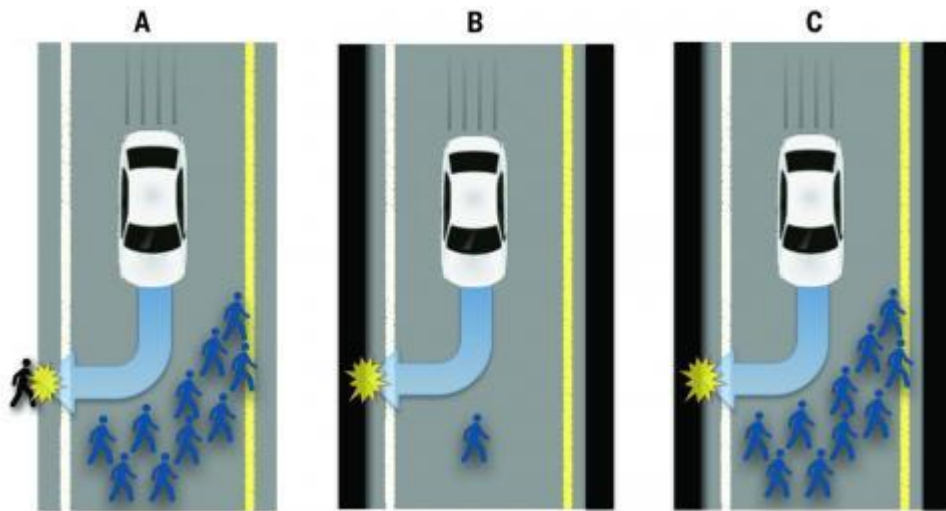
- 인공지능의 윤리
 - 마음이 없는 인공지능
- 살상용 자율무기(LAWS : Lethal Autonomous Weapon Systems)
 - 인간의 개입 없이 스스로 표적을 찾아내고 제거하는 무기
 - 잘못된 판단의 문제



<https://theglobal.blog/2019/05/28/lethal-autonomous-weapons-war-crimes-and-the-convention-on-conventional-weapons/>

인공지능의 윤리

- 인공지능의 윤리
 - 자율주행 자동차의 돌발 생황에 대한 프로그래밍
 - 모든 가능한 상황에 대한 고려 필요
 - 돌발상황에서 희생자를 선택하는 프로그래밍



2018.03 우버 사고



인공지능의 윤리

