

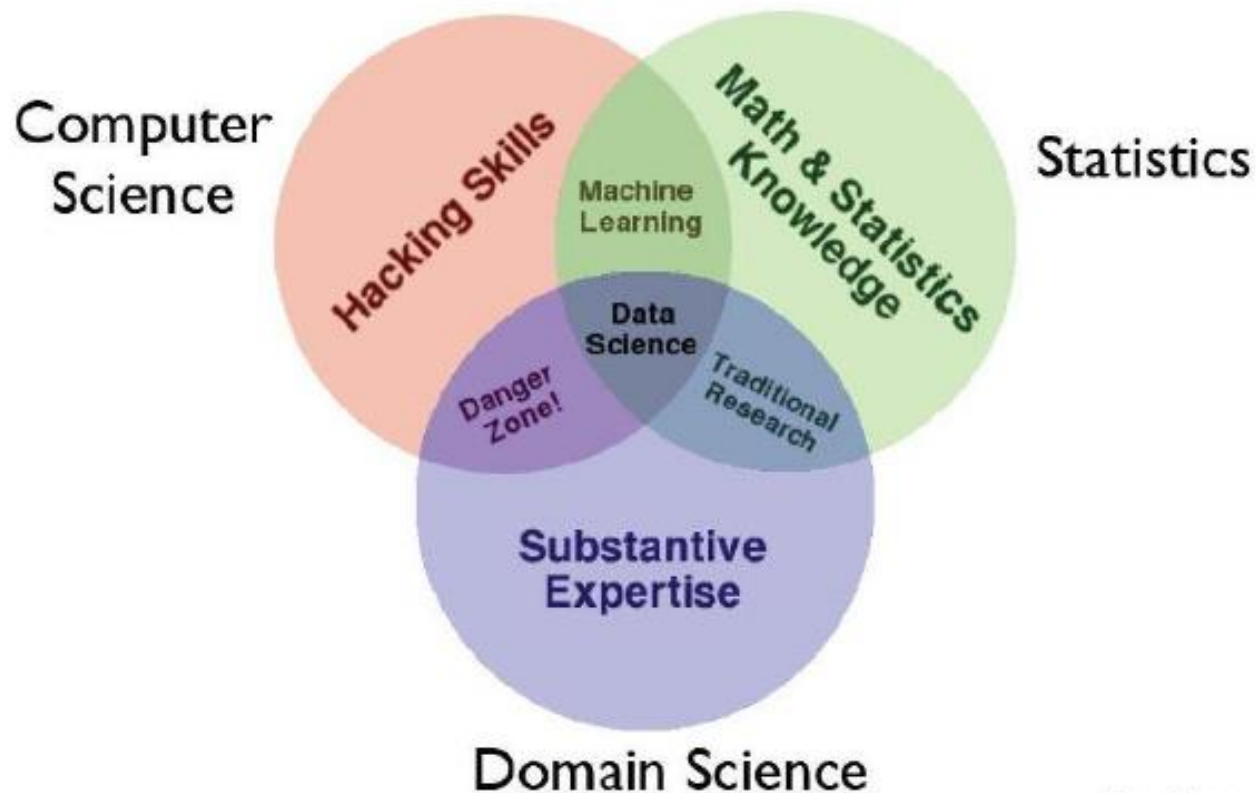
데이터 사이언스(Data Science) 개요

데이터 중심산업

앞으로(현재) 데이터 기술(data technology)의 중요성이 새롭게 부각

- 데이터 사이언스는 정보통신기술(ICT)을 기반으로 하는 최신 기술
- ICT 기술은 전자공학, 컴퓨터공학, 통신공학을 기본으로 발전
- (기존) 음악, 영화, 블로그, 홈쇼핑은 음원, 비디오, 텍스트, 상품 거래, SNS에 관한 데이터를 효과적으로 안전하게 다루는 것이 목적
- (앞으로) 데이터 사이언스에서는 데이터를 효과적으로 다루고 분석하는 기술을 포함, 데이터를 수집하고 변환하고 알고리즘을 적용하는 분석 과정 포함
- 이제 모든 산업이 데이터 기반으로 동작

What is Data Science? – One definition



Drew Conway

“우리는 지금까지 우리가 살아오고 일하던 방식을
송두리째 바꿀 기술 혁명 직전에 와 있다.
제 4차 산업혁명은 그 속도와 파급효과 측면에서
이전의 혁명과 비교도 안 될 정도로
빠르고 광범위하게 일어날 것이다”

— 클라우드 슈밥, 세계경제포럼(WEF · 다보스 포럼) 의장 —



Copyright©2017, Fabmonster co., Ltd. · LEE Gyeongjin All rights reserved.

— 출처, 4차 산업혁명 위기인가 기회인가 뉴시스 2017. 02. 07. —

데이터 사이언스

4차 산업혁명

4차 산업혁명이란 무엇인가?

파괴적 기술과
역사적 산업혁명의 전개



4차 산업혁명



1차 산업혁명

증기기관
18세기



2차 산업혁명

전기·내연기관
19-20세기



3차 산업혁명

컴퓨터·인터넷
20세기 후반

IoT

로봇



AI

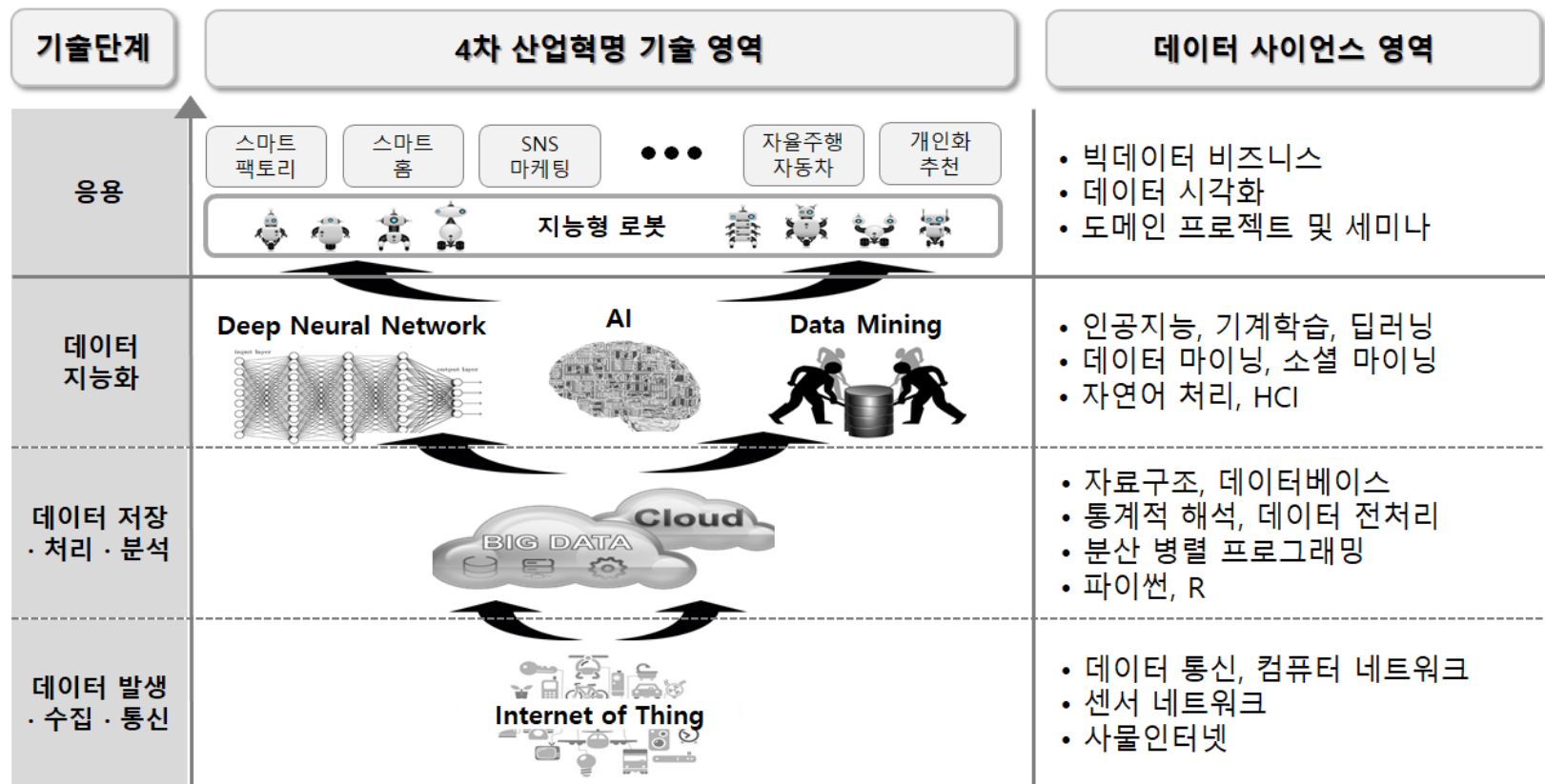
3D
프린팅

AR/VR

AI 기술을 핵심동인으로
상품·서비스의 생산·유통·소비 전 과정에서
모든 것이 연결되고 지능화

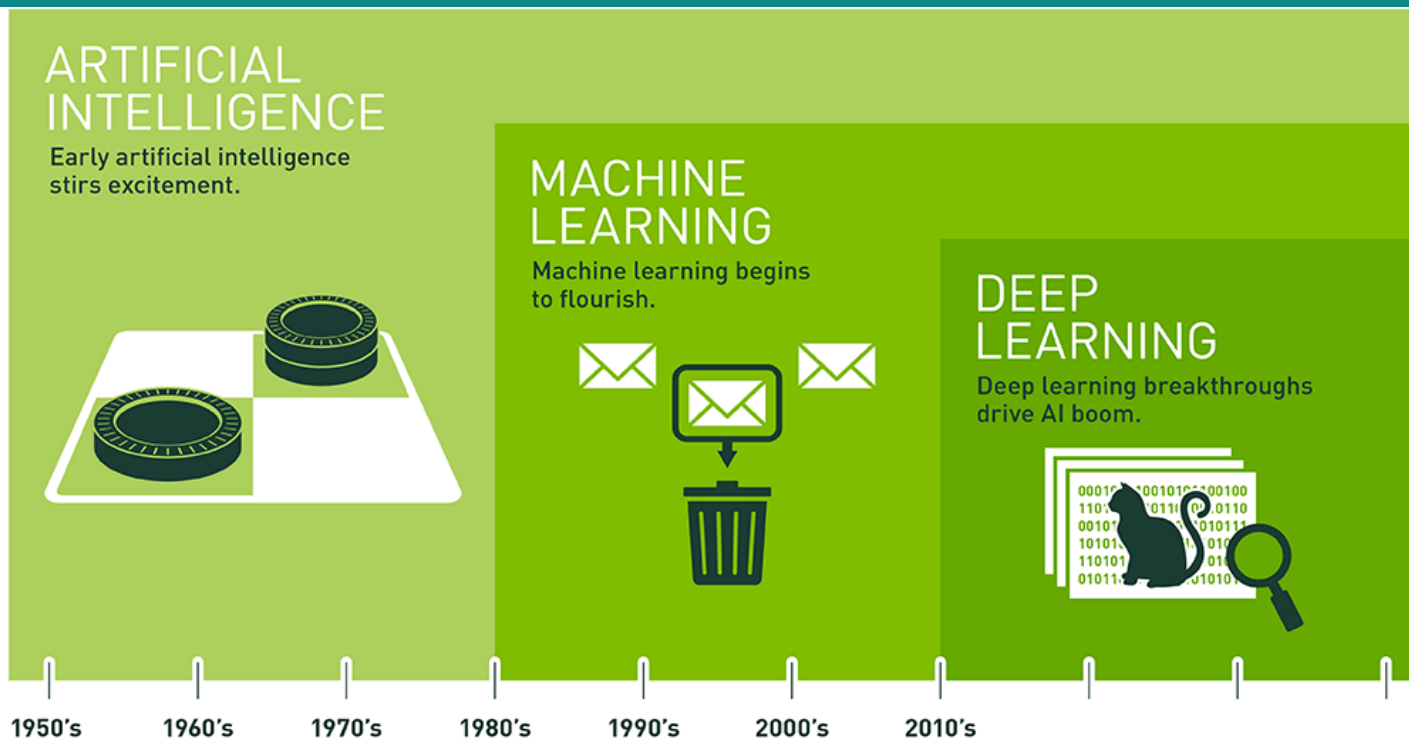
4차 산업혁명 시대 – 데이터 사이언스 영역

- ◆ 4차 산업혁명을 주도하는 주요 영역은 인공지능, 로봇, 사물인터넷, **IoT, Mobile** 등
- ◆ 주요 영역 기반 기술은 데이터를 활용(데이터 사이언스)하는 정보통신기술(ICT)
- ◆ 최근 데이터로부터 가치를 창출하는 데이터 기술 (**Data Technology**) 급 성장 및 확장



기술단계별 4차 산업혁명 Vs. 데이터 사이언스

인공지능 분야는 컴퓨팅, 대용량 스토리지, 인터넷과 같은 핵심기술 발전으로 크게 성장



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

1. 데이터 사이언스

AI의 확대

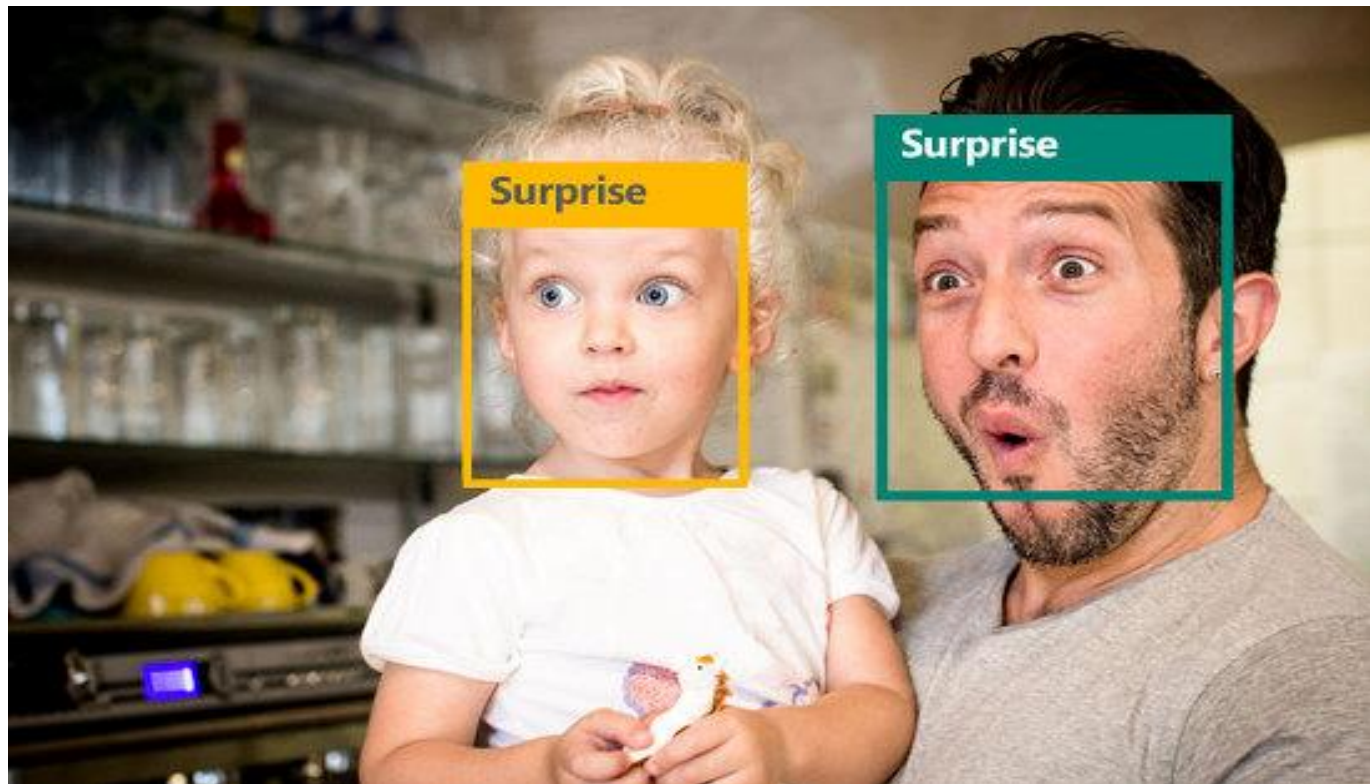
인공지능은 인간의 지능을 기계로 구현한 것



1. 데이터 사이언스

AI의 확대

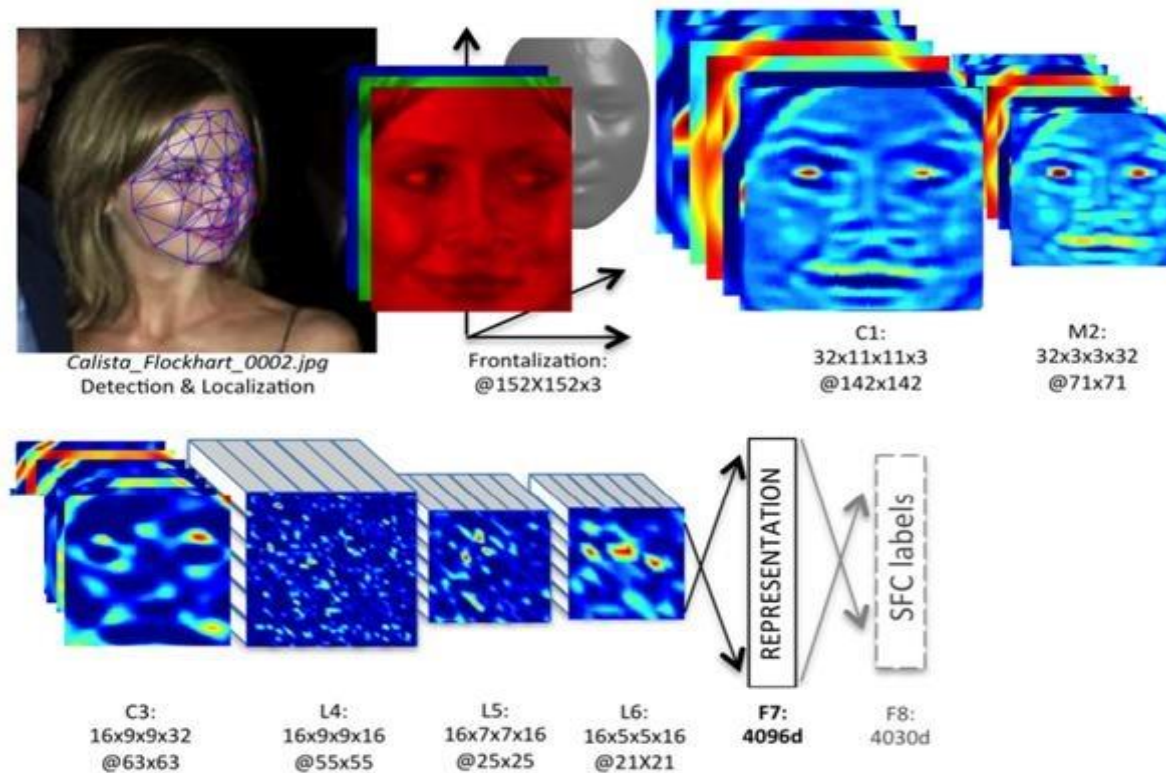
딥 러닝은 완전한 머신러닝을 실현하는 기술



1. 데이터 사이언스

AI의 확대

딥러닝을 이용한 이미지 인식 프로그램은 이미 실생활에 접목되어 사용



1. 데이터사이언스

AI의 미래

다가올 미래에 대비해야 내 일자리가 생긴다.

- 우리는 AI의 미래를 잘 예측하고 준비해서 불필요한 경쟁을 피해야
- 사람 VS 기계, 엔진
- AI로 인해 직업 변화가 크게 나타날 것
- 자율주행차는 아직 불안하다? 배달용 무인 자동차 운영 및 도입



2. 데이터 사이언스 도구

파이썬(Python)



2. 데이터 사이언스 도구

파이썬(Python)

- 일반 programming 언어이면서 자바, C에 비해 문법이 간단, 배우기 쉽고 강력한 기능
- 초보자 교육에서부터 웹 개발, 과학 계산, 데이터 분석, 인공지능 등 다양한 분야에 활용
- 데이터 사이언스에서 사용할 수 있는 도구는 자바, C와 같은 일반 프로그래밍 언어와 R과 같은 통계 분석에 특화된 언어, 그리고 Matlab, SPSS, SAS와 같은 데이터 분석 도구
- R과 유사하게 데이터를 다루는데 편리한 기능을 제공한다. 최근에는 파이썬(python) 언어가 데이터 분석에 가장 많이 사용
- 딥러닝을 이용한 프로그램 작성에서 파이썬을 기본 언어로 채택함으로써 이제 머신러닝 영역뿐만 아니라 일반 데이터 분석에서 폭넓게 파이썬을 도입

2. 데이터 사이언스 도구



파이썬(Python)이란?

- 오픈 소스
- 크로스 플랫폼(Windows, Mac OS X, Linux 등)
- Interpreter형 컴파일 방식을 채택하여 동적 변환이 쉽고, 여러 분야의 어플리케이션 개발과 스크립팅에 이상적인 언어
- 대화형 언어
- 파이썬은 기능 위주의 프로그래밍, 절차지향적 프로그래밍, 객체지향적 프로그래밍 등을 채택하여 다중 프로그래밍 패턴을 지원
- 방대한 사용자와 활발한 커뮤니티 : 파이썬은 풍부한 라이브러리를 보유, 파이썬 웹사이트에서 무료로 소스를 얻을 수 있고 자유로운 배포가 가능한 장점이 있음
- 구글, 유튜브, 인스타그램, 드롭박스, NASA 등에서 사용

2. 데이터 사이언스 도구

○ 파이썬(Python)으로 할 수 있는 것

- 웹 개발
- 게임 개발
- 업무 자동화
- 데이터 분석
- 기계 학습
- 과학 계산
- 텍스트 처리
- 이미지 처리
- 멀티 미디어
- 데이터 베이스
- 기타

2. 데이터 사이언스 도구

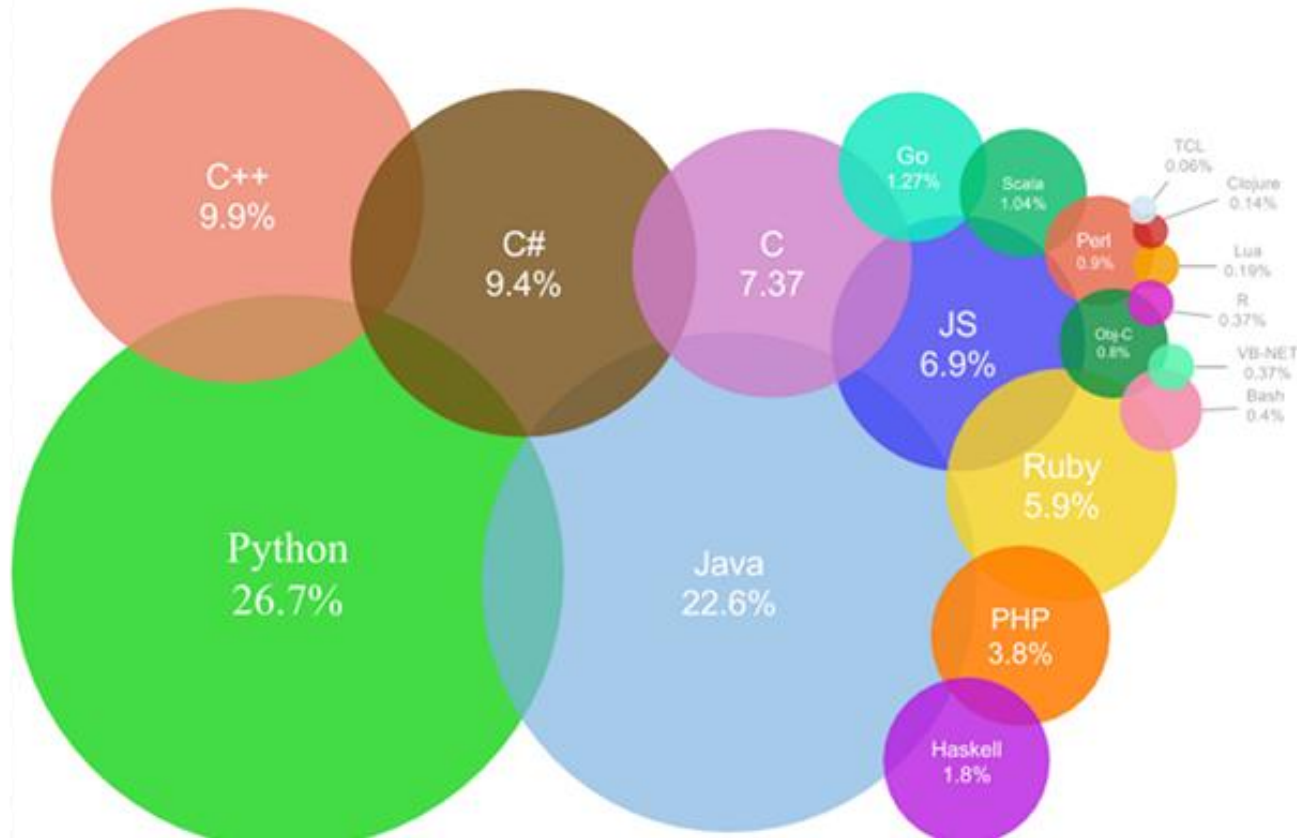
파이썬(Python)의 단점

- 느린 속도
- 모바일 앱 개발에서 사용하기 어렵다.

2. 데이터 사이언스 도구

파이썬(Python)이란?

Most Popular Coding Languages of 2016



4차 산업혁명 시대 – 미래 혁신 성장 기술 중심



Hyper Cycle 2017 – Gartner

AI 관련 업종별 권장 과목

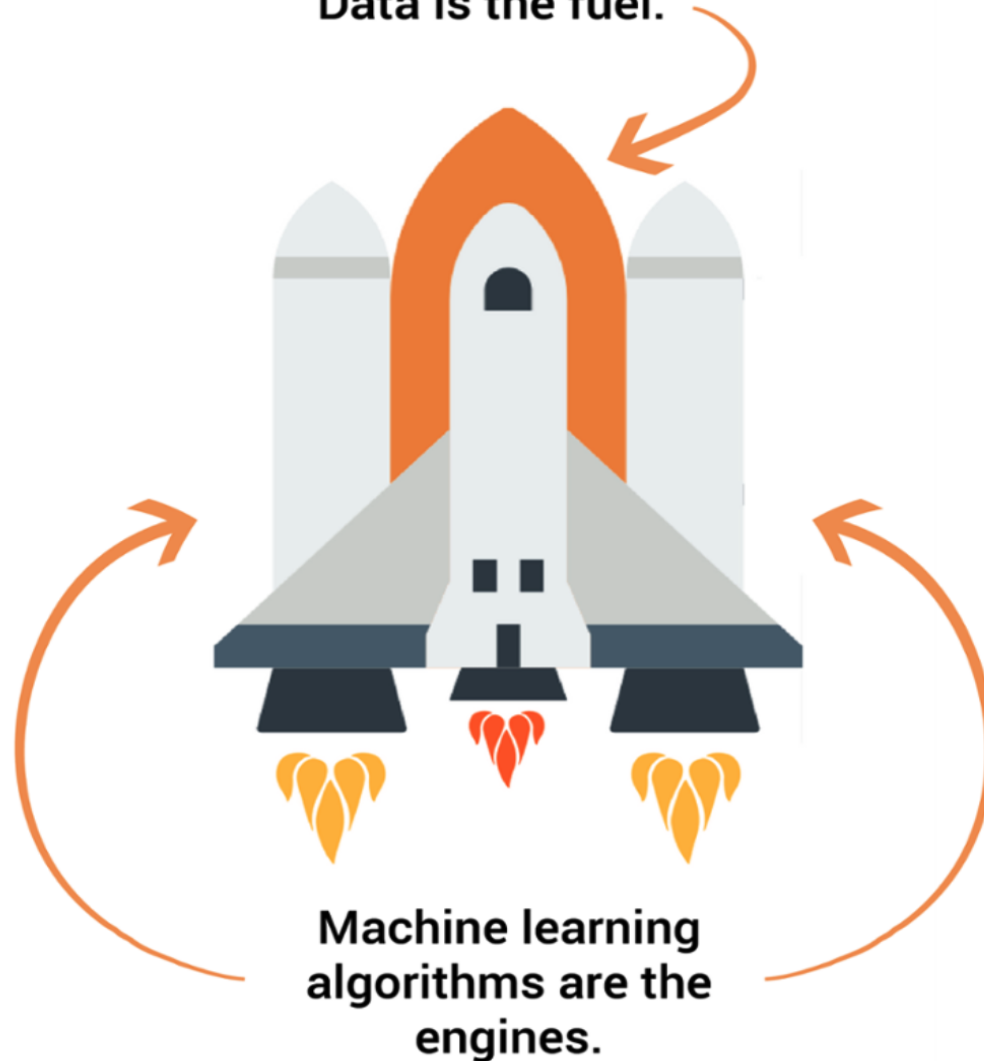
– 관련 업종(미래 일자리) 설문 조사

단계	과목(모듈)	업종				비고
		금융	제조	마케팅	물류/유통	
기초	통계 및 프로그래밍	○	○	○	○	업종 공통 영역 (분석 역량 중심)
	데이터 분석	○	○	○	○	업종 공통 영역 (플랫폼/기술 역량 중심)
코어*	머신러닝	○	○	○	○	업종 공통 영역 (데이터 사이언스 역량 필수 영역)
	딥러닝	○	○	○	△	업종 선별 영역
응용	이미지 분석		○	△	△	업종 선별 영역
	자연어 처리	○		○	△	업종 선별 영역
실전	도메인 프로젝트	△	△	△	△	개인별 선택 영역

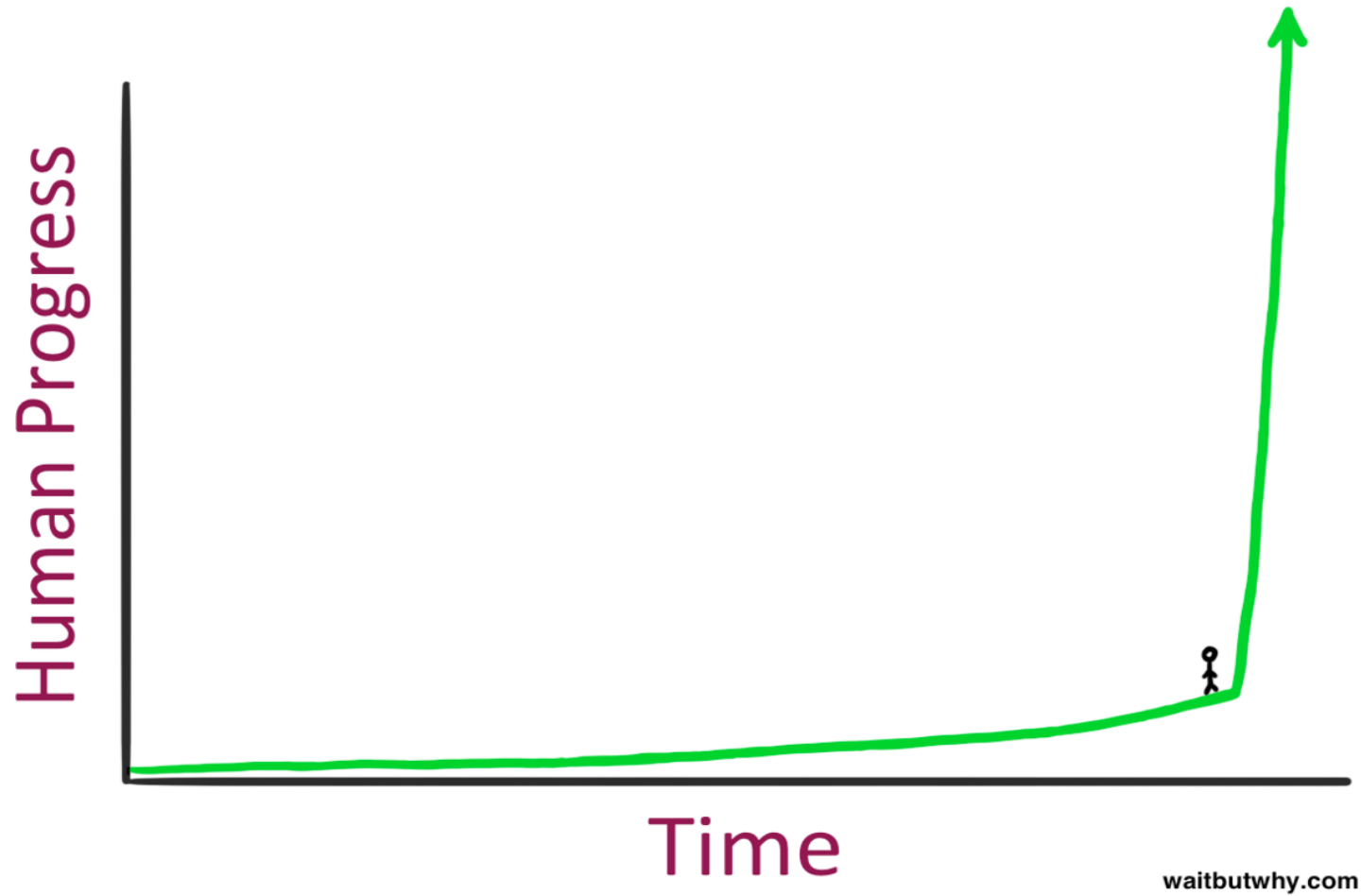
빅데이터와 AI 트렌드

빅데이터와 머신러닝

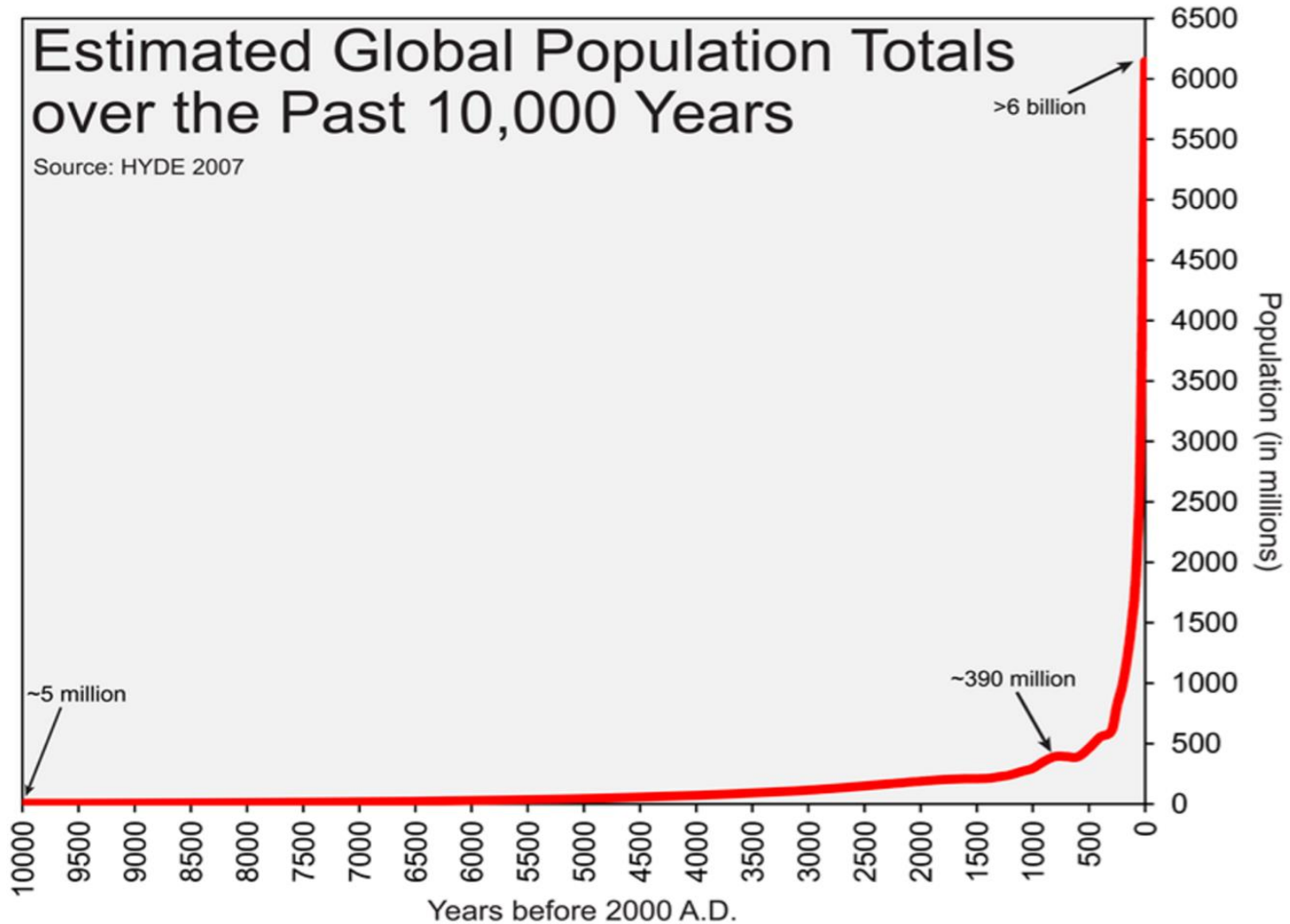
Data is the fuel.



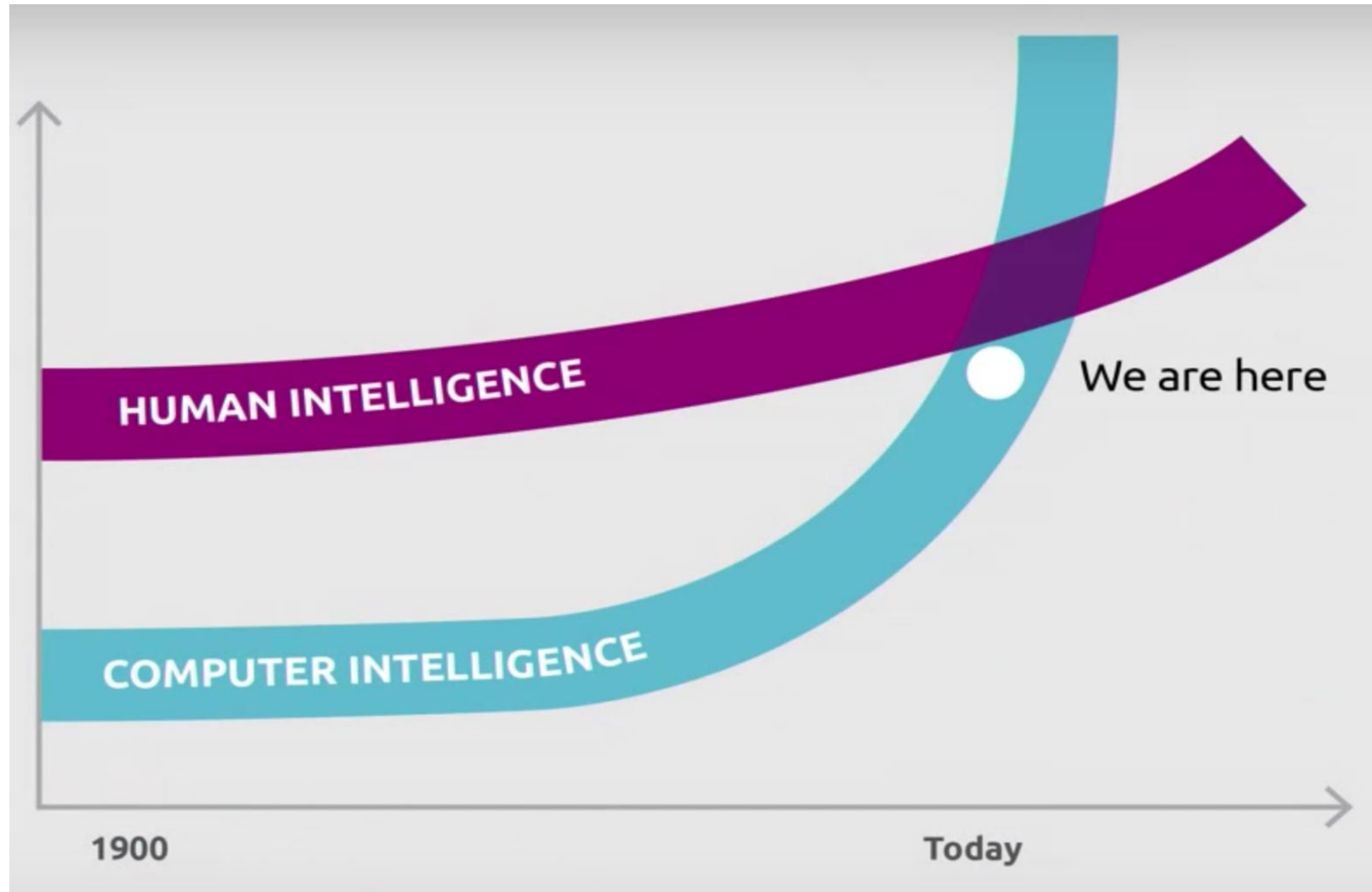
AI 임팩트



변화 사례



컴퓨터 지능



4차 산업혁명

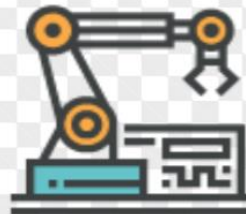
- 1, 2차 산업혁명
 - 에너지 생산과 에너지 전달의 혁명
 - 인간 **근육 노동력**의 한계
- 3, 4차 산업혁명
 - 정보(IT) 기술 혁명
 - 인간 **지적 노동력**의 한계



1st Industrial Revolution
WATER & STEAM



2nd Industrial Revolution
ELECTRICITY

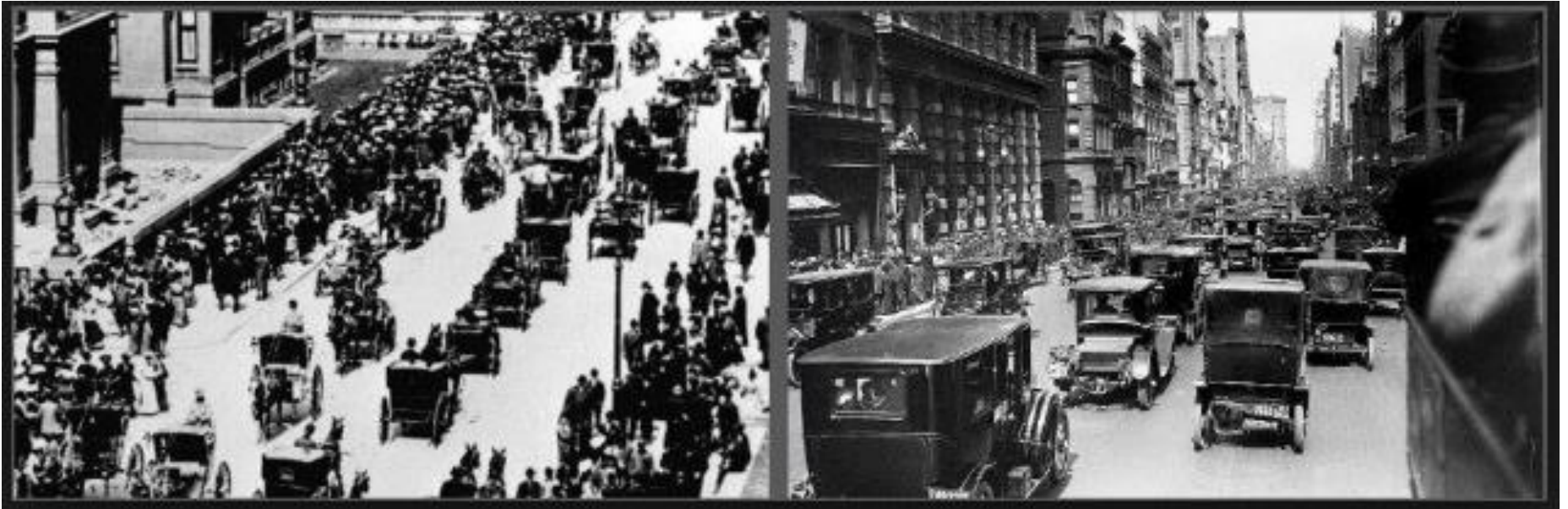


3rd Industrial Revolution
AUTOMATION



4th Industrial Revolution
CYBER-PHYSICAL SYSTEMS

1900년초 뉴욕



AI 영역 확대 -> 최종 모델(1 2)

- 대량의 숫자 데이터 분석 – 빅데이터 분석
- 이미지 인식 – 보는 능력
- 음성인식, 텍스트 인식 – 듣는 능력
- 실시간 번역 – 말하는 능력
- 이미지 캡션, 언어 모델링 – 쓰는 능력
- 감성 능력까지 ?



빅데이터(AI) 도입 사례

자동차 보험사

- 보험사는 교통사고를 낼 확률이 높은 사람을 찾아 이들의 보험료를 올리고 이들이 다른 회사로 가게 한다



응급환자 예측

- 응급환자를 미리 찾아내어 구급차 이용을 줄인다



마케팅

- 개인별 맞춤형으로 마케팅을 하기 위해서 고객의 과거 상품구매 이력, SNS 분석
 - 고객이 매장을 둘러보는 도중에 고객의 성향과 현재의 욕구를 파악하여 실시간 추천
- 새로운 영화가 소개되면 얼마나 투자를 해야할지 또는 몇 개의 개봉관에서 이 영화를 상영하는 것이 적절할지를 예측
 - 영화 시나리오 대본, 주인공 분석, 예고편에 대한 고객 반응을 분석



콜센터

- 고객 성향에 맞게 잘 대응할 적합한 직원을 배정
- 고객의 과거 이용 기록, 목소리 감성분석
 - 까다로운 고객인지, 또는 지금 무슨 용건으로 전화를 했는지를 예측
 - 예를 들어 평소에 카드 납입액을 자주 물어본 고객이라면 기다리는 동안 미리 카드 납입액을 자동으로 알려준다.
 - 고객이 평소에 자주 궁금해 하는 정보를 바로 자동응답기로 안내

콜센터

- 응답 최적 루트
- 콜 요청량 예측
- 다음 행동(질의) 추정
- 실시간 고객 감성 분석
- 실시간 응대 제안 (상급 응대자 판정)
- 스크립트 최적화
- Script optimization
- 음성 인식 (상습 불법 신고자 파악)

인재 관리

- 인재 필요 역량 예측
- 작업량 예측
- 인재 유출 예측 (인재의 가치 평가, Talent Analytics)

마케팅

- 고객 세분화 (행동 세분화, 구매 패턴 분석)
- 고객 재방문 예측
- 고객 이탈 예측
- 개인 맞춤형 광고
- 크로스셀링, 업셀링 아이템 추천
- 소셜 데이터 분석
- 고객 불만 분석
- 매출 예측
- 수요 예측

금융

- 개인이나 기업에 대출을 위한 신용 평가
- 과거의 대출 사고 사례, 개인의 금융 활동을 분석
- 도난된 신용 카드 사용을 찾아내기 위해서
 - 평소 정상적인 거래를 분석하고 이상 현상을 발견
- 보험사는 사기성 보험 청구를 찾거나 불법자금 세탁을 찾아내는데 사용
- 자동차 보험회사에서는 차량에 블랙박스를 달면 보험료를 감해 준다.
 - 블랙박스 데이터를 보험사로 보내주면, 예를 들어 급정거를 하는지, 급커브를 하는지 등의 정보를 제공하면 보험료를 더 감해준다.
 - 고객의 안전한 운전을 유도하며 동시에 사고를 줄일 방법을 찾는다.

공장 자동화

- 생산 라인의 최적화, 상품의 물류와 재고 관리의 정보화에 추가로
- 이 과정에서 발생하는 데이터를 상세히 분석함으로써 더 효과적인 생산, 유통, 재고관리
 - 제품 자체의 품질 향상에도 이용하고 있다.

에너지 이슈

- 에너지의 역할
 - 에너지는 경제, 안전, 산업의 기반
 - 기후변화, 재해, 가난에 대응
 - clean, affordable, reliable 에너지 필요
- AI가 에너지 생산, 전송, 분배, 사용 전분야의 지능화 수단으로 도입
 - 에너지 생산 방법의 다양화
 - 에너지 사용의 효율화

에너지 이슈

- 전력 수요 예측
- 신재생 에너지의 불규칙한 공급 예측
- 도전 감시
- 정전 예측 (시점 및 기간)
- 전기자동차의 충전 시간, 장소, 용량 예측
- AMI 분석을 통한 에너지 이용 패턴 분석
- 에너지 다변화 대응 (Peaker plant 운영)

Peaker Plant

- 피크 타임은 시기, 장소에 따라 다르다
 - 일반적으로 저녁시간(가전제품 사용),
 - 더운날은 오후에(에어컨),
 - 추운날은 아침에(히터),
- Peaker로는 개스터빈, 화력발전이 주로 사용된다.
 - 기반 전력생산비보다 높다.

Smart Grid

- 양방향 통신, 제어, 지능형 컴퓨팅으로 구현
- 주요 구성 장치
 - Phasor Measurement Units (PMUs) – 그리드 안정성 센서
 - Digital Meters – 정밀한 전력사용 측정
 - Relays – 장애 우회 릴레이
 - Automated feeder switches
 - Battery – 에너지 저장 장치
- [애니메이션](https://goo.gl/c6zEq9)
 - <https://goo.gl/c6zEq9>

AI 스마트 그리드

- 에너지 센서 데이터를 분석하여 에너지 자원의 적절한 배분
- 딥러닝 기술 도입
 - 기존의 모델 중심의 머신러닝에서, 빅데이터 중심의 머신러닝으로 진화
 - 정교한 예측 및 관리
 - 소지역 중심의 관리, 축전 장치의 활용
 - 에너지 절감
 - 구글 데이터
 - 센터의 에너지 절약



에너지 이슈

- “돌이 다 떨어져서 석기시대가 끝난 것이 아니다 (yamani)”

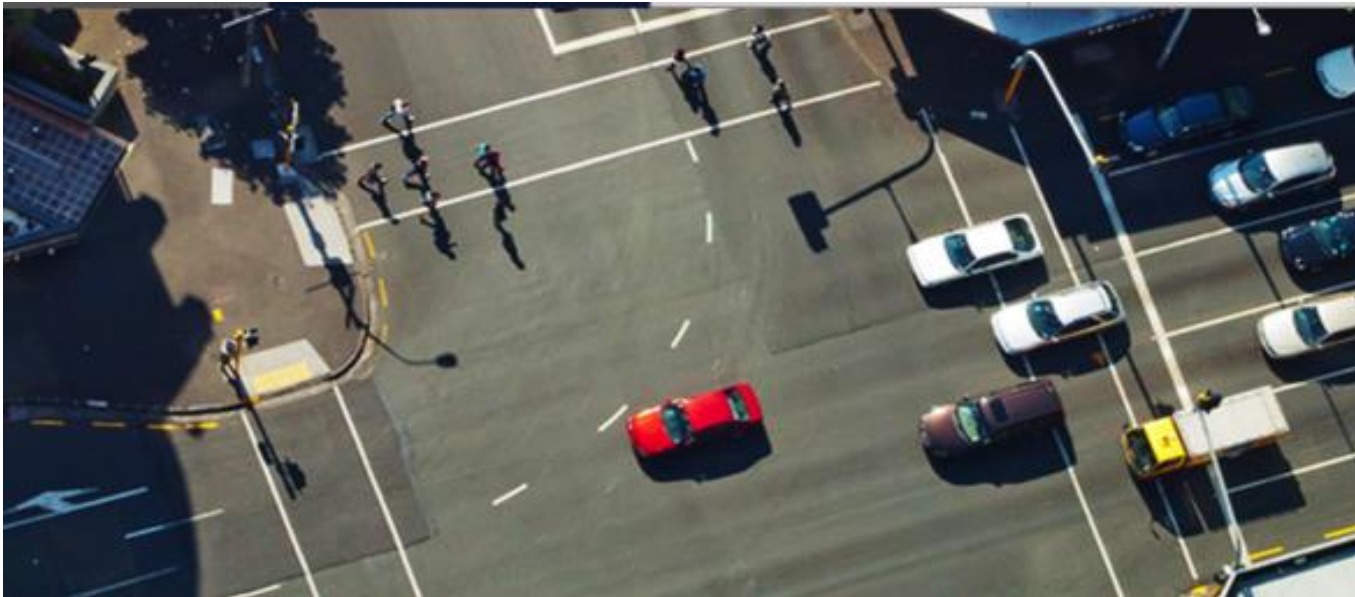


자율주행차

- 자율주행차의 상용화 대비([PathProject](#))
- 부정적인 면
 - 새로운 물체나 상황 인식이 어렵다
 - 공사현장, 수신호 등을 파악하기 어렵다
 - 돌발상황에 실시간으로 대응
- 긍정적인 면
 - 버스, 트럭부터 운행한다 (길을 잘 아는 상황)
 - 360도 관찰이 가능하다
 - 기술 발전이 빠르다
 - Path Project

적응형 신호등

- 대기시간 10~50% 단축



Self-Driving Car Timeline (18/7)

Ford – True Self-Driving by 2021

Volvo – Self-Driving on the Highway by 2021

BMW – Fully self-driving possible by 2021

Honda – Self-Driving on the Highway by 2020

Toyota – Self-Driving on the Highway by 2020

Hyundai – Highway 2020, Urban Driving 2030

건강

- 혈압, 심장박동수, 수면 패턴, 운동 패턴과 같은 개인 데이터를 종합적으로 분석
- 개인의 DNA 염기서열, 유전자 발현 분석을 통해서 개인별 질병 예측과 치료방법 제시도 상당히 정확해질 것
- 개인간의 약 복용에 대한 효과 차이를 구분할 수 있다면 맞춤형 처방 또는 맞춤형 신약개발이 가능

심혈관질환 예측

- 현재 가이드라인
 - 나이, 콜레스테롤 수치, 혈압, 흡연, 당뇨 등에 기반
- 영국 노팅햄대학교 연구
 - 37만명의 EMR 분석
- 실제로는
 - 인종차이, 정신질환, 경구용 스테로이드 복용 등이 큰 영향

Sedasys

- 존슨앤존슨의 자동 수면 마취기
 - 마취약 자동 주사
 - 심박수, 산소포화도, 심전도, 혈압을 보고
 - 수면 내시경 의료비를 1/10로 낮춤
 - 더 빠르게 회복,
 - 저산소증문제도 적게 겪음
- 2016년 시장에서 철수



SimSensei

- 심리 상담:
 - 인간 의사보다 AI의사에게 더 솔직하게 상담
- 서던 캘리포니아대에서 개발
 - 가상의 여성 상담사
 - 환자의 시선, 미소, 머리 움직임, 표정 분석



정신과 상담

- 목소리로 감성 분석 (Beyond Verbal)
- 앱 Moodies
 - 언어적 요소 linguistics 뿐 아니라 비언어적 요소 acoudtics도 분석
- 조현병, 자살위험군 예측
- 우울증: 6초간 목소리 분석으로 AUC 0.93 (PureTech Health)
- 조울증 90% 진단

AI 우선 도입 분야 예 (의료)

- 판독에 많은 시간과 노력이 드는 분야
- 문제의 난이도와 위험도가 적은 분야
- 의사들이 하기 싫어 하는 분야
- 보험 수가가 낮은 분야
- 비용감소와 효율이 향상되는 분야

AI는 보편적인 도구가 될 것



“ AI 와 인간이 경쟁하는 것이 아니라 AI 를 사용하는 인간들 사이의 경쟁이다.”

AI 도입 전략

AI의 정의

- General (일반적인) AI – 종합적인 처리능력을 가진 범용 AI
- Specual (특수한) AI – 특정한 업무를 잘 수행하는 AI
 - 직업과 업무, 비즈니스의 변화

왜 AI인가?

- 의사결정을 돕는다
 - 프로세스의 자동화
 - 개인화된 제안(추천)이 가능
- AI가 실용적이 된 배경
 - 하드웨어, 소프트웨어, 알고리즘, 빅데이터
 - IOT, 딥러닝, 5G 통신과 같은 플랫폼
- AI-주도 비즈니스
 - 구글, 페이스북, 네이버

AI의 본격적 실용화

- 2012 ImageNet
 - 컴퓨터 비전 경진대회 (보는 능력)
 - 토론토대학 팀의 성능이 급격히 향상
- 2017년 이미지넷 인식률 97.3%
 - AI은 이제 시작일 뿐

AI의 미래

- AI가 잘 하는 영역
 - 기억 memorization,
 - 처리 transactions,
 - 예측 prediction
- 사람이 잘 하는 영역
 - 감성적 능력 emotive capability,
 - 소통 능력 communication,
 - 즉흥성 improvisation,
 - 일반화 generalization
- 협력 지능(collaborative intelligence) – Augmented Intelligence

우선순위 배정

- 작은 성과를 낼 수 있도록 시작한다
 - 한번에 성공하기 어렵다
 - 전체의 동작을 파악하고 새로운 문제를 찾을 수 있도록 한다
- 초기 전략회의에 엔지니어가 참석해야
 - 어떤 문제가 발생할지 미리 예측해야
 - 같은 기술이 모든 문제에 적용되지 않는다
- 성과측정 지표를 정해야
 - 머신러닝을 하는 최종 목적을 달성할 수 있는지가 중요하다

파일럿 프로젝트

- 전체를 파악
 - 처음에는 가능한 간단한 모델을 목표로 만들어야 한다
 - 성과를 바로 얻는 것 보다 경험을 쌓는 것이 목적
- 최신의 기술을 적용해야만 하는 것이 아니다
 - 원하는 성과 목표만 달성하면 된다.
 - 엔지니어는 최신의 기술 적용을 추구하는 경향이 있다.

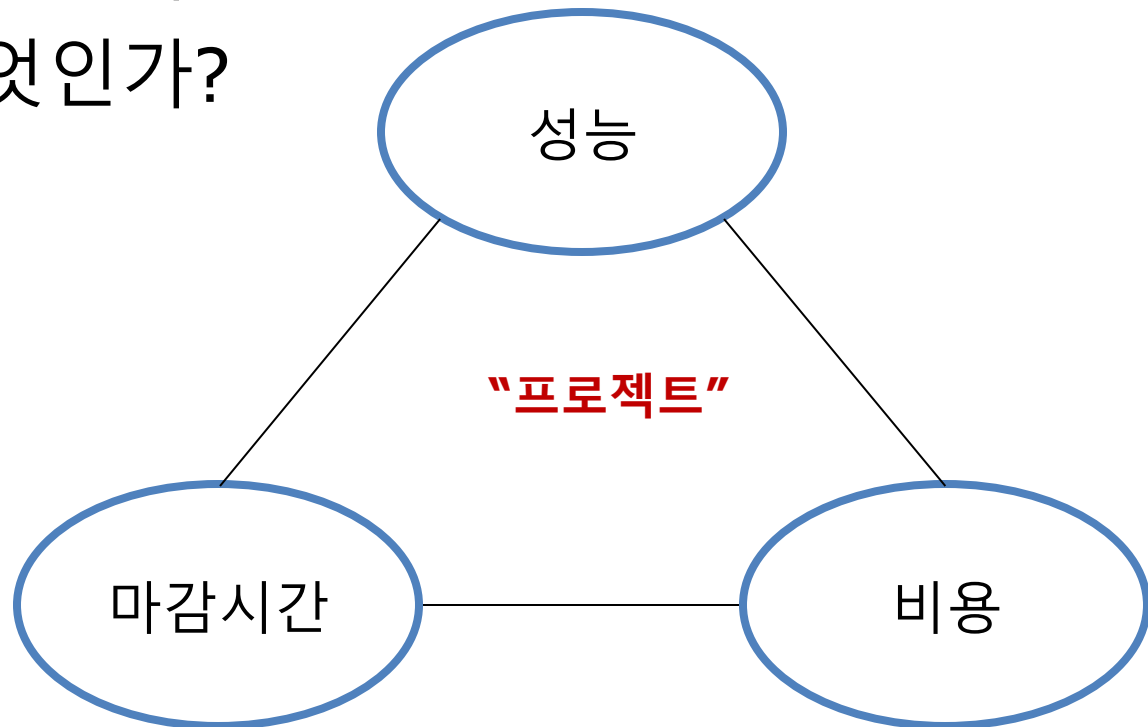
적용 전략

- 사람들은 인공지능의 완벽한 동작을 기대한다
 - 투자자(지원자)는 특히 기대가 크다
- 그러나 적용전에 품질제어와 테스트가 필요하다
 - 여러 단계의 테스트가 필요

AI 프로젝트 전략

머신러닝 '프로젝트' 고려사항

- 최대 얼마까지 비용을 투자할 수 있는가?
- 언제가 마감 시간인가?
- 성능 기준은 무엇인가?



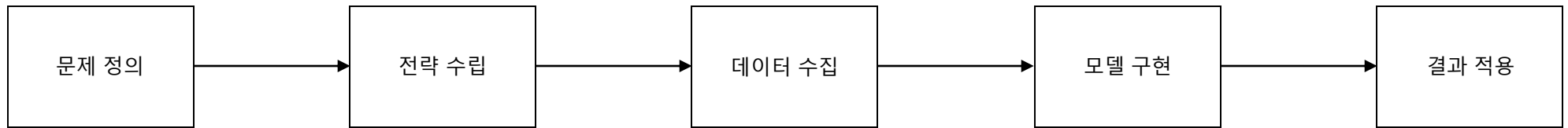
프로젝트 검토 사항

- 근본 문제 검토(고객의 근본적 고민이 무엇인지)
- 현재까지 왜 해결을 못했는지
- 목적을 분명히 파악하고 구체적(정량적) 목표 수립
- 프로젝트 종료 조건 명시 (목표 달성 여부, 시간 등)

가용 자원의 파악

- 사용할 수 있는 데이터의 종류는?
- 인적 자원은? (내부 인력 또는 외부 전문가 활용)
- 컴퓨팅 자원은? (클라우드 사용)
- 데이터 수집 비용은?(무료, 유료)
- 관계자와의 협조 및 갈등 요소는?
- 보안 대책은?

AI 프로젝트 프로세스



- 문제 정의 - 해결하려는 문제를 명확히 정의하는 것
- 전략 수립 - 문제 해결을 위해 어떤 데이터를 어떻게 사용할지
를 정함
- 데이터 수집 - 머신러닝에 필요한 데이터를 수집하는 것
- 모델 구현 - 분류, 회귀, 설명, 추천 등을 위한 머신러닝 모델을
구현
- 결과 적용 - 머신러닝 모델을 실제 상황에 적용하고 성능을 개
선하는 것

문제정의

- 문제를 명확하게 구체적으로 정의하는 것
- 가장 중요한 단계
- 문제가 아닌 것을 해결하려고 시도하는 경우가 의외로 많음
- 문제는 조건, 조직, 시기, 장소 등에 따라 다름

문제정의 방법

- 큰 문제를 한번에 해결하기 어려움
- 해결할 수 있는 크기의 작은 문제로 나누어 해결하는 전략
- 예) 큰 문제: 기업의 수익 감소
 - 배송이 늦어지는 이유 분석
 - 고객 불만이 발생하는 원인 분석
 - 반품이 많은 이유 분석

관련 기술



The diagram consists of three concentric circles. The outermost circle is blue and represents Artificial Intelligence. Inside it is a teal circle representing Machine Learning. The innermost circle is orange and represents Deep Learning. The circles are nested, indicating that Deep Learning is a subset of Machine Learning, which is a subset of Artificial Intelligence.

ARTIFICIAL INTELLIGENCE

Programs with the ability to learn and reason like humans

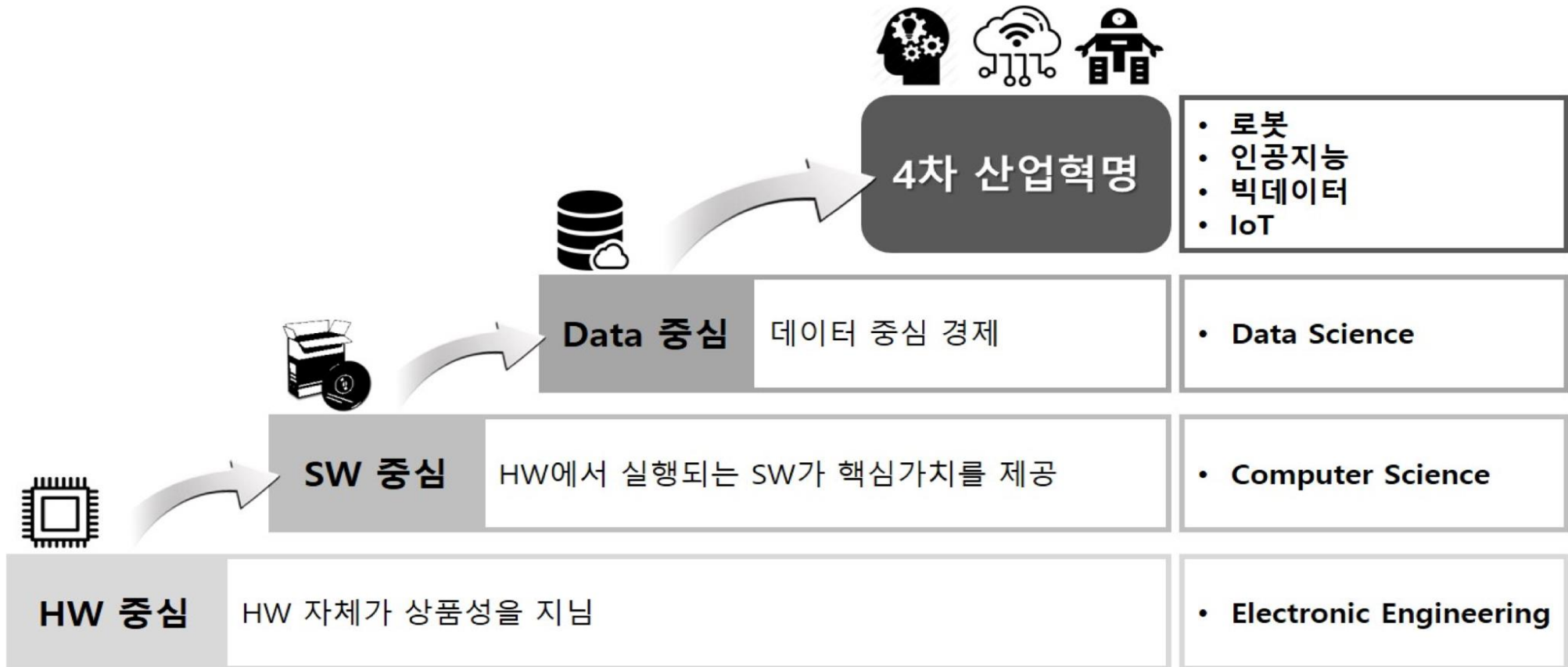
MACHINE LEARNING

Algorithms with the ability to learn without being explicitly programmed

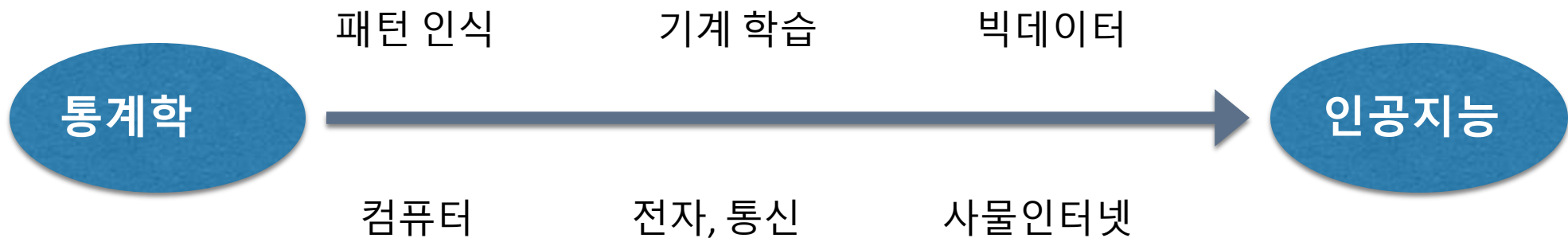
DEEP LEARNING

Subset of machine learning in which artificial neural networks adapt and learn from vast amounts of data

데이터 사이언스



인공지능 구현 기술



인공지능 관련 기술

- 인공지능:
 - 지능이 있는 것처럼 컴퓨터가 **똑똑하게** 일을 처리하는 것
- 데이터마이닝:
 - 데이터베이스에서 **새로운** 지식을 얻는 것
- 비즈니스 인텔리전스:
 - 데이터 분석을 통해 새로운 **비즈니스** 전략을 얻는 것
- 통계분석:
 - **샘플 데이터**로부터 전체 데이터의 속성을 파악하는 것
- 빅데이터 분석:
 - **대량**의 데이터를 분석하여 일반적인 분석으로는 찾지 못하던 **새로운 가치**를 얻는 것
- 머신러닝:
 - 데이터로부터 새로운 지식을 얻는 모델을 만들고 학습으로 **모델의 성능을 개선**
- 딥러닝:
 - 머신러닝의 한 방법으로 신경망을 사용

심슨 패러독스(Simpson's Paradox)

- 각 그룹 데이터에서 나타나는 특징이 그룹들이 결합되었을 때 달라지는 현상

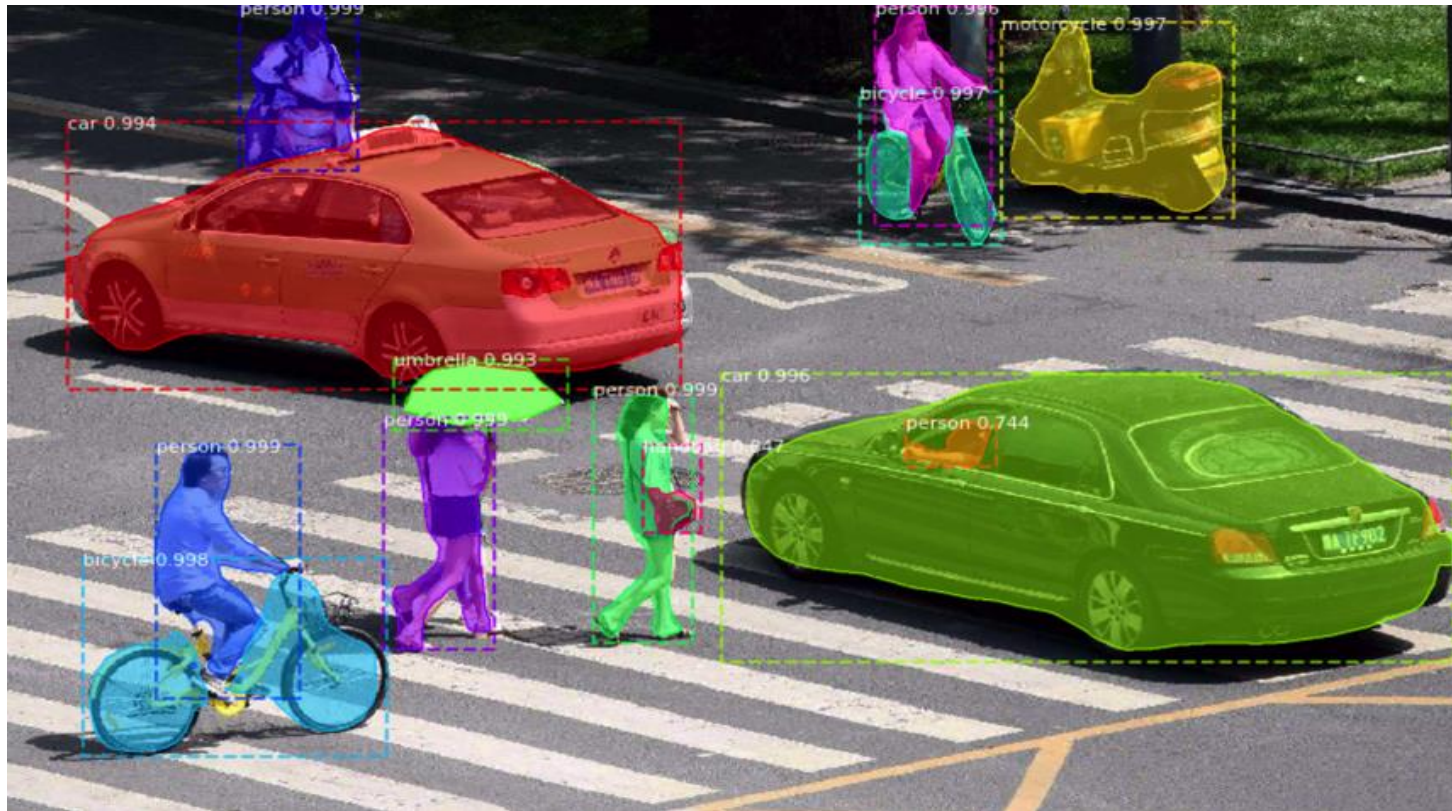
도시	A사	B사
서울	정상품 90 불량품 10 (불량률 10%)	정상품 920 불량품 80 (불량률 8%)
춘천	정상품 980 불량품 20 (불량률 2%)	정상품 99 불량품 1 (불량률 1%)
전체	A사 총 불량률 $30/1,100$ = 3%	B사 총 불량률 $81/1,100$ = 8%

딥러닝 예

- 동영상을 보고 요리법을 스스로 배우는 로봇



딥러닝 예



IOT

- 사물 인터넷 (IOT: Internet of Things)
 - TV, 냉장고, 세탁기, 보안장치, 난방장치
 - 각종 센서를 부착하여 가축의 이동, 소화, 건강상태도 모니터
 - 데이터의 실시간 분석, 다양한 센서 데이터를 융합 처리
- 지금까지 가치 있는 정보는 주로 사람이 만들어냈다.
 - 뉴스, SNS 데이터, 블로그, 통계자료, 음악, 영화 등
- 앞으로는 사람이 직접 만드는 데이터보다 센서와 정보 기기들이 생산하는 데이터가 급격히 늘어날 것
 - 온도, 습도와 같은 과학적 측정 데이터
 - 소음, 카메라, 오염도, 교통상황, 인구밀도 변화, 약물사용 통계 등
- 전혀 새로운 서비스가 나타날 것
 - 기계 장치의 고장을 미리 알아내고 이용자의 패턴을 쉽게 찾아내며, 위험을 조기에 예측하게 될 것이다.

데이터 분석

데이터 수집

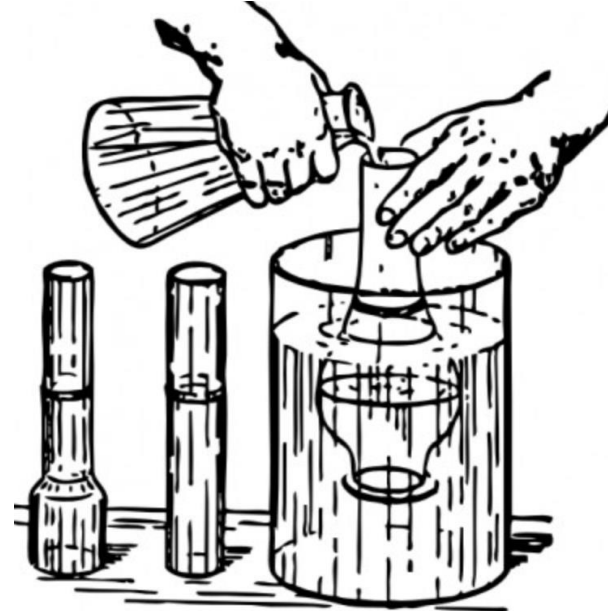
- 전체 과정에서 70~80%의 시간을 소모함
- 핵심 데이터를 확보했는지 여부
- 데이터 품질
- 잘못된 데이터 사용은 잘못된 결과를 도출

데이터 수집 기술

- 수집 가능 여부 (보유 기관의 데이터 정책)
- 수집 주기 (일회성, 한시간/하루/한달에 한번 등)
- 비용 (무료 또는 유상, 통신 비용 등)
- 데이터 포맷 변경, 호환성, 처리 비용
- 정답 데이터 셋 확보 여부
- 수집 아이디어...

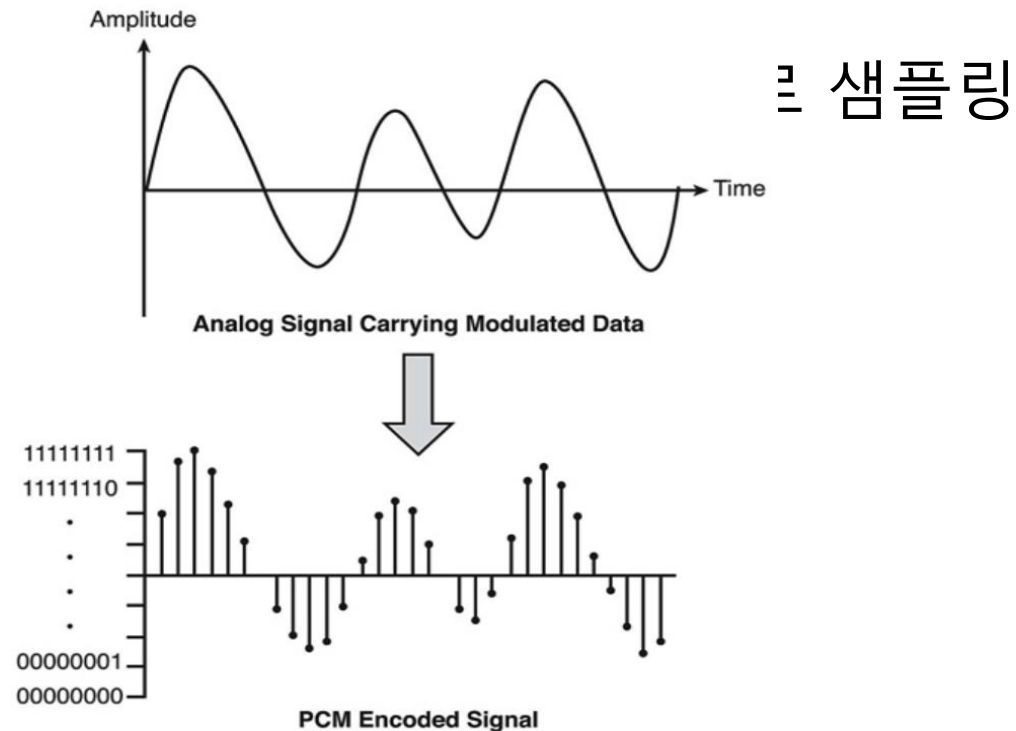


Data, Data, Data



신호 샘플링

- 데이터 수집의 수준 결정
- 나이퀴스트의 샘플링 이론 (Nyquist sampling theorem)
 - 신호의 최고



데이터 분석

- 데이터분석 타입
 - descriptive – 설명적 분석
 - predictive – 예측 분석
 - recommendation - 추천

Descriptive 분석

- 고객, 비즈니스 프로세스, 성과, 데이터를 이해하는 것
- 예
 - market research,
 - funnel analysis (고객의 행동전환이 이루어지는 단계 분석)
 - dashboard reporting
 - exploratory data analysis (EDA)
- 결과는 인사이트를 얻는 것
 - 비즈니스 인텔리전스

Predictive 분석

- 새로운 샘플에 대한 미래 값을 예측하는 것
- 예
 - 날씨, 주가, 판매량 예측
 - 스팸 메일 예측, 질명 예측
 - 고객 세그멘테이션 예측
 - 사이버 공격의 위험 예측

추천

- 최종적으로 의사결정을 돕는 것
 - 단순히 정보(인사이트)를 주는 것을 넘어서 최적의 추천
 - 다양한 설명 및 예측 모델을 종합적으로 활용
- 예
 - 약의 처방, 네비게이터, 검색엔진, 상품 추천
 - 자율차의 운행
 - 알파고와 같은 게임 플레이어,
 - 보험 사기청구 거절 등