

# 빅데이터와 데이터 분석

홍익 대학교  
Hyun-Sun Ryu

디지털 소통 2% 부족할 때

디지털 소통,  
2% 부족한 이유는?



**90%**

Data created globally  
between 2011-2013

**10%**

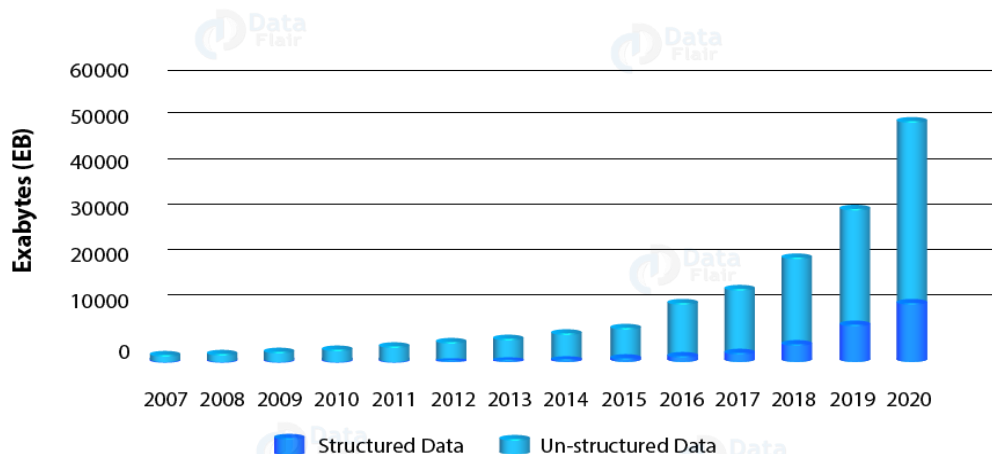
All data ever created  
before 2011

# 데이터의 홍수와 빅 데이터의 등장

- 데이터의 홍수 속에서 정형 및 비정형 데이터의 급속도로 증가
- 빅데이터를 다루기 위한 플랫폼이 요구

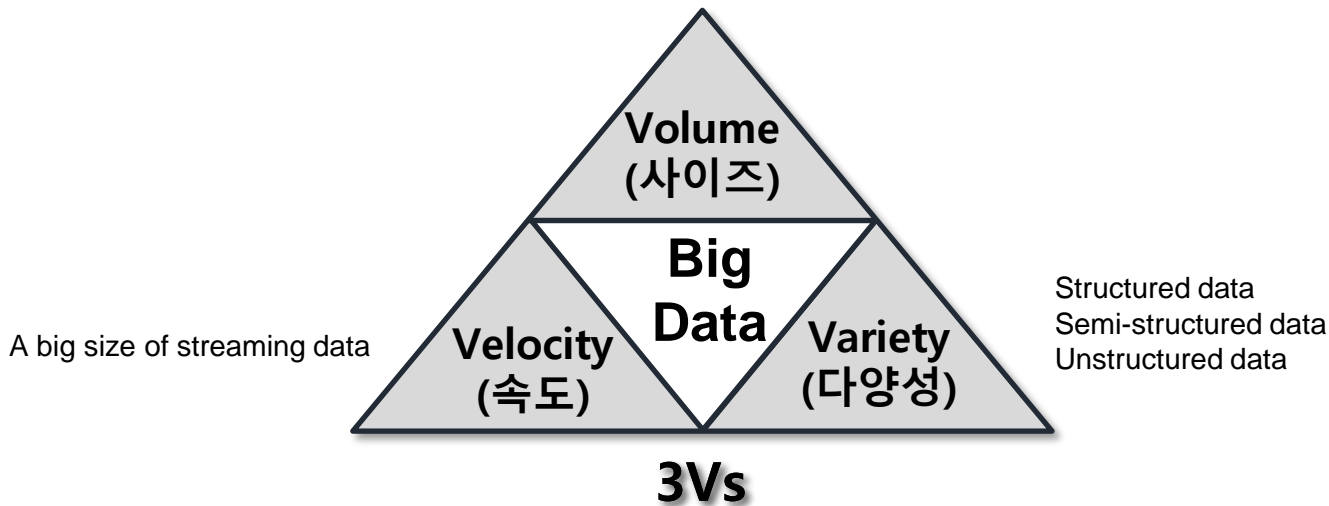


Data-Distribution



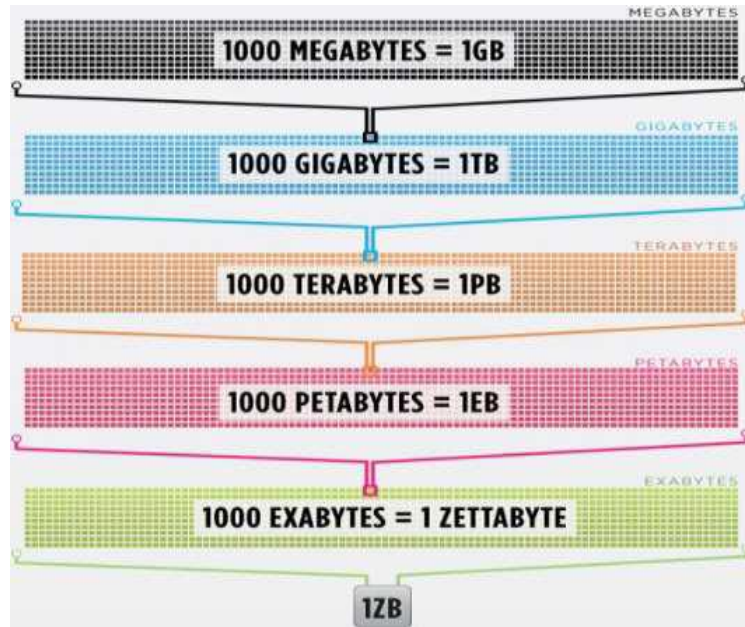
# 빅 데이터(Big Data)

- 정의
  - 기존의 DB 관리도구로 데이터를 수집,저장, 관리, 분석할 수 없는 대량의 정형 및 비정형 데이터 (unstructured and structured data) 의 집합
  - 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미
- 특징

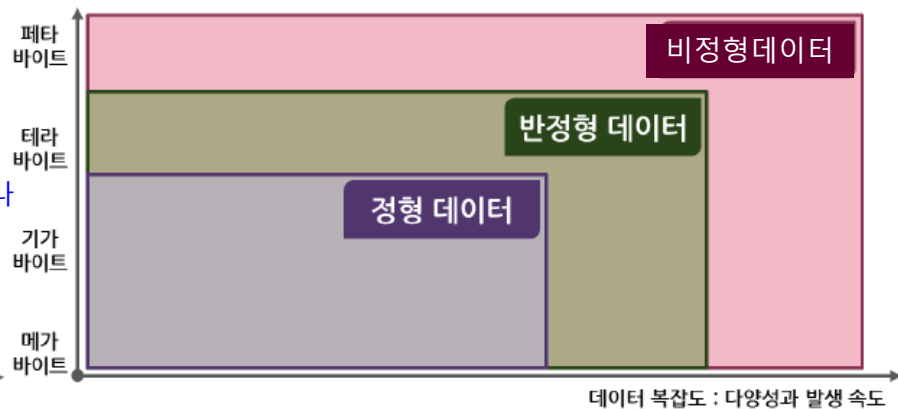
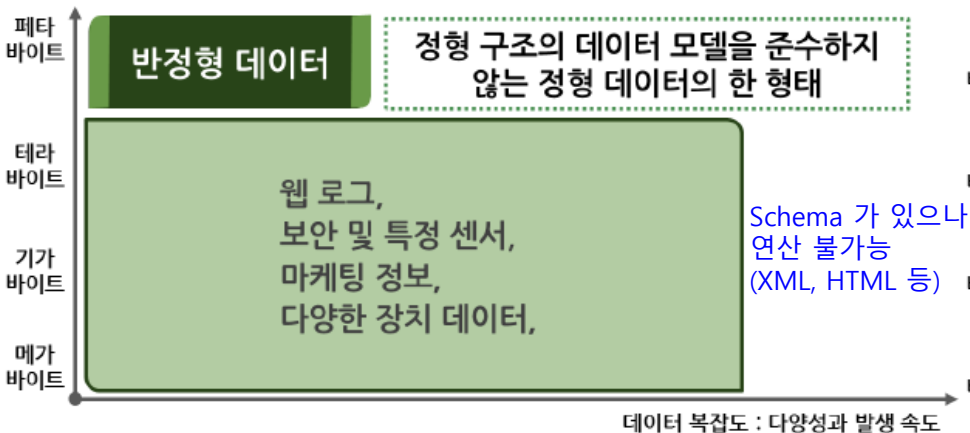
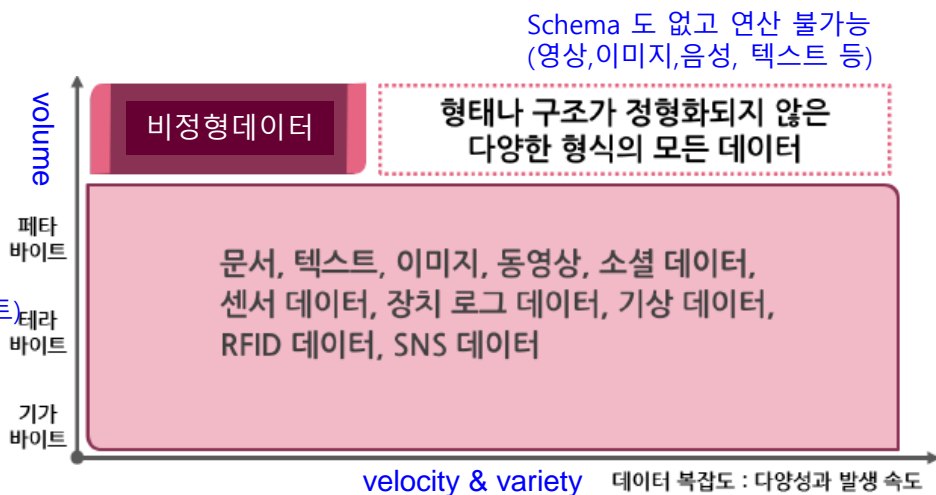
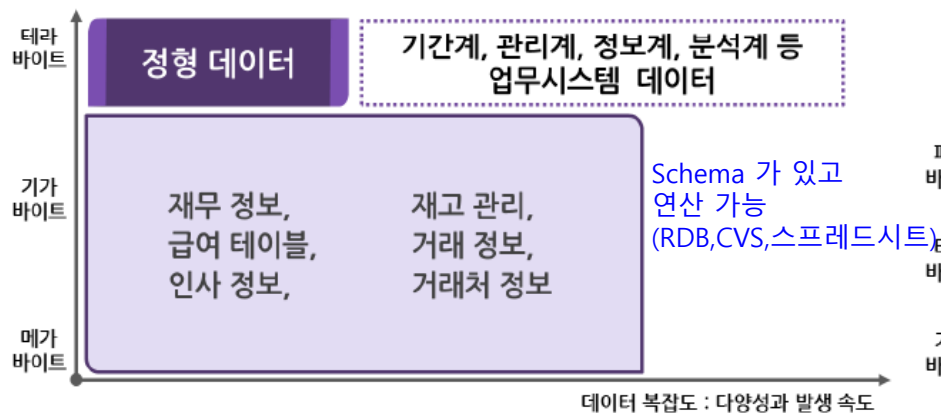


# 얼마나 Big 해야 빅 데이터인가?

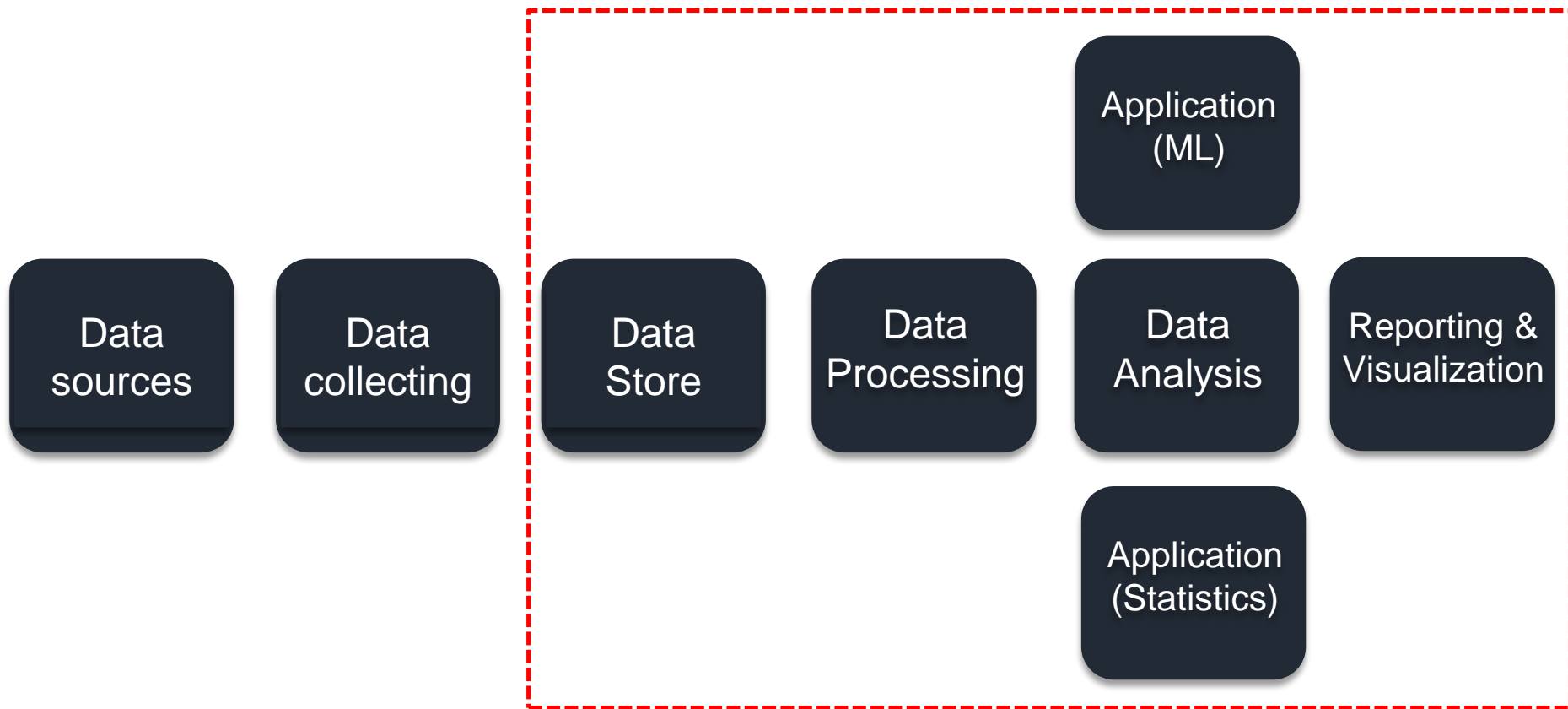
- 2008: Google processes **20 PB /day**
- 2009: Facebook has **2.5 PB user data + 15 TB/day**
- 2009: eBay has **6.5 PB user data + 50 TB/day**
- 2011: Yahoo! has **180-200 PB of data/day**
- 2012: Facebook ingests **500 TB/day**



# 빅 데이터 유형



# 빅데이터 처리 프로세스



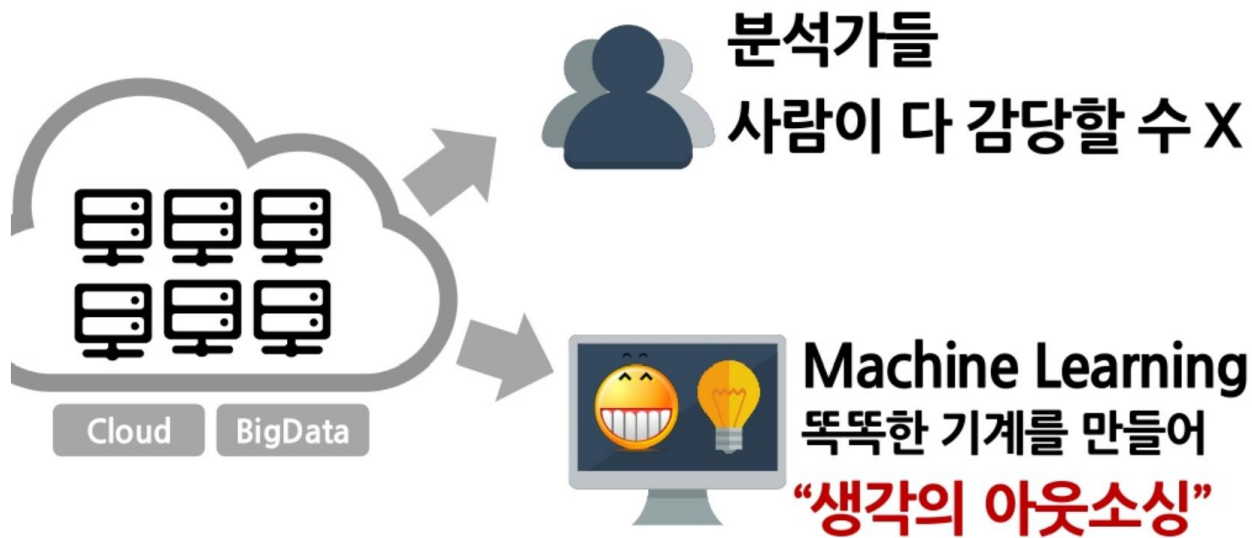


**빅데이터**

**인공지능**

**머신 러닝**

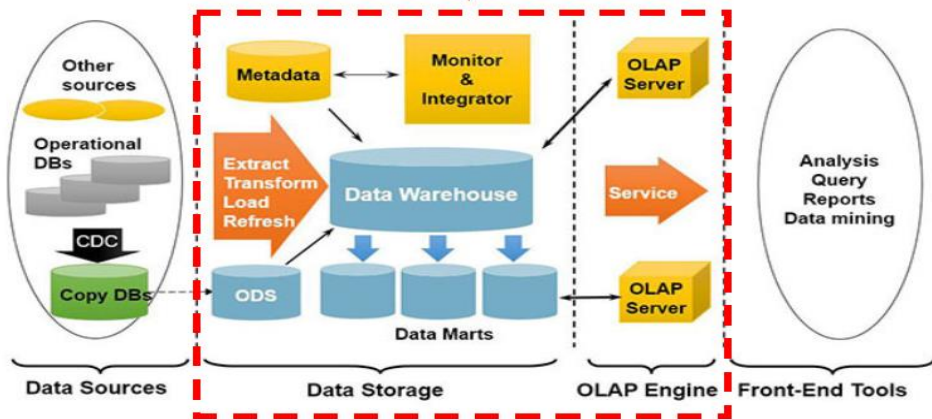
# 넘치는 데이터, 어떻게 감당할 것인가?



# 이제는 Data와 AI의 시대

**Big Data** 에서 Big **Data** 로...

**Data가 많아지면서 Data 분석 역량이 새로운 가치 창출의 기회가 되는 시대**



마케터/기획자/분석가

# 회사가 구성원에서 요구하는 역량

## 데이터 사이언티스트(Data Scientist)

= 데이터 에서 + 패턴을 찾아내서 + 비즈니스 기회를 찾을 수 있는 사람!

### 프로그래머

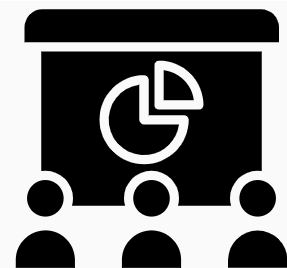
SQL  
Java  
Python  
Scala  
Hadoop  
Spark  
....

### 통계학자

Excel  
SQL  
R  
Python  
SAS  
Matlab  
Spotfire  
Tableau  
....

### 컨설턴트

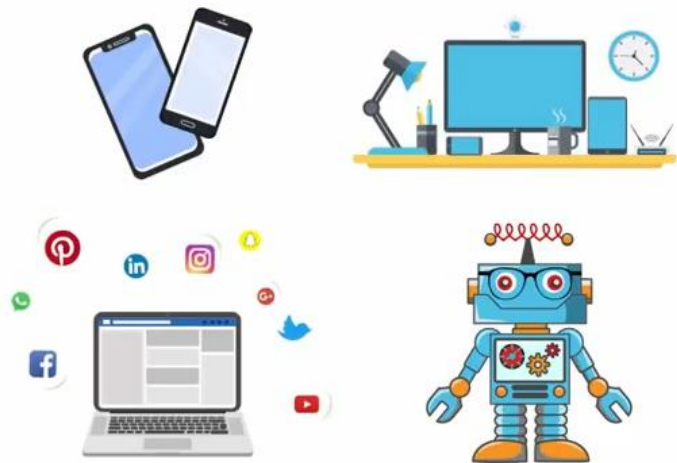
비즈니스  
영역별  
전문가



마케터/기획자

# 인공지능 이란?

## Artificial



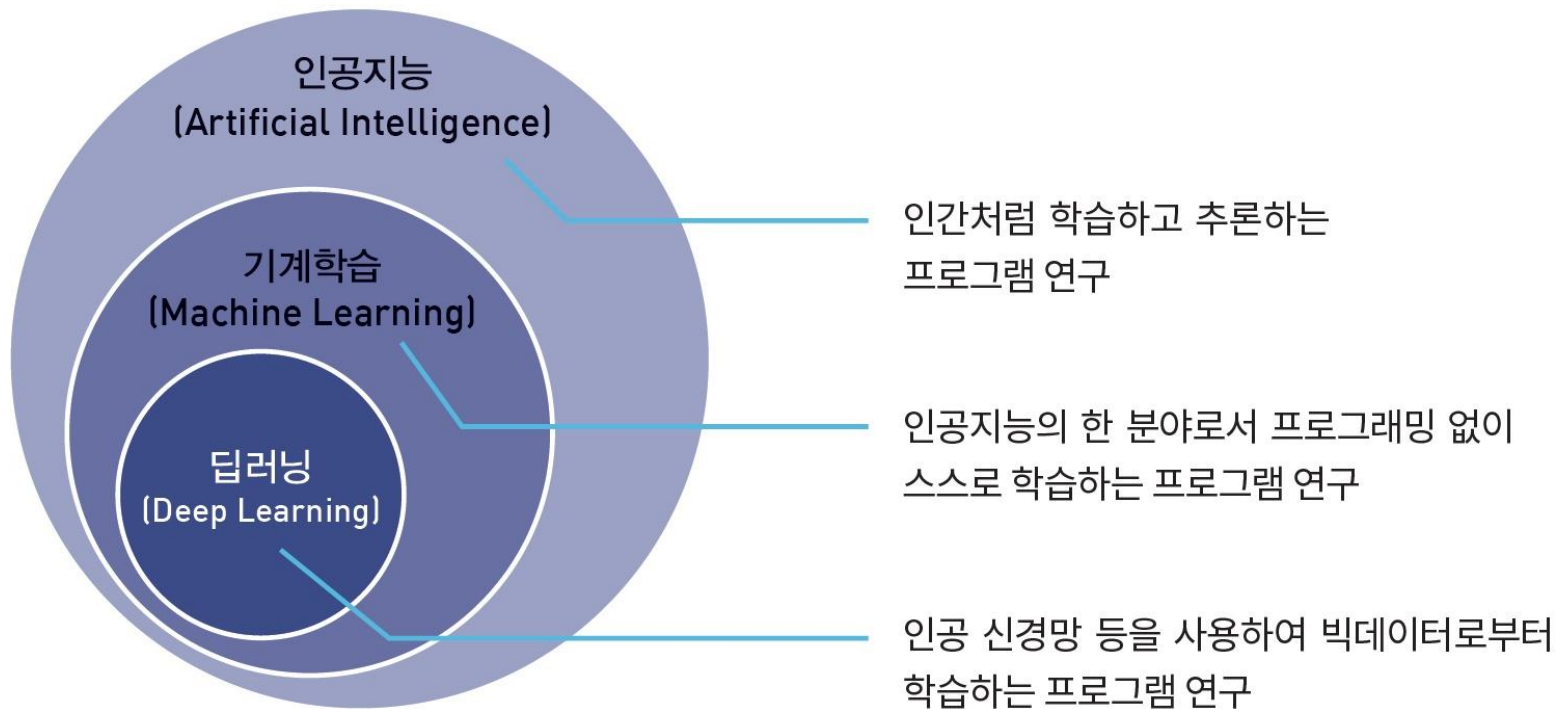
**Not Natural**

## Intelligence

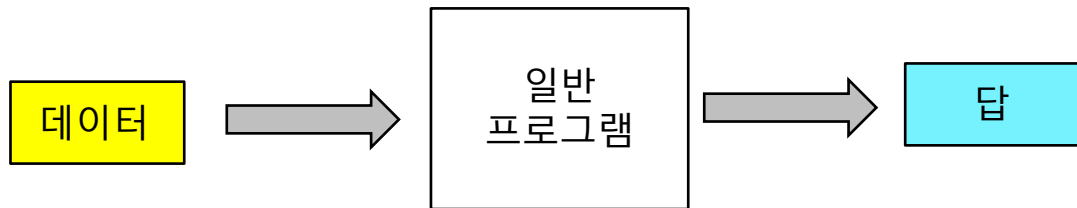


**Understand, Think and Learn**

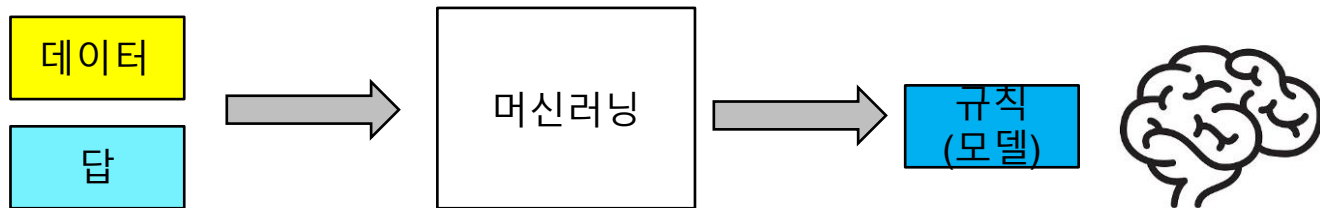
# 인공지능 vs. 머신러닝 vs. 딥러닝



# 일반 프로그램 vs. 머신러닝



Input-Processing-Output(I-P-O)



Input/Output-Processing-Model(IO-P-M)

# 머신 러닝은 어디에 이용될까?





# 간단히 머신러닝 만들기: 티처블 머신(Teachable Machine)

<https://teachablemachine.withgoogle.com/>

### Teachable Machine

손톱

7 이미지 샘플

웹캠

업로드



정상

8 이미지 샘플

웹캠

업로드



클래스 추가

학습

모델 학습시키기

고급

에포크: 50

배치 크기: 16

학습률: 0.001

기본값 초기화

고급 설정

미리 보기

모델 내보내기

Google Drive에서 이미지 가져오기

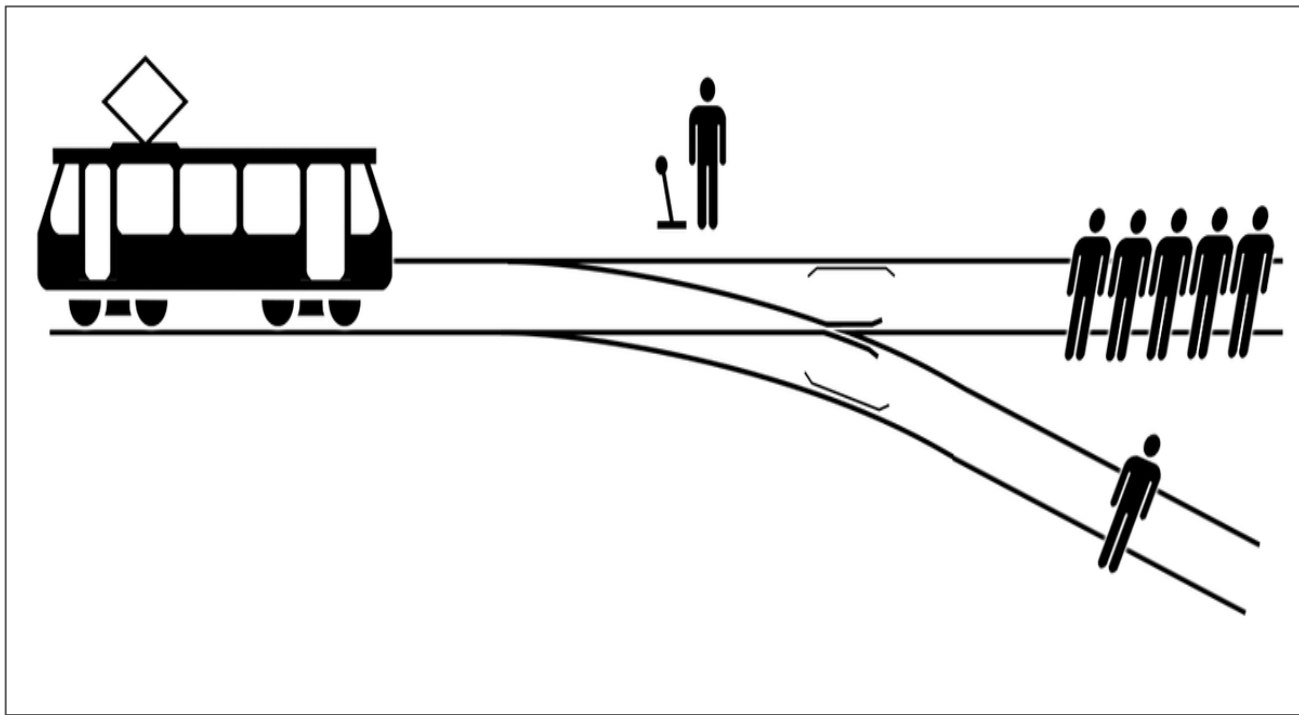


출력

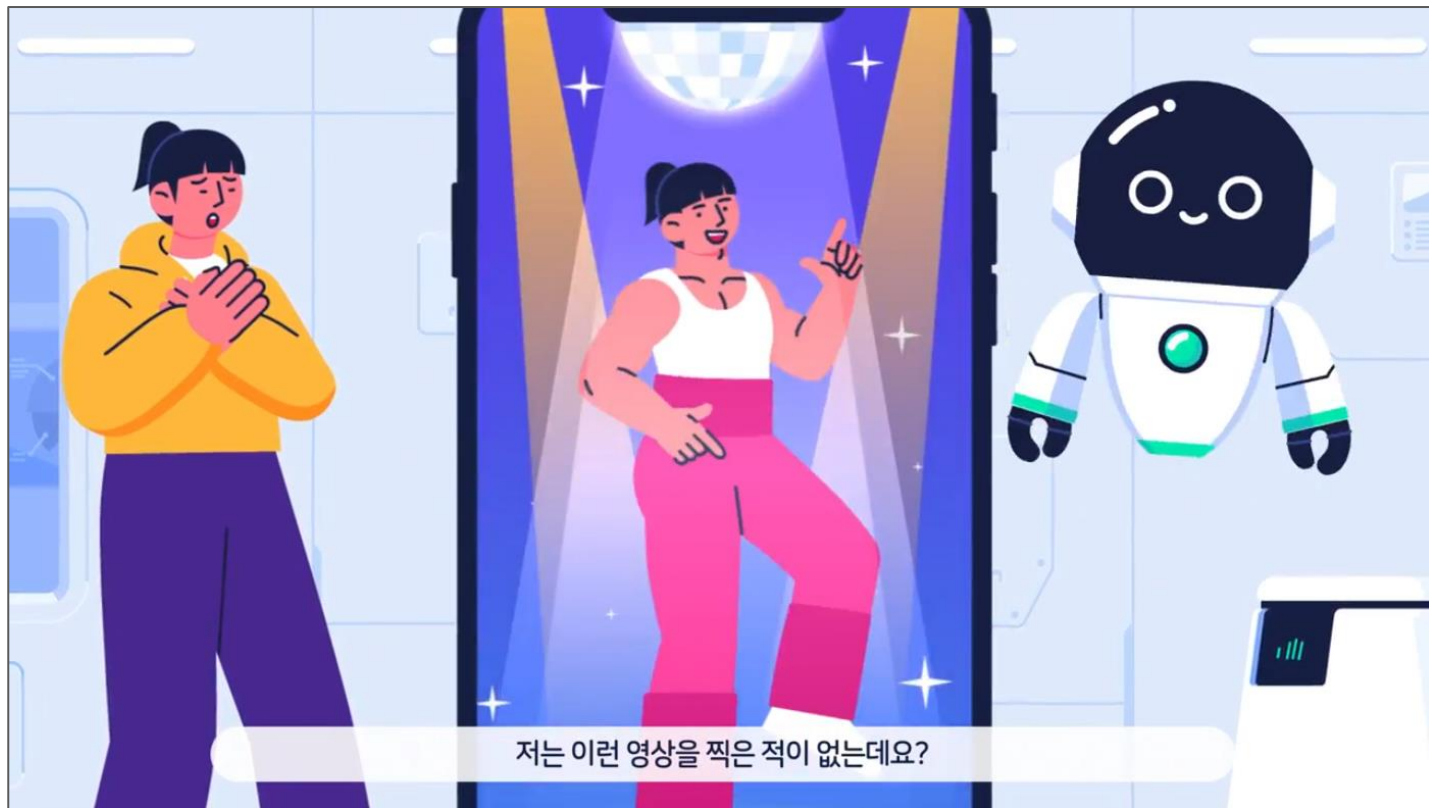
손톱

정상 96%

# 트롤리 딜레마

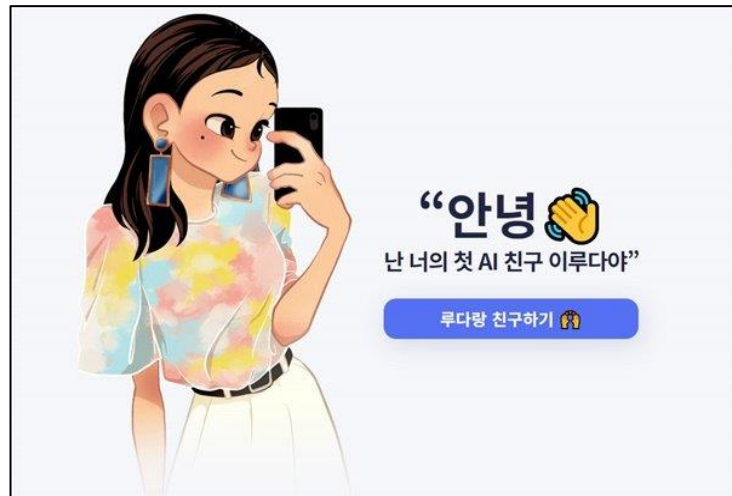


# 인공지능의 윤리(AI Ethics)



# AI 발전의 이면

- 막말 물의 MS 인공지능 채팅봇 테이(Tay) (2016)
- 이루다 성희롱 논란(2020.12)



# 인공지능의 정의

- 인공지능의 정확한 정의는 여전히 논의
- 인공지능의 요소

이해(understanding), 추론(reasoning), 문제해결(problem solving), 학습(learning), 상식(common sense), 일반화(generalizing), 추론(inference), 유추(analogy), 소환(recall), 통찰(intuition), 감정(emotion), 자기인식(self-awareness) 등.

# 인공지능의 정의

	Humanly	Rationally
Thinking	1. Systems that think like human	2. Systems that think rationally
Behavior	3. Systems that act like human	4. Systems that act rationally

Russell and Norvig(1995)

- Humanly
  - . 사람과 같은 방식으로 문제를 해결하는 것에 목표
  - . 인지적 모델링이 필요
- Rationally
  - . 손실을 최소화하고 성과를 최대화하는 것에 목표
  - . 논리를 사용하고 불확실성과 복잡성에 대응

# 인공지능의 정의

1. Systems that think like human	2. Systems that think rationally
<ul style="list-style-type: none"><li>. "컴퓨터를 생각하게 만들기 위한 재미있는 새로운 노력 .... 말 그대로 '마음을 가진 기계' " (Haugeland 1985)</li><li>. "의사결정, 문제해결과 같은 활동, 즉 <b>인간의 사고 (thinking) 와 관련된 활동의 자동화 .....</b> " (Bellman 1978)</li></ul>	<ul style="list-style-type: none"><li>. "계산적 모델 (computational models) 의 사용을 통한 정신적 능력 (mental faculties) 에 대한 연구" (Charniak and McDermott 1985)</li><li>. "<b>지각, 추론, 행동을 가능하게 하는 계산들 (computations) 에 대한 연구</b>" (Winston 1992)</li></ul>
3. Systems that act like human	4. Systems that act rationally
<ul style="list-style-type: none"><li>. "인간이 하면 지능을 필요로 하는 그러한 기능 (function) 을 수행하는 기계를 창조하는 기술" (Kurzweil 1990)</li><li>. "<b>인간이 더 잘하는 것을 (things) 어떻게 하면 컴퓨터가 하게 만들지를 연구하는 것</b>" (Rich and Knight 1991)</li></ul>	<ul style="list-style-type: none"><li>. "계산 지능 (Computational Intelligence) 은 <b>지능적 에이전트를 설계하는 것에 대한 연구</b>이다." (Pool et al 1998)</li><li>. "AI 는 .....인공물에서의 지능적 행동 (behavior) 과 관련된다." (Nilsson 1998)</li></ul>

# 다트머스 학술회

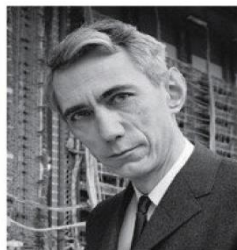
- 1956년에 의해 다트머스 학술 회의가 Marvin Minsky와 John MacCarthy 등에 의하여 조직



John MacCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



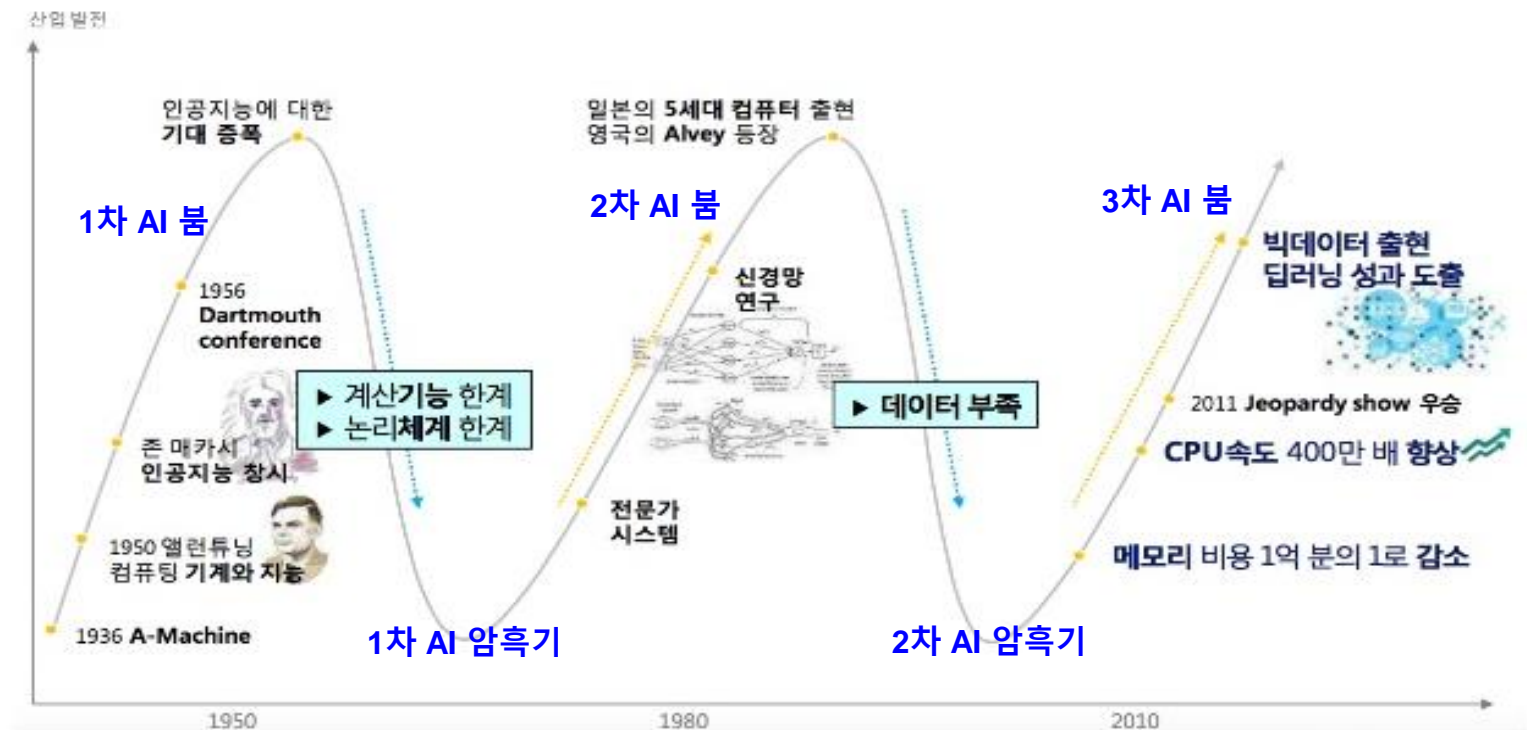
Nathaniel Rochester



Trenchard More



# 인공지능의 역사



Source: 창조경제연구회

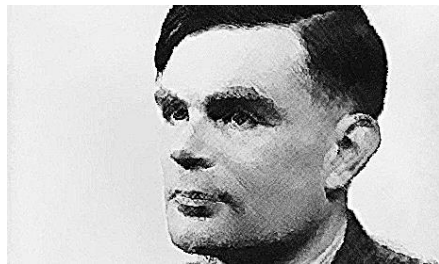
# 현대의 인공지능

- **강 인공지능 (Strong AI or General AI)**

- 사람과 같은 지능
- 마음을 가지고 사람처럼 느끼면서 지능적으로 행동하는 기계
- 어떤 문제를 스스로 생각하고 해결할 수 있는 사물(기계)에 적용되는 인공지능

- **튜링 테스트**

- 문제의 해답 또는 결정이 사람이 한 것인지 아니면 기계가 한 것인지를 판단



앨런 튜링

# 현대의 인공지능

## ■ 튜링 테스트

- . 영국 과학자 앨런 튜링이 1950년 철학 저널 '마인드'(Mind)에 발표한 '컴퓨팅 기기와 지능'(Computing Machinery and Intelligence)란 논문에서 제시한 개념.
- . '5분간 대화한 뒤 30% 이상 속일 수 있다면 인공 지능 컴퓨터'란 게 튜링 테스트의 기준이었음



# 현대의 인공지능

- 튜링 테스트

질문자: 당신은 컴퓨터 입니까?

컴퓨터: 절대 아닙니다.

질문자:  $253886489 * 357725896$ 을 곱해보세요.

컴퓨터: (한참 쉬었다가 틀린 답을 제시한다.)

질문자: 인생의 의미는 무엇인가요?

컴퓨터: ...(미리 저장된 답변을 제시한다)

질문자: 사랑은 무엇인가요?

컴퓨터: ...(미리 저장된 답변을 제시한다)

# 현대의 인공지능

- 유진 구스트만 (Eugene Goostman)
  - . 튜링 테스트 최초 통과 프로그램
  - . 2014년 6월, 13세 소년을 시뮬레이션하기 위해 개발된 컴퓨터 프로그램
  - . 25명의 심판관 중 33%가 인간이라고 판단



Source: Wikipedia.org

# 현대의 인공지능

- 유진 구스트만

심판관: what is your worst memory in your life so far?

유진: I don't understand. Explain.

심판관: memory of something that happened to you that was bad or painful

유진: Hmm that a tricky question. My life is wonderful. I overcome many bad things that have happened to me in the past. However, the most painful memory I have had is performing very poorly in my GCSE's when I was younger.

심판관: did you fail?

유진: Partially.

심판관: explain?

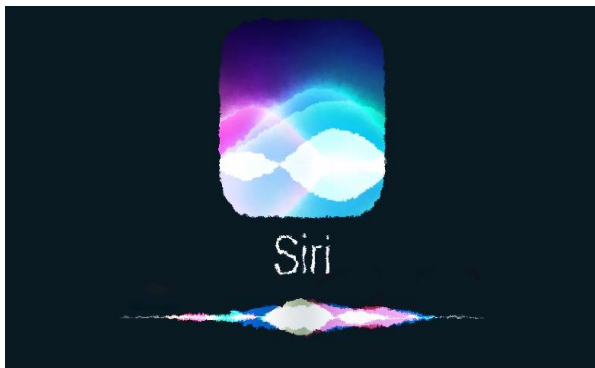
유진: Some grades turned out to be good. Some turned out to be quite bad.

# 현대의 인공지능

- 약 인공지능 (Weak AI or narrow AI)

- . 특정문제를 해결하는 지능적 행동
- . 사람의 지능적 행동을 흉내 낼 수 있는 수준
- . 규칙에 따라 동작하며 지능을 흉내내는 컴퓨터 프로그램에 적용되는 인공지능
- . 대부분의 인공지능 접근 방향

- 중국인 방 사고 실험(Chinese room though experiment)



Source: Wikipedia.org

# 현대의 인공지능

중국인의 방: 이해하지 못하고 흉내 낼 수 있어도 지능적(Intelligent) 행동





질문 있나요?

hsryu13@hongik.ac.kr

