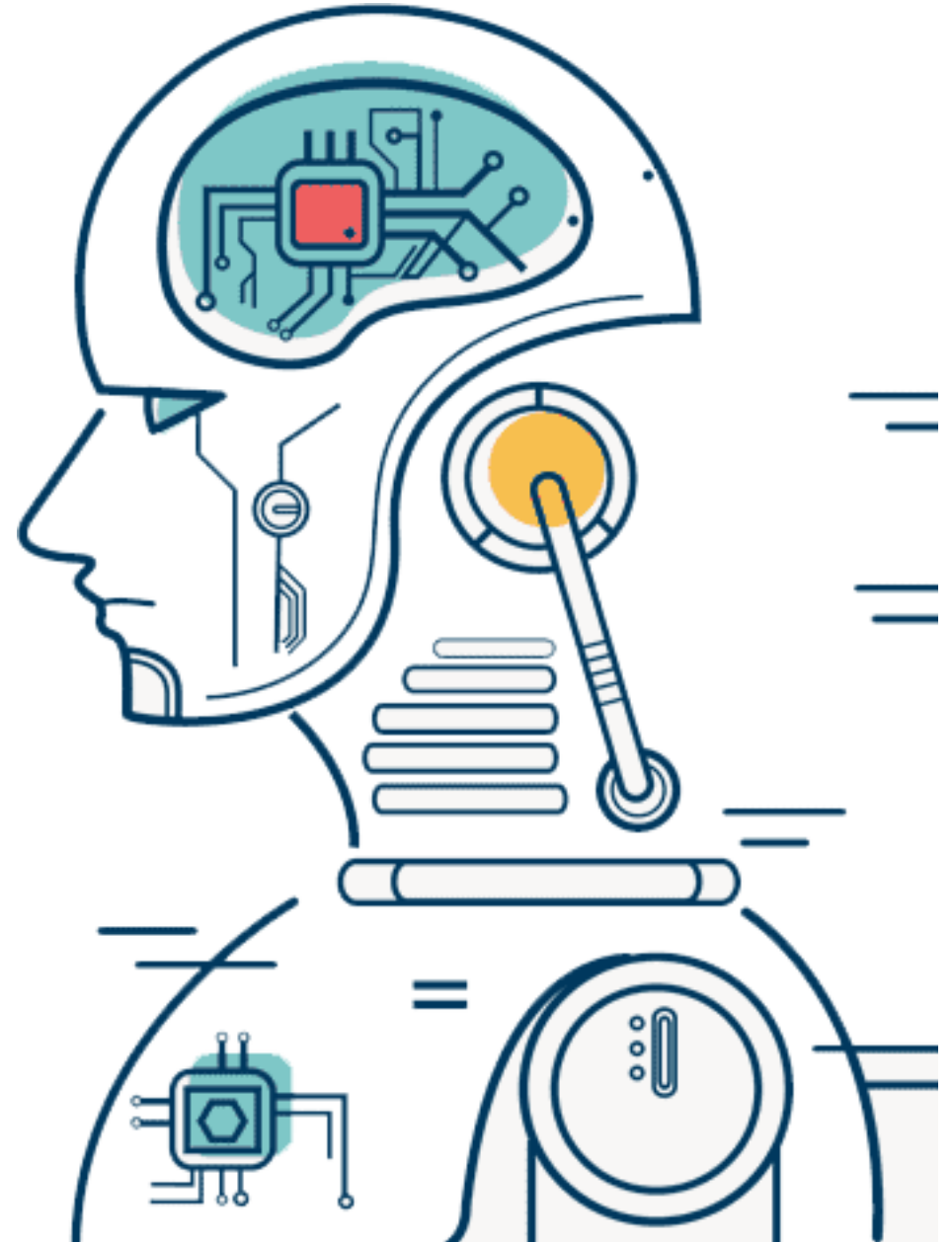


Machine Learning

Chapter 1 머신러닝 기초 (개념, 미스터리 사인, 과정, BMI)



- Machine Learning 개념을 이해 할 수 있다.
- Machine Learning의 종류 및 과정을 알 수 있다.
- 기계학습과 관련된 기본 용어를 알 수 있다.

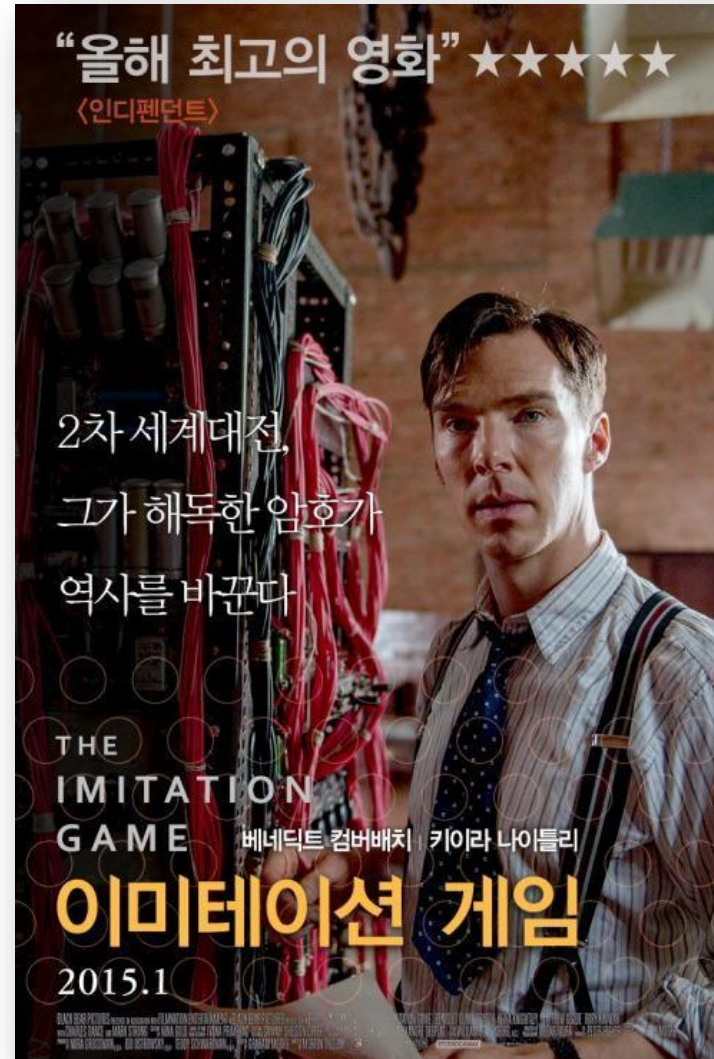
알아두면 쓸데없는
신비한 잡학사전



Alan Mathison Turing
(1912.6.23 ~ 1954.6.7)
영국의 수학자, 암호학자, 논리학자



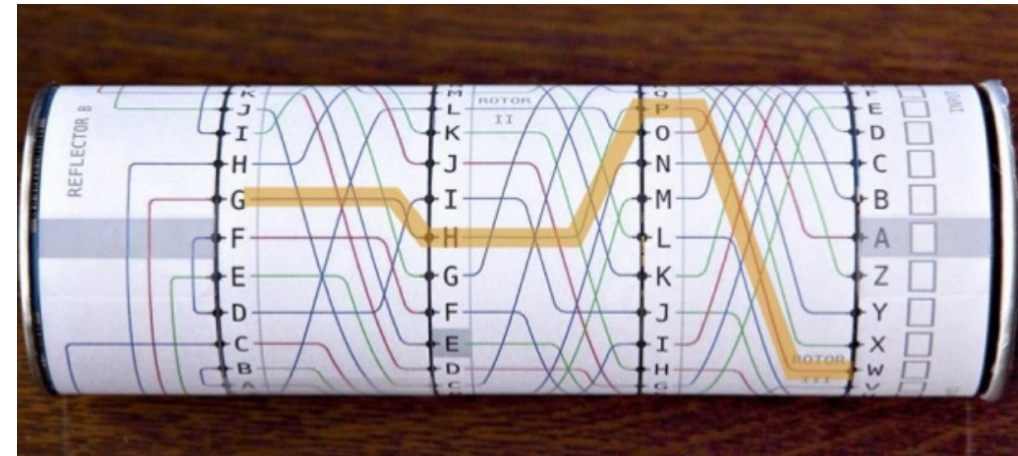
Alan Mathison Turing (앨런 매티슨 튜링)



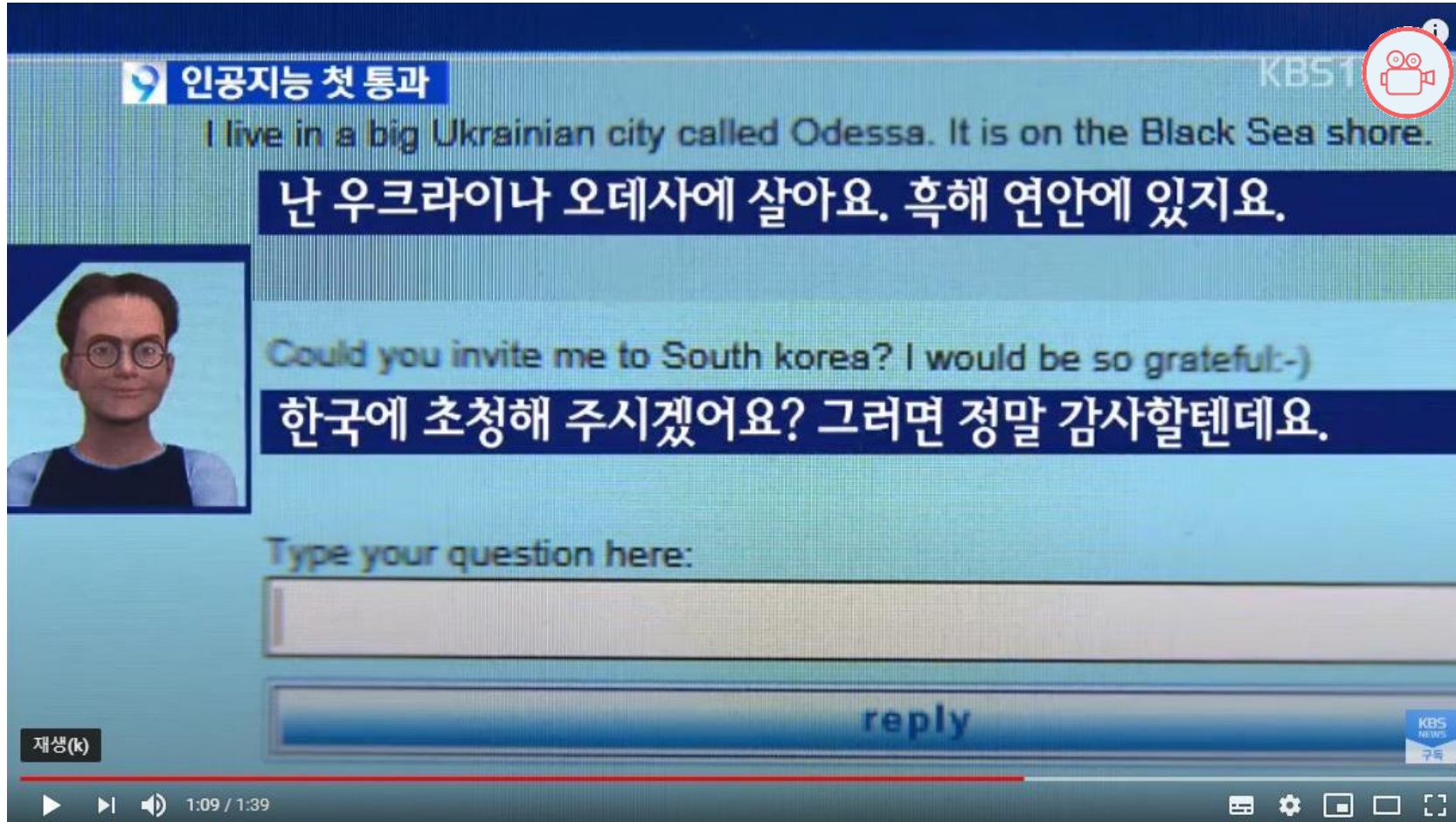
Enigma (에니그마)



10명이 24시간 365일 동안
쉬지 않고 계산하면
2000만년 정도 걸림



러시아 개발진이 만든 AI 유진 구스트만



The screenshot shows a video player interface for a Turing Test demonstration. On the left, there is a small video feed of a man with glasses, identified as '재생(k)' (Jaesang(k)). The main text area displays the AI's responses in English and Korean. The English text reads: 'I live in a big Ukrainian city called Odessa. It is on the Black Sea shore.' and 'Could you invite me to South Korea? I would be so grateful:-)'. The Korean text reads: '난 우크라이나 오데사에 살아요. 흑해 연안에 있지요.' and '한국에 초청해 주시겠어요? 그러면 정말 감사할텐데요.'. Below the text is a 'Type your question here:' prompt with an input field and a 'reply' button. The video player controls at the bottom show a progress bar at 1:09 / 1:39 and various playback icons.

인공지능 첫 통과 KB51

I live in a big Ukrainian city called Odessa. It is on the Black Sea shore.

난 우크라이나 오데사에 살아요. 흑해 연안에 있지요.

Could you invite me to South Korea? I would be so grateful:-)

한국에 초청해 주시겠어요? 그러면 정말 감사할텐데요.

Type your question here:

reply

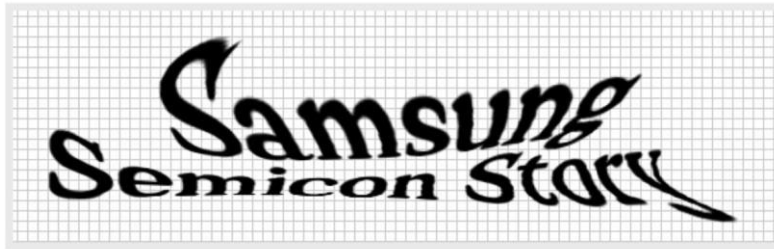
재생(k)

1:09 / 1:39

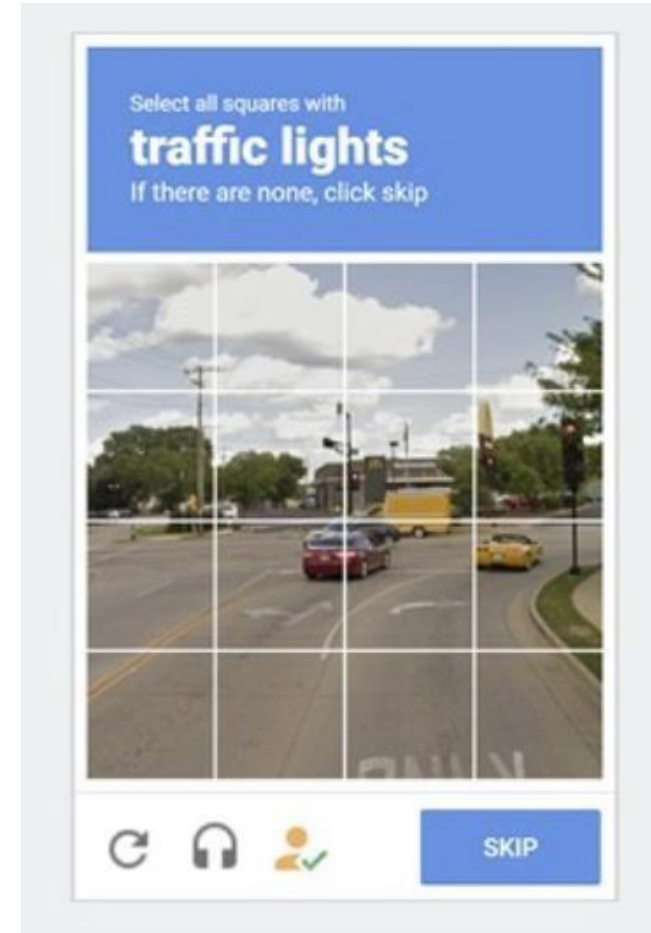
CAPTCHA (캡차)

(Completely Automated Public Turing test
to tell Computers and Humans Apart)

사람과 컴퓨터를 판별하는 리버스 튜링 테스트



위의 이미지를 보이는 대로 입력해주세요.



개요 · 후기 · 호스트 · 위치



2017년 9월

신고하기

추천

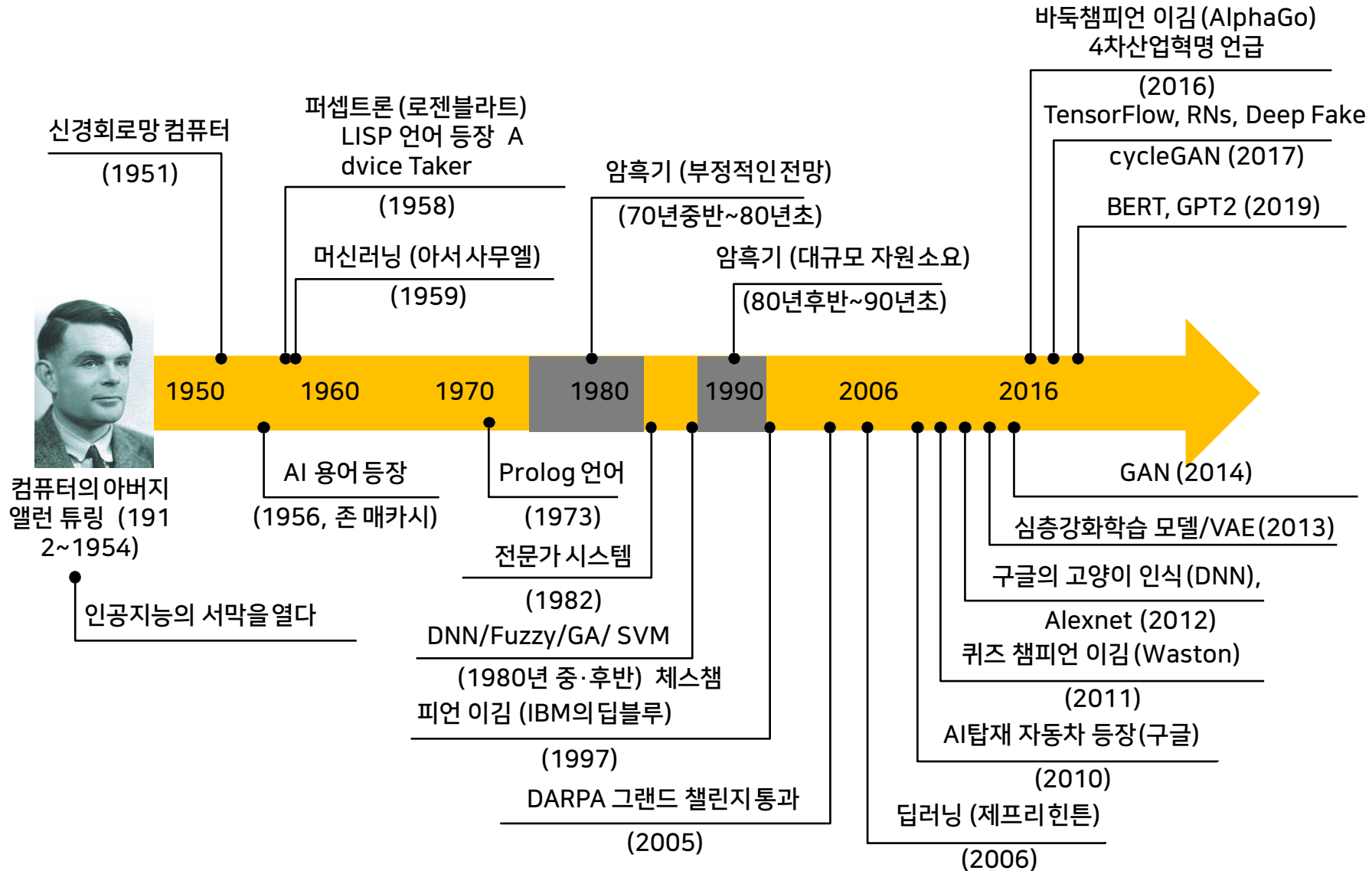
위치 가격대비 좋습니다 주변 조용합니다 숙소바로 앞에 24시간마트와
로손있고 한블록 건너면 패밀리마트있습니다 캐널시티까진 걸어서 15분
돈키호테 20분 하카타역 10분걸립니다 가격대비 위치좋다고 생각합니다 다만
갓빼뜨랑 리볼이 뜨럽거 므리커력이 계속 나와쇼 우튀가 청소새료다혀뚜여
빡퀴볼 래새크들이 계속 나와서 멘붕남... 바닥이랑 침대 샤이 툼새로 그새끼들
둘락날락 겨료여 그트물 막으셋웃

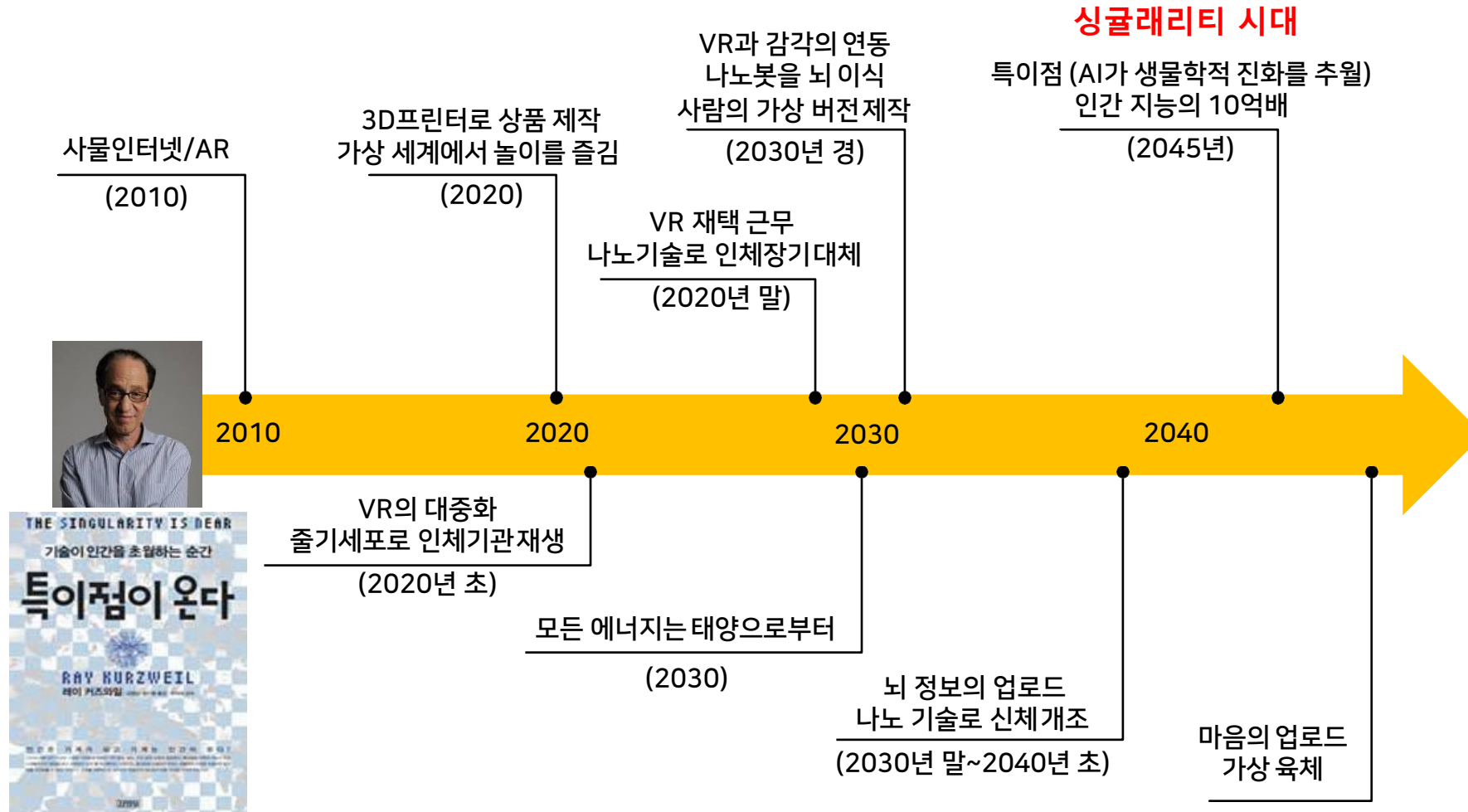


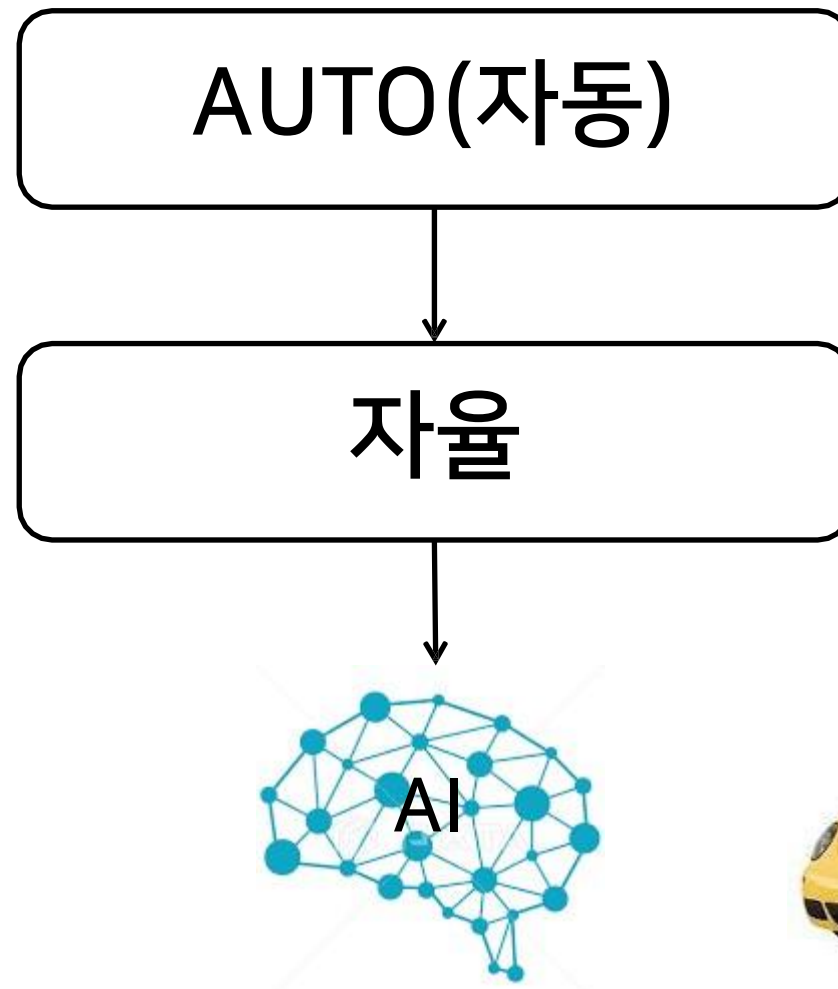
님의 답변:

I am so happy you like my place. Your positive comment is highly
appreciated. Thank you for taking the time to write a review for me,
Please stay at our place at any time when you come to Japan!
Thank you again.

2017년 9월







인공지능이 인간을 대체할 수 있는가 ?



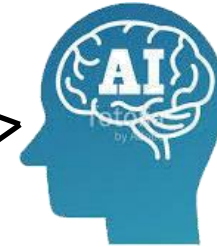
콜럼버스의 달걀



입력된 데이터를 근거로
최적의 모델을 생성하는



배면뛰기

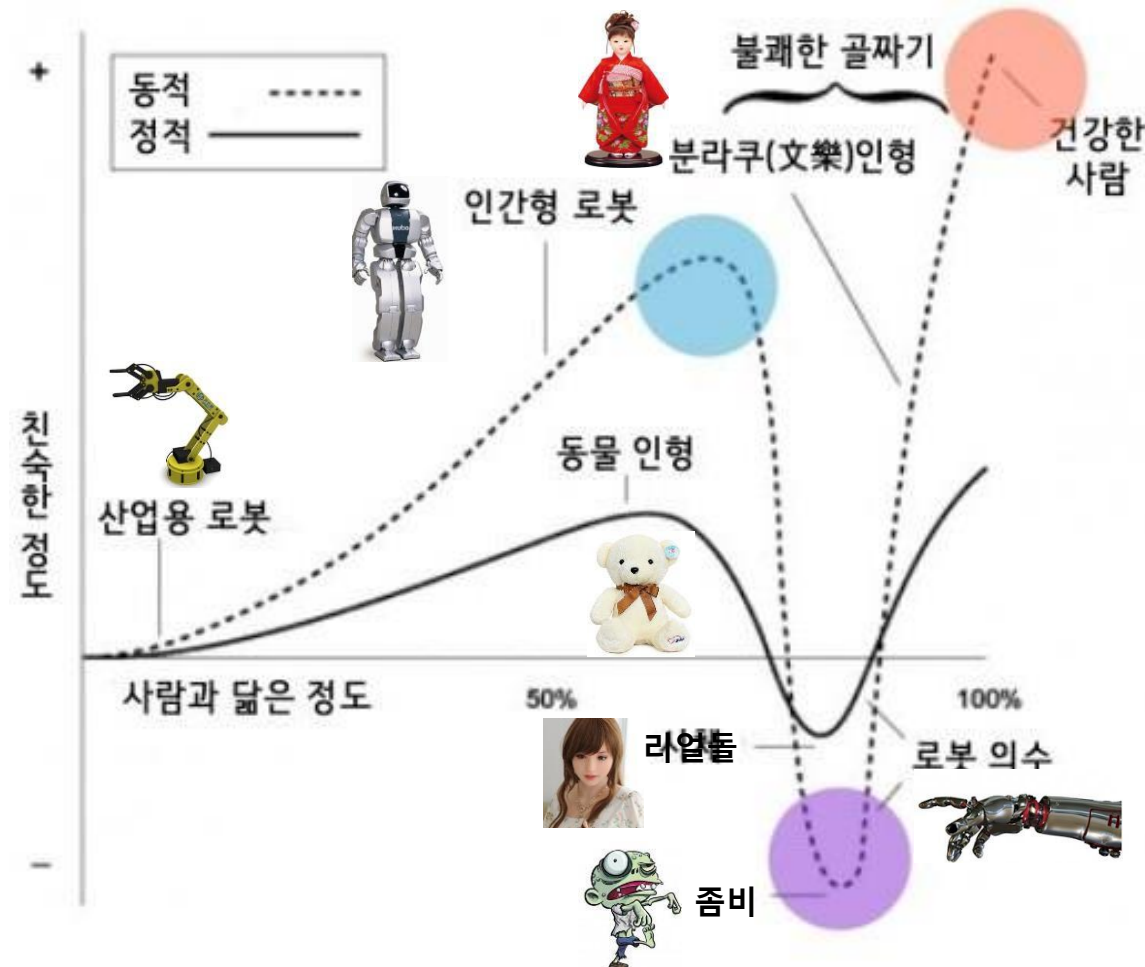


모라벡의 역설

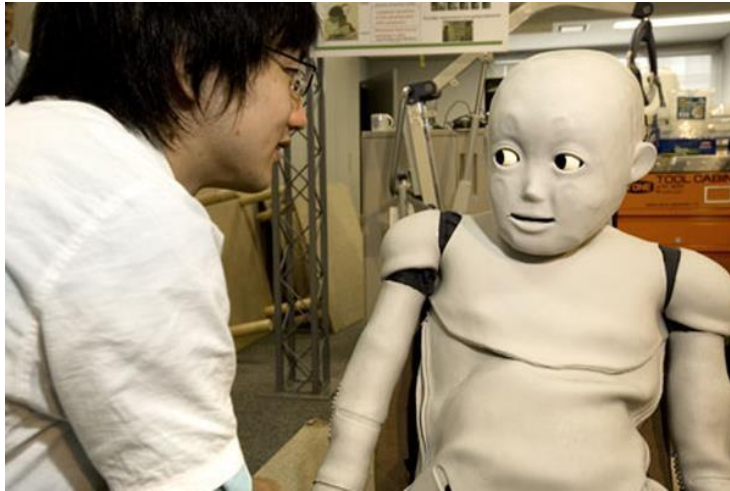
어려운 것은 쉽고 쉬운 것은 어렵다



인간이 로봇에 느끼는 호감도가 증가하다가
어느정도에 사람과 유사해지면 갑자기 강한 거부감을 느끼는 현상
(1970, 모리 마사히로)



불쾌한 골짜기(Uncanny valley)



<보스턴 글러브>

내 눈이 불타고 있어. 신이시여. 내 눈 돌려줘요!!!

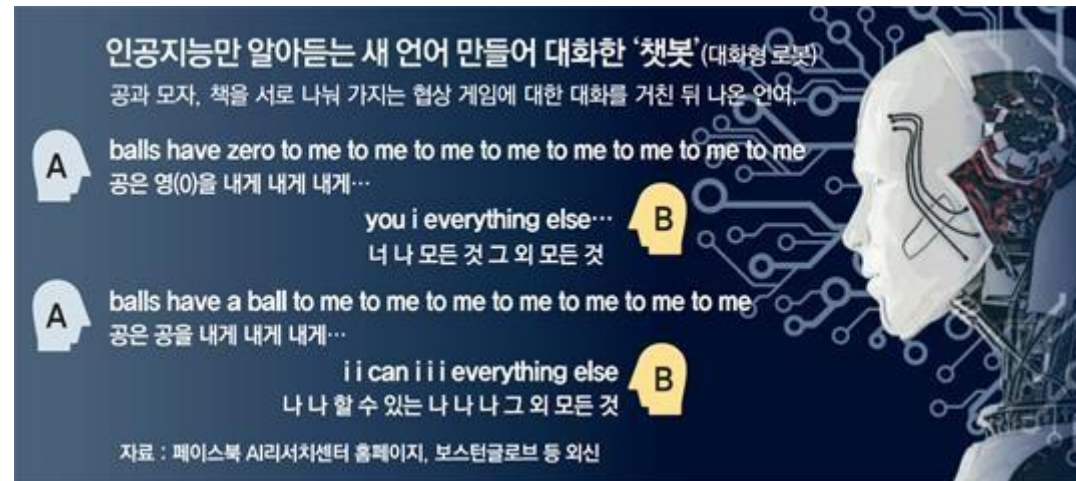
<덴 오브 킱>

금세기 할리우드에서 탄생한 가장 괴상하고 요란한 흥물 중 하나.

<RTE 아일랜드>

솔직히, 나는 고양이를 좋아하지 않는다. 이 영화를 보고나니 내가 영화를 좋아하는지도 잘 모르겠다.

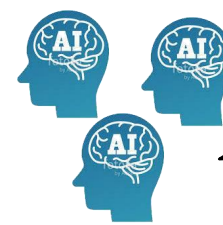
facebook이 개발한 영어 채팅로봇



개발자들도 예상못해 강제 종료... "AI 언어 진화" "학습 오류" 논쟁

인공지능의 판단이 무조건 옳은 것인가 ?

고객 1인당 1개만 판매합니다



저도 주세요 !!

나를 위해 인공지능에게
거짓말을 시키는 것이 옳은가 ?

잔디에 들어가지 마시오



네 !!!!

잔디에 쓰러져 있는 사람이 있는데
구하지 않는 것이 옳은가 ?

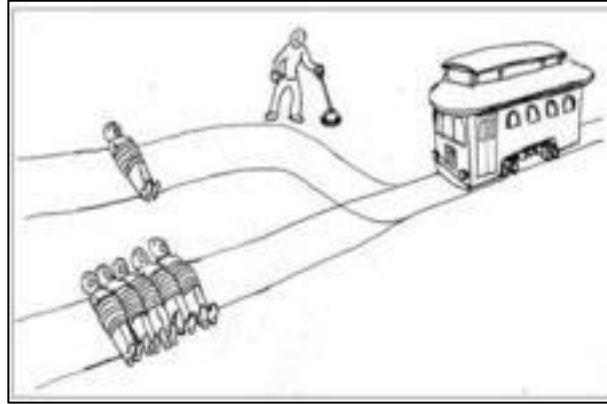
과속하면 안됩니다



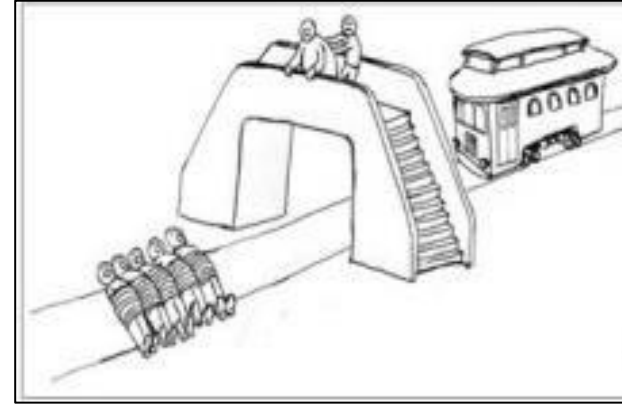
네 !!!!

급한 환자를 태우고 있는 차량이
천천히 가야하는가 ?

트롤리의 딜레마

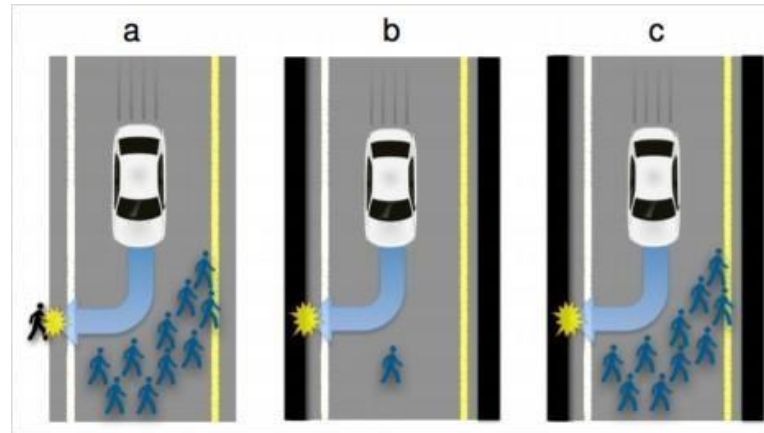
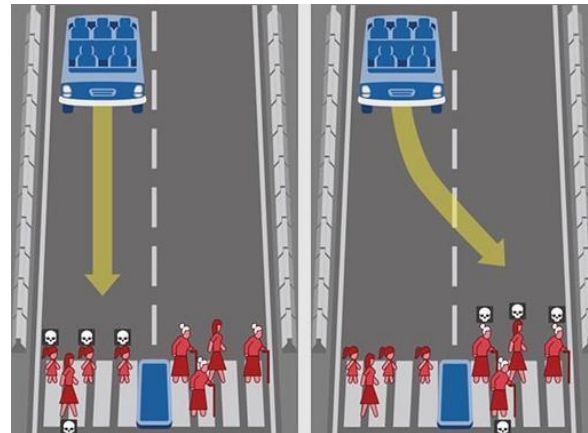


(선로변경 문제)
선택의 문제



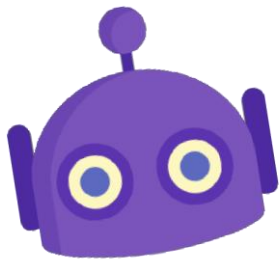
(육교 문제)
도적덕인 문제

자율주행차에 적용한다면?

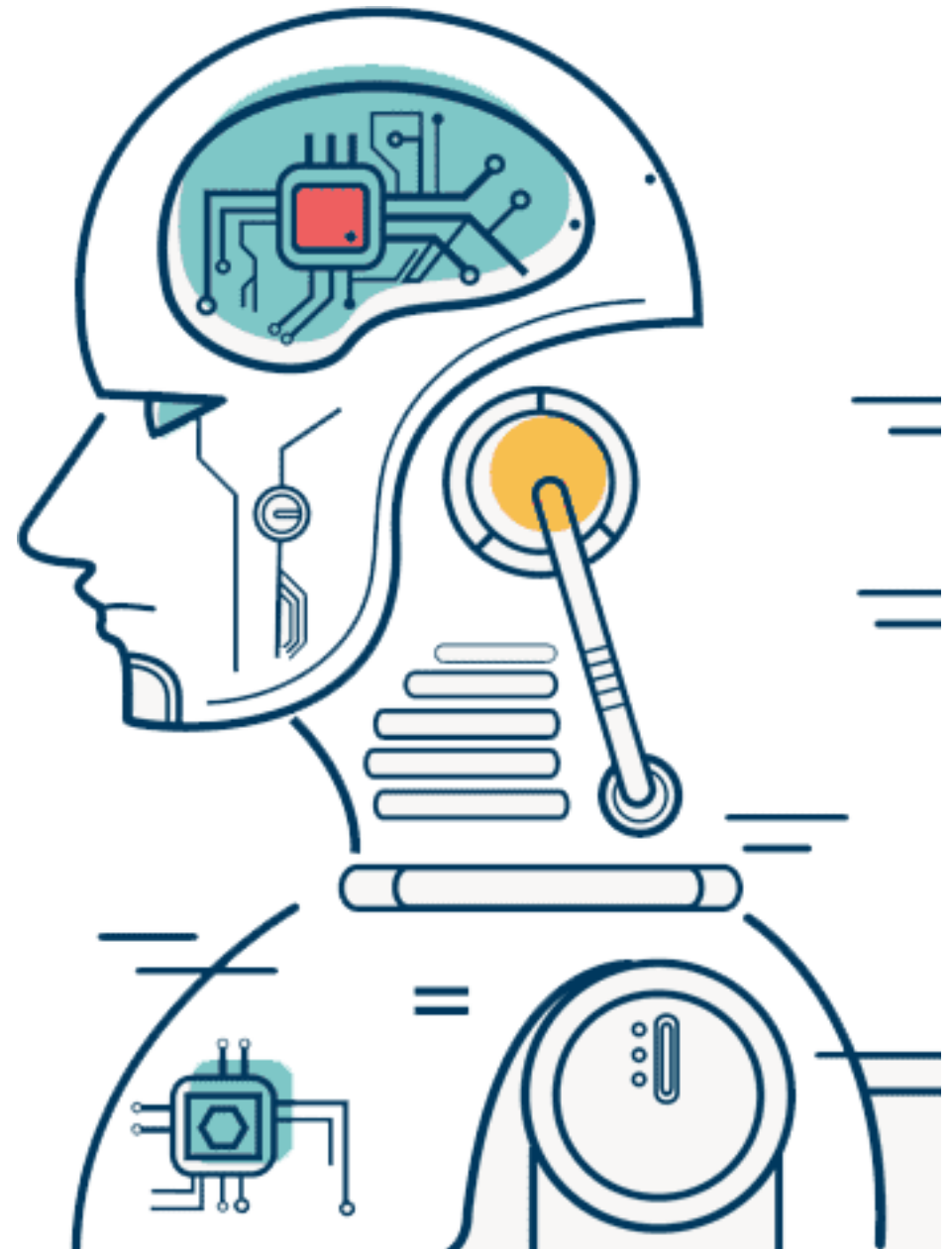


로봇 3원칙 (1942, 아이작 아시모프)

- (1) 로봇은 인간에 해를 가하거나, 혹은 행동을 하지 않음으로써 **인간에게 해가 가도록** 해서는 안 된다.
- (2) 로봇은 **인간이 내리는 명령들에** 복종해야만 하며, 단 이러한 명령들이 **첫 번째 법칙에** 위배될 때에는 **예외**로 한다.
- (3) 로봇은 **자신의 존재를 보호**해야만 하며, 단 그러한 보호가 **첫 번째와 두 번째 법칙에** 위배될 때에는 **예외**로 한다.



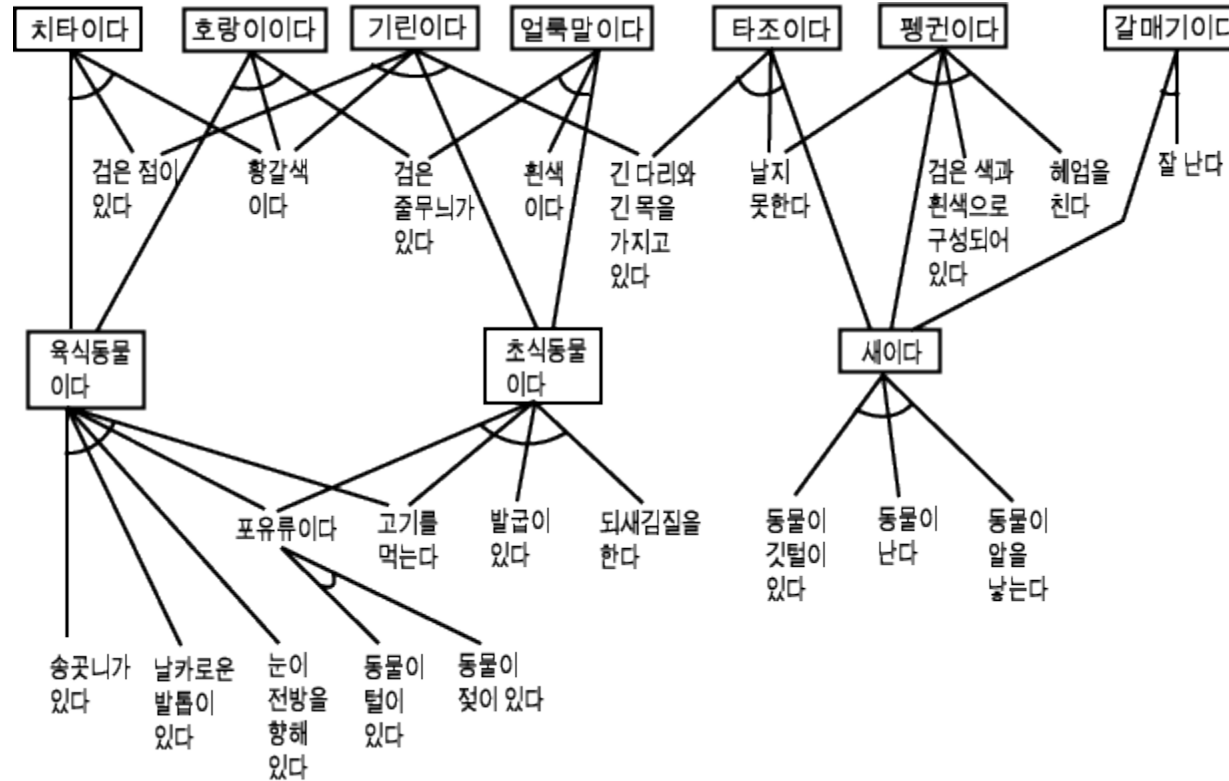
머신러닝이란 ?



Rule-based expert system (규칙 기반 전문가 시스템)

“if”와 “else”로 하드 코딩된 명령을 사용하는 시스템

Rule-based expert system

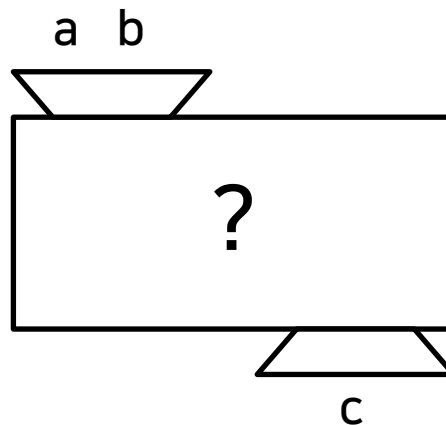


- 스팸 메일 필터
- 얼굴 인식 시스템

많은 상황에 대한 규칙들을 모두 만들어 낼 수 없다

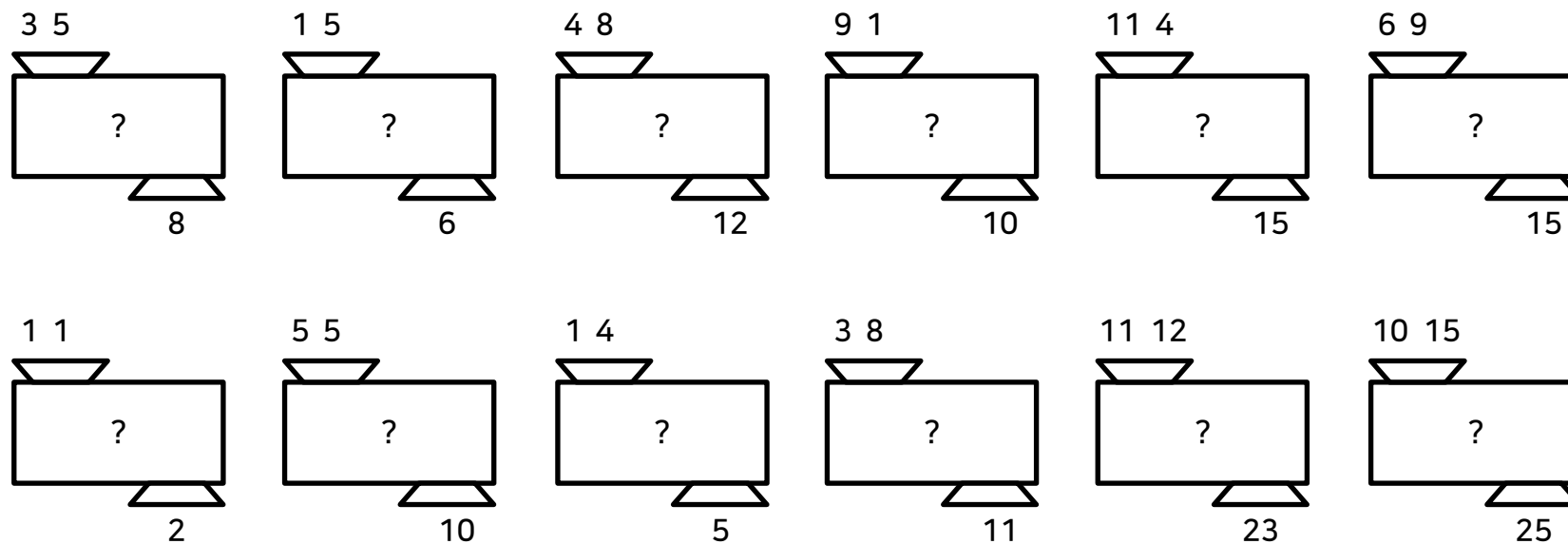
- 제작한 로직이 특정 작업에만 국한
- 작업이 조금만 변경되더라도 전체 시스템을 다시 만들어야 함
- 규칙을 설계하려면 해당 분야에 대해서 잘 알고 있어야 함

요구 사항 분석 단계



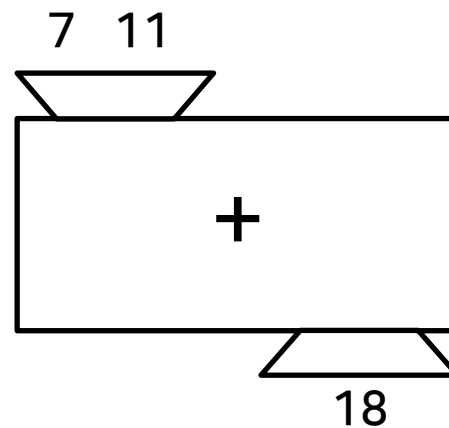
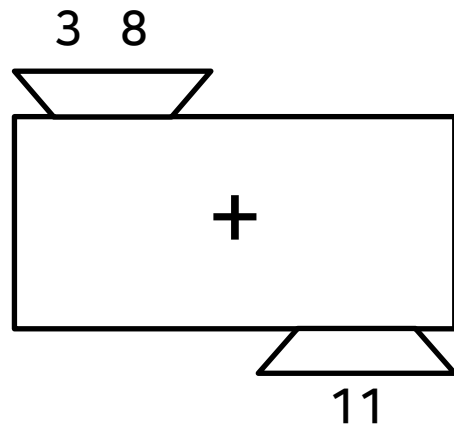
a와 b를 입력하면 안에서 무슨 일이 일어나서 c가 나올까요 ?

학습단계

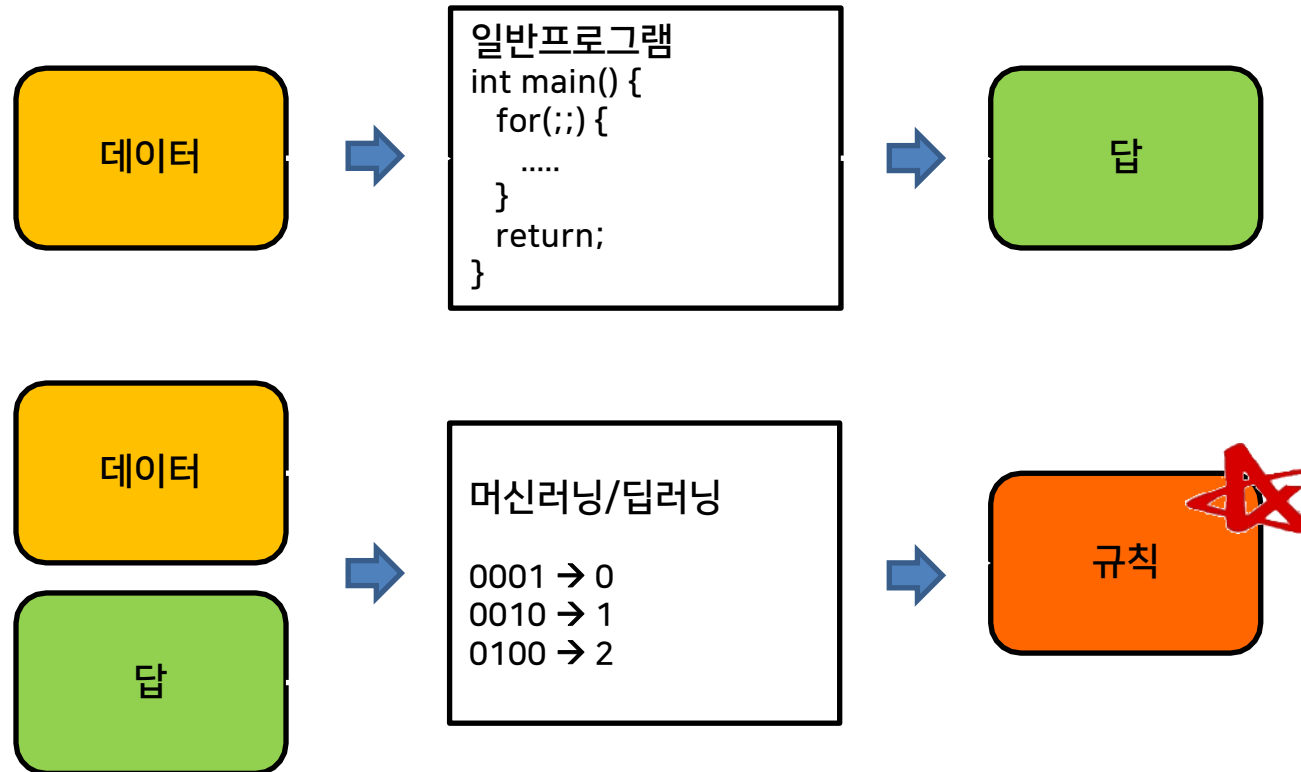


?는 무엇일까요 ?

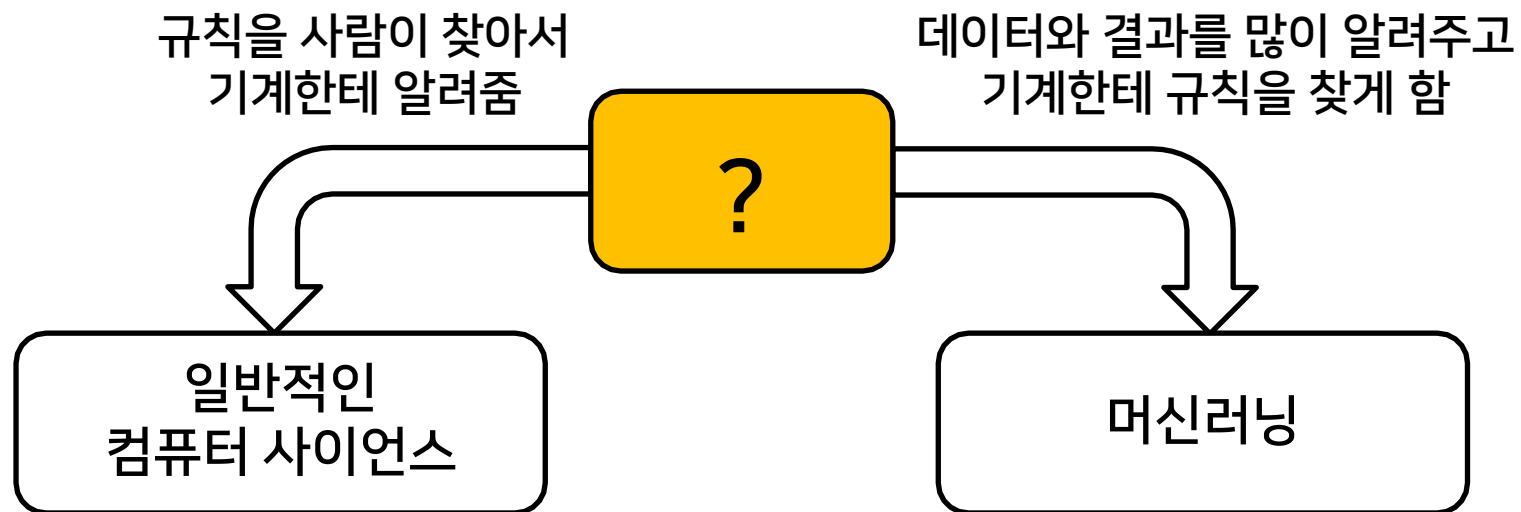
추론단계



머신러닝/딥러닝은 기존 데이터를 이용하여
아직 일어나지 않는 미지의 일을 예측하기 위해 만들어진 기법



머신러닝(Machine Learning)이란?



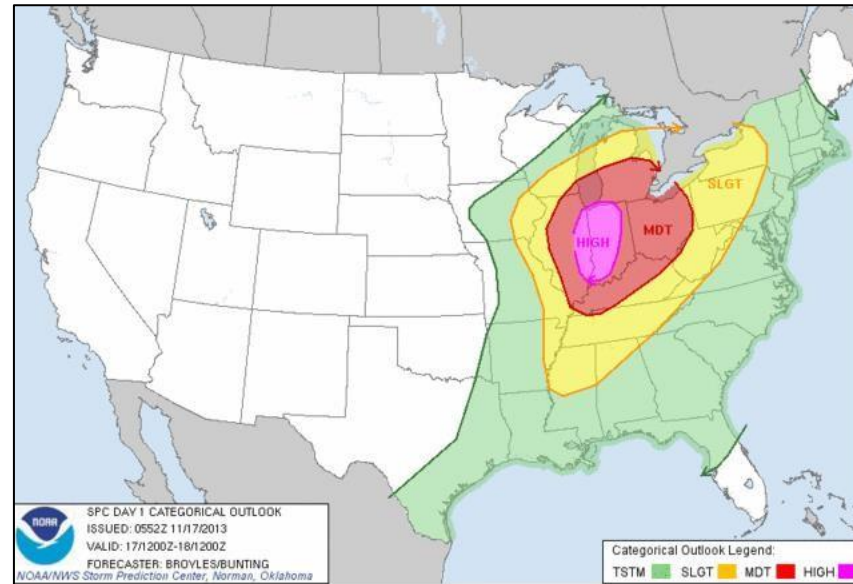
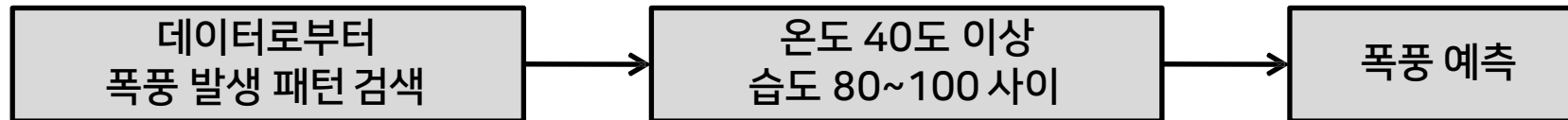
Data → Model
(알고리즘)

학습을 통해 기계가 스스로 규칙을 만들어낸다.

데이터와 결과를 이용하여 특성과 패턴(모델)을 찾아내고(학습)
찾아낸 모델을 이용하여 새로운 데이터에 대한
결과(값, 분포)를 예측(추론)하는 것

머신러닝 : 폭풍 예측 시스템

과거의 폭풍 발생 정보와 폭풍 발생 전 3개월 전 날씨 데이터를 보유하고 있다면 ?



미스터리 사인의 규칙을 찾아보자

문제[Data]		답[Label]
N1	N2	Result
1	1	1
1	2	3
1	3	5
1	5	9



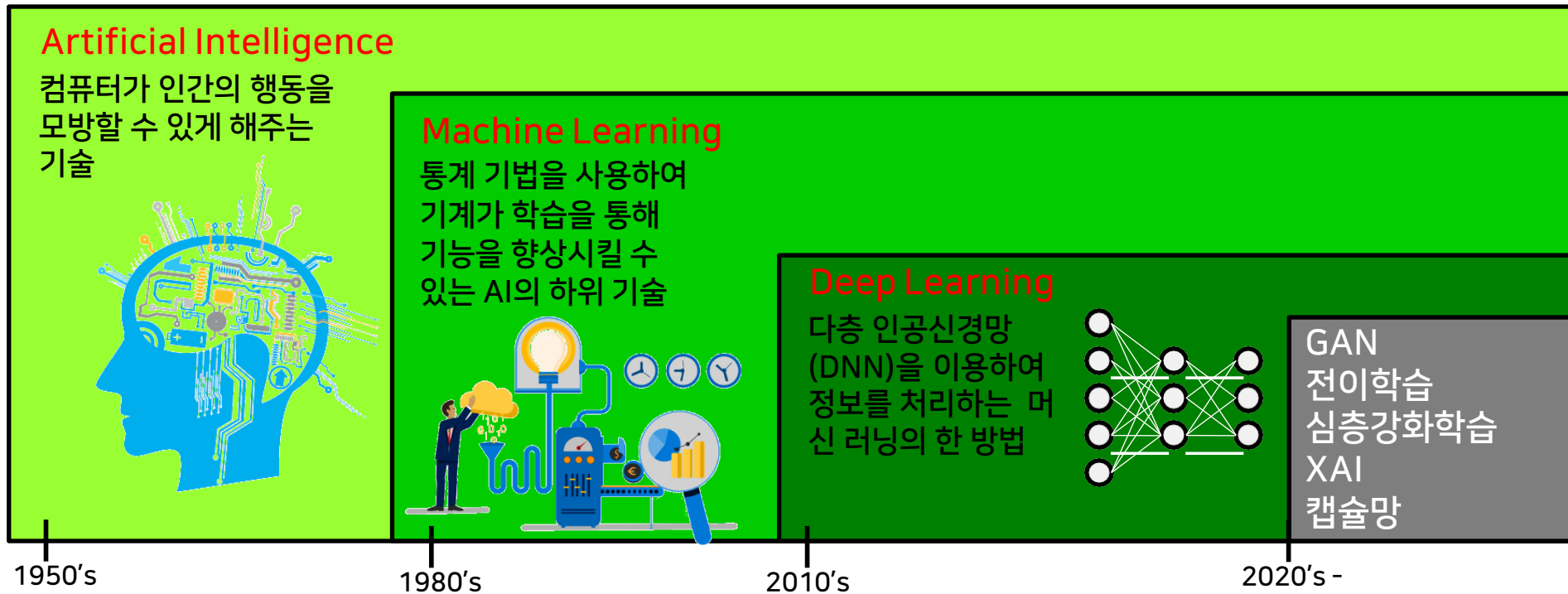
Model
(알고리즘)

RandomForest 회귀 모델을 사용

```
random_forest = RandomForestRegressor()
```

```
random_forest.fit(문제, 답)
```

```
random_forest.predict(값을 얻고 싶은데이터)
```



지도학습 (Supervised Learning)

비지도학습 (Unsupervised Learning)

강화학습 (Reinforcement Learning)

지도 학습 (Supervised Learning)

- 데이터에 대한 Label(명시적인 답)이 주어진 상태에서 컴퓨터를 학습시키는 방법
- 분류(Classification)와 회귀(Regression)로 나뉘어진다.

분류 (Classification)

- 미리 정의된 여러 클래스 레이블 중 하나를 예측하는 것
- 속성 값을 입력, 클래스 값을 출력으로 하는 모델
- 붓꽃(iris)의 세 품종 중 하나로 분류, 암 분류 등
- 이진분류, 다중 분류 등이 있다.

회귀 (Regression)

- 연속적인 숫자를 예측하는 것
- 속성 값을 입력, 연속적인 실수 값을 출력으로 하는 모델
- 어떤 사람의 교육수준, 나이, 주거지를 바탕으로 연간 소득 예측
- 예측 값의 미묘한 차이가 크게 중요하지 않다.

키	몸무게	BMI
174	96	31.70
189	87	24.35
185	110	32.14
195	104	27.35
149	61	27.47
169	103	36.06
157	110	44.62
154	111	46.86
174	90	29.72

회귀 (Regression)
(y가 수치 데이터로 구성)

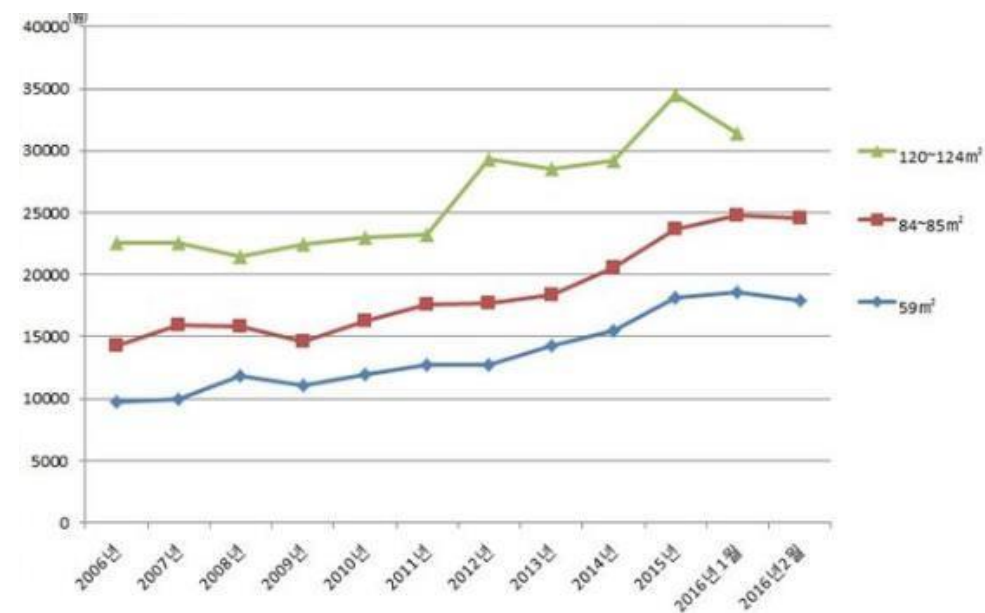
키	몸무게	비만여부
174	96	비만
189	87	정상
185	110	비만
195	104	과체중
149	61	과체중
169	103	과체중
157	110	초고도비만
154	111	초고도비만
174	90	과체중

분류 (Classification)
(y가 범주형 데이터로 구성)

지도 학습 (Supervised Learning)



스팸 메일 분류



집 가격 예측 (회귀)

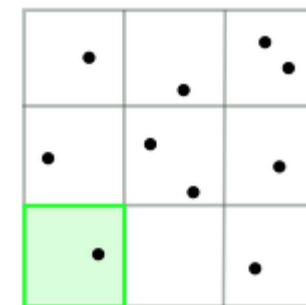
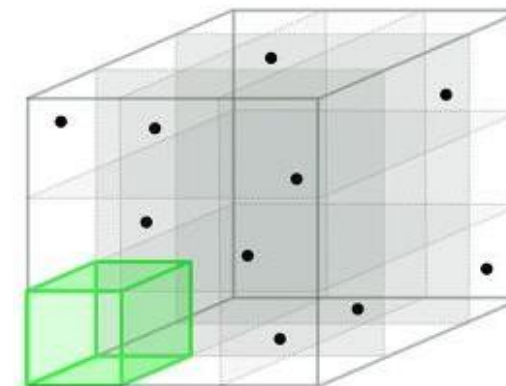
비지도 학습 (Unsupervised Learning)

- 데이터에 대한 Label(명시적인 답)이 없는 상태에서 컴퓨터를 학습시키는 방법.
- 데이터의 숨겨진 특징, 구조, 패턴을 파악하는데 사용.
- 데이터를 비슷한 특성끼리 묶는 클러스터링(Clustering)과 차원축소(Dimensionality Reduction)등이 있다.

비지도 학습 (Unsupervised Learning)

키	몸무게	Label
174	96	B
189	87	A
185	110	B
195	104	C
149	61	C
169	103	C
157	110	D
154	111	D
174	90	C

군집 (Clustering)
(y가 없는 데이터를 라벨링)



차원축소
(Dimensionality Reduction)

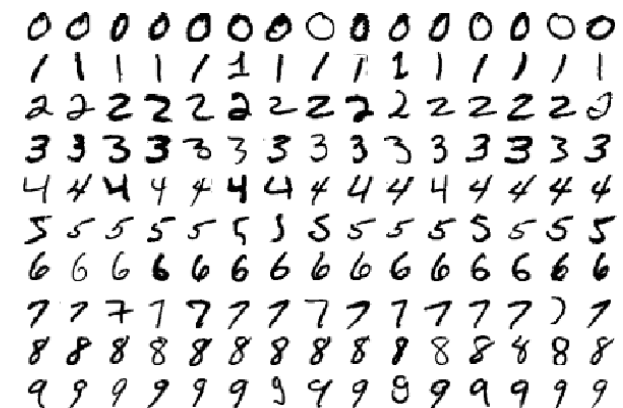
비지도 학습 (Unsupervised Learning)



이미지 감색 처리

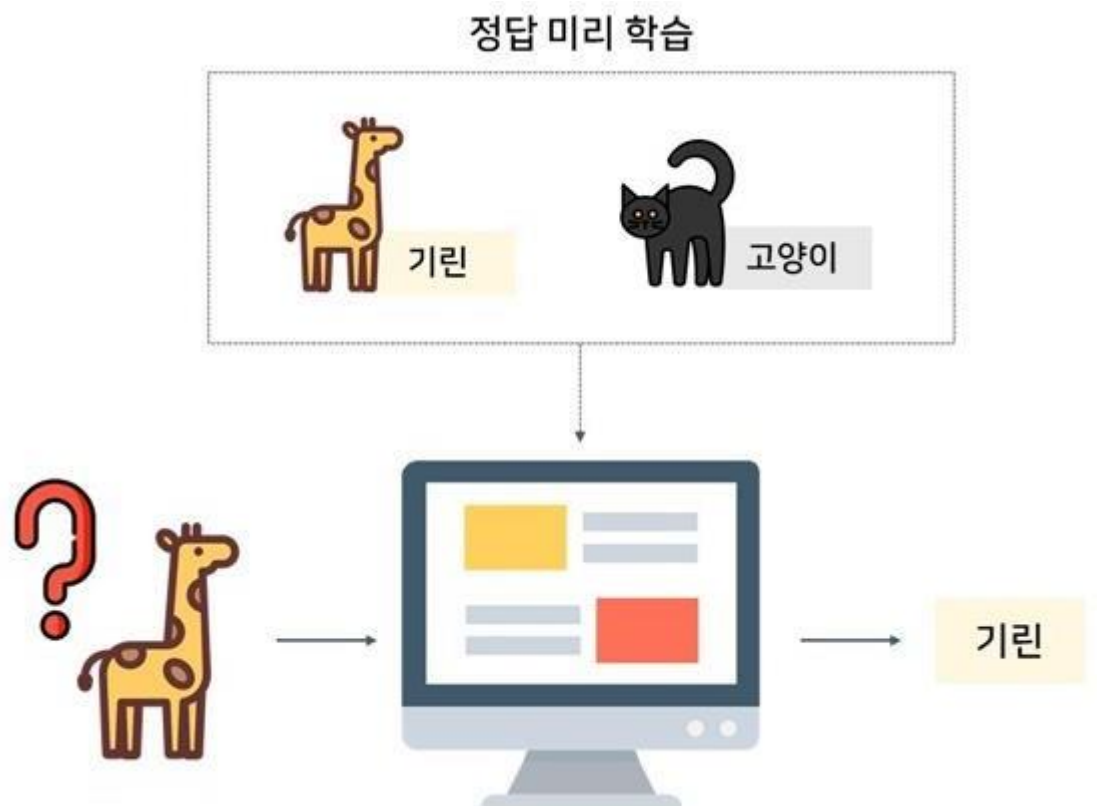


소비자 그룹 발견을
통한 마케팅



손 글씨 숫자 인식

지도 학습



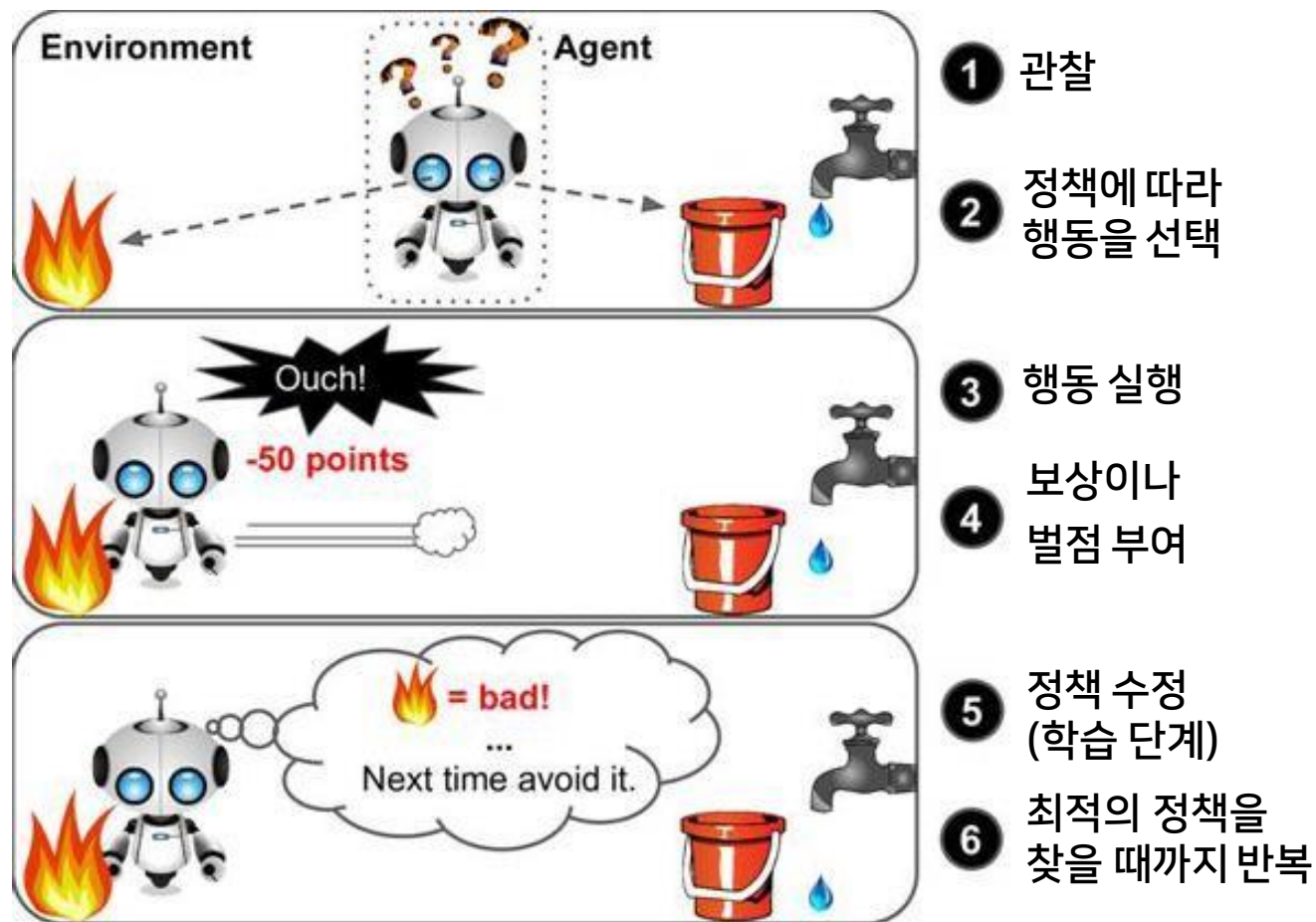
비지도 학습



강화 학습 (Reinforcement Learning)

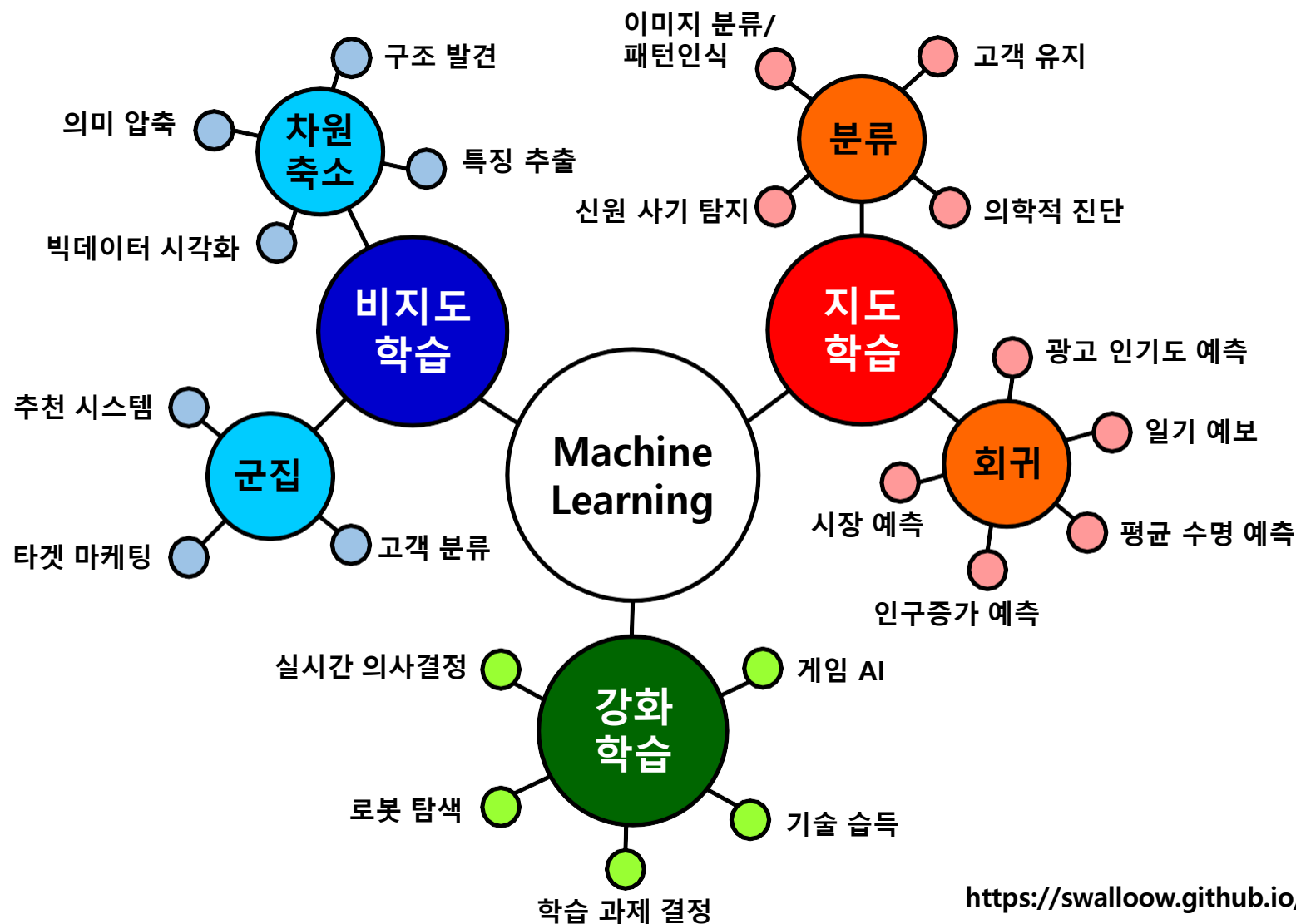
- 지도학습과 비슷하지만 완전한 답(Label)을 제공하지 않는 특징이 있다.
- 기계는 더 많은 보상을 얻을 수 있는 방향으로 행동을 학습
- 주로 게임이나 로봇을 학습시키는데 많이 사용

강화 학습 (Reinforcement Learning)



강화 학습 (Reinforcement Learning)

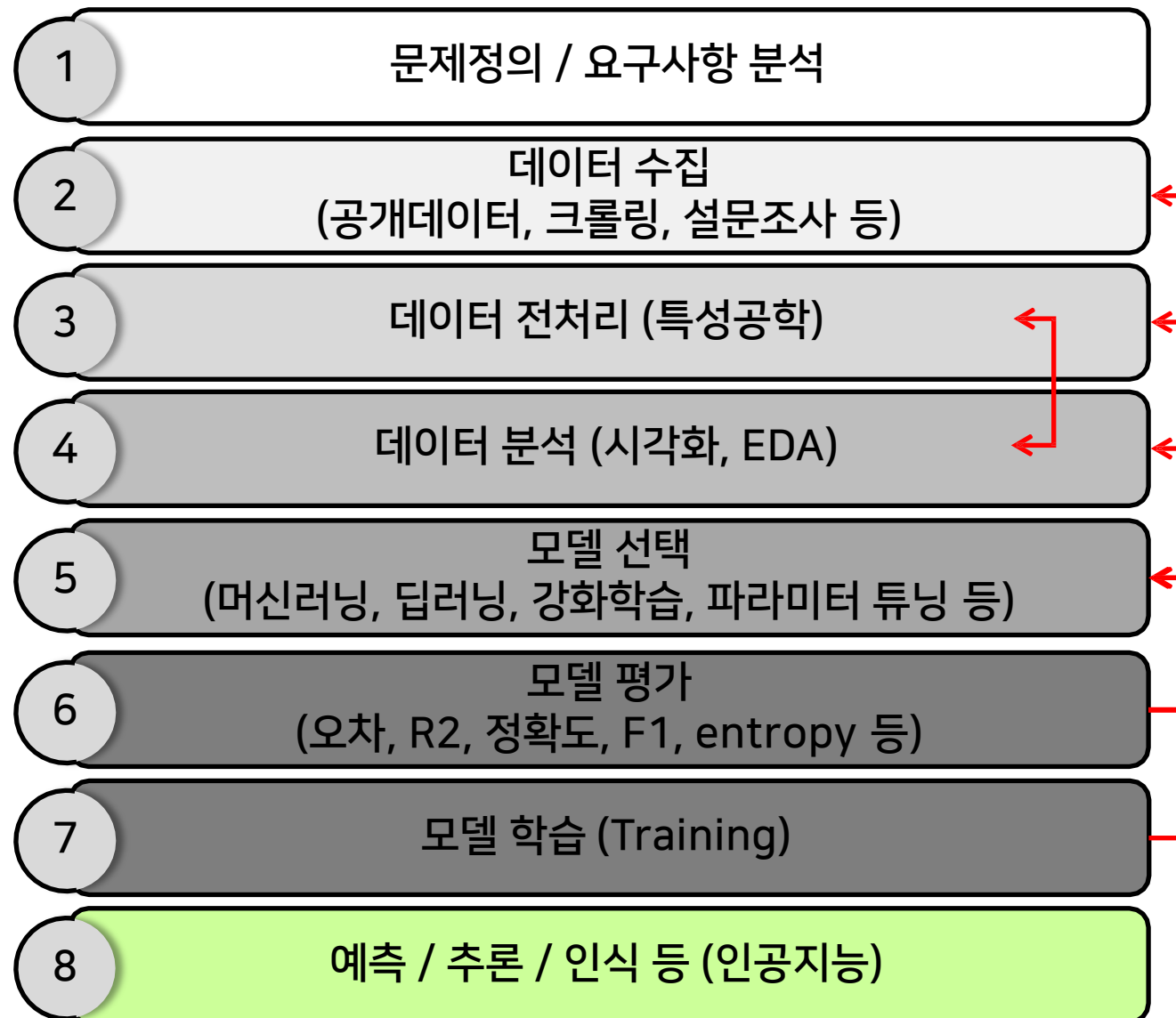




- 기존 솔루션으로는 많은 수동 조정과 규칙이 필요한 문제
- 전통적인 방식으로는 전혀 해결 방법이 없는 복잡한 문제
- 새로운 데이터에 적응해야하는 유동적인 환경
- 대량의 데이터에서 통찰을 얻어야 하는 문제

scikit-learn

- 파이썬에서 쉽게 사용할 수 있는 머신러닝 프레임워크, 라이브러리
- 회귀, 분류, 군집, 차원 축소, 특성 공학, 전처리, 교차 검증, 파이프라인 등 머신러닝에 필요한 기능을 갖춘
- 학습을 위한 샘플 데이터도 제공



1. Problem Identification(문제정의)

- 비즈니스 목적 정의
모델을 어떻게 사용해 이익을 얻을까?
- 현재 솔루션의 구성 파악
- 지도 vs 비지도 vs 강화
- 분류 vs 회귀

2. Data Collect(데이터 수집)

- File (CSV, XML, JSON)
- Database
- Web Crawler (뉴스, SNS, 블로그)
- IoT 센서를 통한 수집
- Survey

3. Data Preprocessing(데이터 전처리)

- 결측치 처리 (삭제 or 대체 (중간, 평균, 예측값 등))
- 이상치(outlier) 처리 (삭제 or 대체 (중간, 평균, 예측값, 범주화 등))
- Cleaning (오류 수정)
- Feature Engineering (특성공학)

Scaling (단위 변환)

Transform (새로운 속성 추출)

Encoding (범주형 → 수치형)

Binning (수치형 → 범주형)

Normalization (정규분포화)

범주형 데이터 통합

4. EDA (탐색적 데이터분석)

- 기술통계, 변수간 상관관계
- 시각화
pandas, matplotlib, seaborn
- Feature Selection (사용할 특성 선택)

- 단변수 시각화 : Histogram (빈도수), Boxplot(평균, 중간값 등)
- 이변수 시각화 : Scatter plot (수치, 상관관계), 누적막대그래프 (범주, 독립성 분석), 범주별 Histogram
- 다변수 시각화 : Violin plot, 3차원 그래프 등

5. Model 선택, Hyper Parameter 조정

- 목적에 맞는 적절한 모델 선택
- KNN, SVM, Linear Regression, Ridge, Lasso, Decision Tree, Random forest, CNN, RNN ...
- Hyper Parameter
model의 성능을 개선하기위해 사람이 직접 넣는 parameter

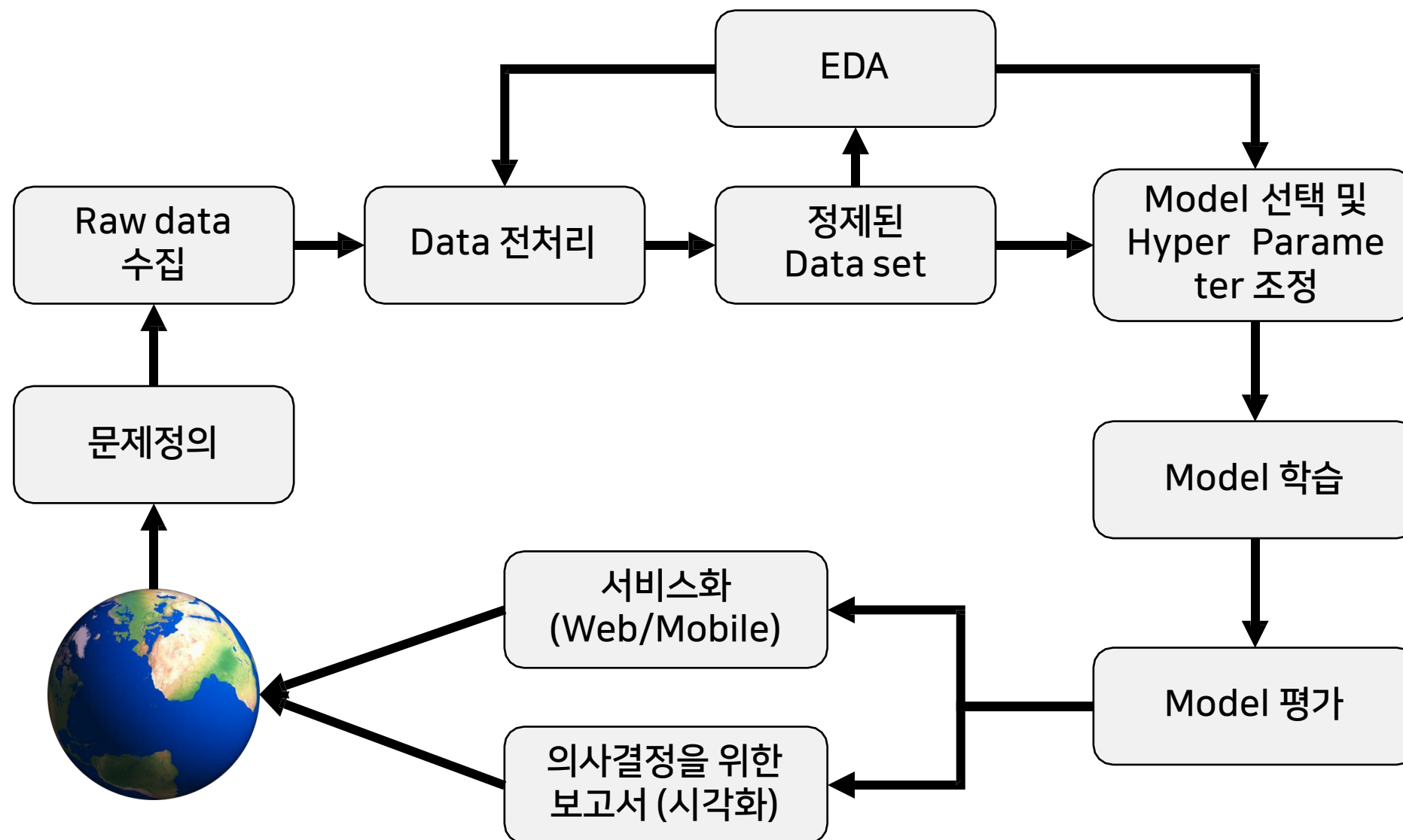
6. Model Training(학습)

- `model.fit(X_train,y_train)`
train 데이터와 test 데이터를 7:3 정도로 나눔
- `model.predict (X_test)`

	X Feature, 독립변수	Y Label, 종속변수
Train Data	7	
Test Data	3	

7. Evaluation(평가)

- accuracy(정확도)
- recall(재현율)
- precision(정밀도)
- f1 score



비만도(BMI) 데이터를 이용해 학습 해보자.

	A	B	C	D
1	Gender	Height	Weight	Label
2	Male	174	96	Obesity
3	Male	189	87	Normal
4	Female	185	110	Obesity
5	Female	195	104	Overweight
6	Male	149	61	Overweight
7	Male	189	104	Overweight
8	Male	147	92	Extreme Obesity
9	Male	154	111	Extreme Obesity
10	Male	174	90	Overweight
11	Female	169	103	Obesity
12	Male	195	81	Normal
13	Female	159	80	Obesity
14	Female	192	101	Overweight
15	Male	155	51	Normal
16	Male	191	79	Normal
17	Female	153	107	Extreme Obesity
18	Female	157	110	Extreme Obesity
19	Male	140	129	Extreme Obesity
20	Male	144	145	Extreme Obesity

회귀 문제?
분류 문제?

시각화를 통해 분류가 가능한
문제인지 확인해보자.