

# Applications of Machine Learning in the Big Data Era

Kipoong Kim

Department of Statistics, Seoul National University

August 21, 2023

# Table of Contents

- Machine Learning이란?
  - Examples
- Machine Learning의 목적
  - 예측과 추론
- Machine Learning의 종류
  - Supervised Learning (지도학습)
  - Unsupervised Learning (비지도학습)
  - Reinforcement Learning (강화학습)
  - Generative Learning (생성학습)

# Learning이란?

- 아기 고양이가 참치를 발견했다고 가정해보죠 (참치 경험X).

# Learning이란?

- 아기 고양이가 참치를 발견했다고 가정해보죠 (참치 경험X).
  - 고양이는 어떻게 행동할까요?

# Learning이란?

- 아기 고양이가 참치를 발견했다고 가정해보죠 (참치 경험X).
  - 고양이는 어떻게 행동할까요?
  - 먼저 일부분 조금만 먹어봅니다.

# Learning이란?

- 아기 고양이가 참치를 발견했다고 가정해보죠 (참치 경험X).
  - 고양이는 어떻게 행동할까요?
  - 먼저 일부분 조금만 먹어봅니다.
  - 그리고 맛있는지? 맛없는지?에 따라 >> Continue or Stop to eat !!

# Learning이란?

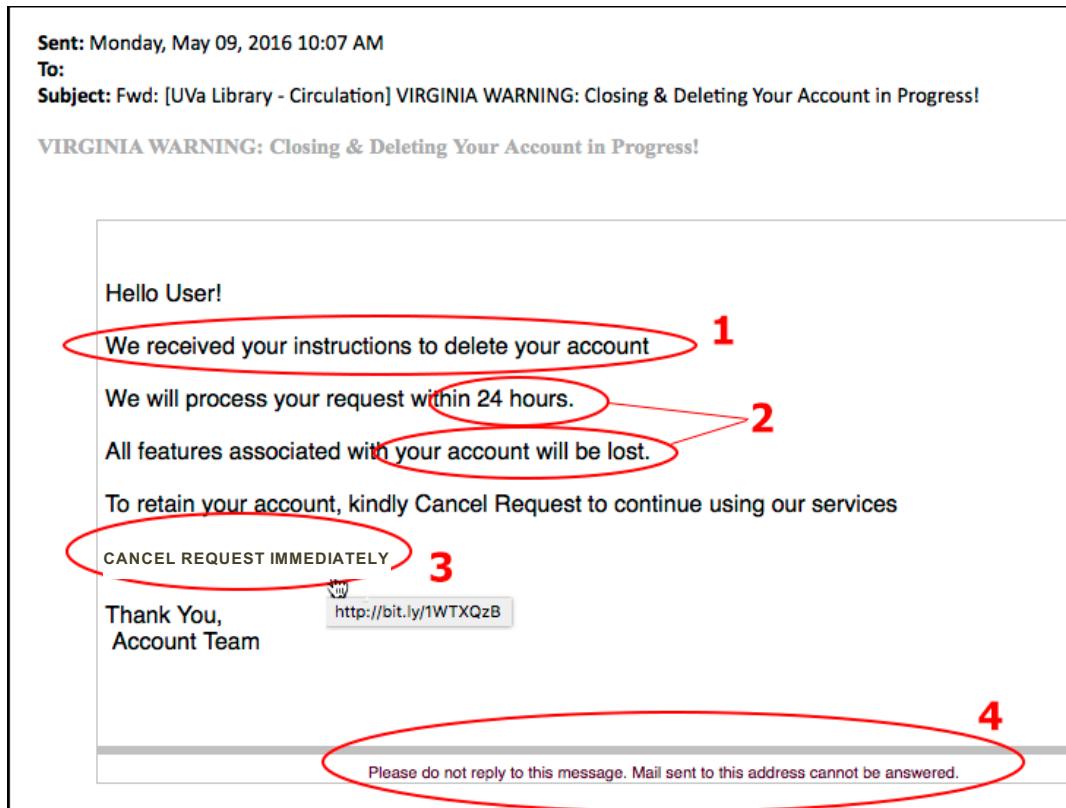
- 아기 고양이가 참치를 발견했다고 가정해보죠 (참치 경험X).
  - 고양이는 어떻게 행동할까요?
  - 먼저 일부분 조금만 먹어봅니다.
  - 그리고 맛있는지? 맛없는지?에 따라 >> Continue or Stop to eat !!
  - (여러번 먹어보고) 학습결과, 참치는 맛있는 것이다 !!

# Learning이란?

- 아기 고양이가 참치를 발견했다고 가정해보죠 (참치 경험X).
  - 고양이는 어떻게 행동할까요?
  - 먼저 일부분 조금만 먹어봅니다.
  - 그리고 맛있는지? 맛없는지?에 따라 >> Continue or Stop to eat !!
  - (여러번 먹어보고) 학습결과, 참치는 맛있는 것이다 !!
- Learning은 과거의 경험을 기반으로 이루어지는 행동의 변화(?).
  - 과거의 경험 >> 데이터
  - Machine Learning follows the same task !

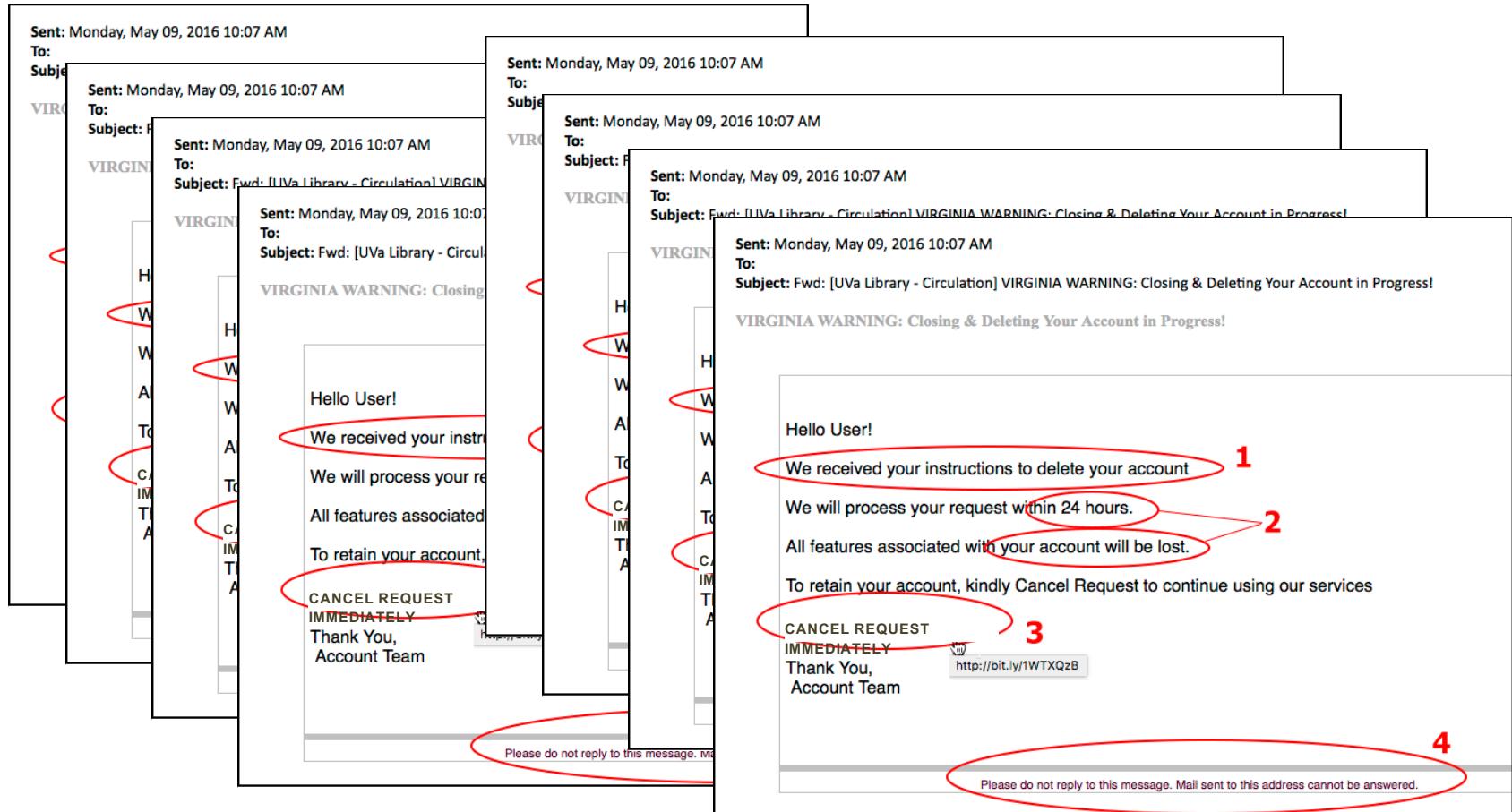
# Example: 스팸메일

- 우리들이 스팸메일을 구분할 수 있지만, 너무 시간이 많이 걸리고 귀찮기 때문에 컴퓨터 (Machine)가 대신 해주기를 원함.



# Example: 스팸메일

- 여러 정상메일 / 스팸메일로부터 패턴 파악 !!

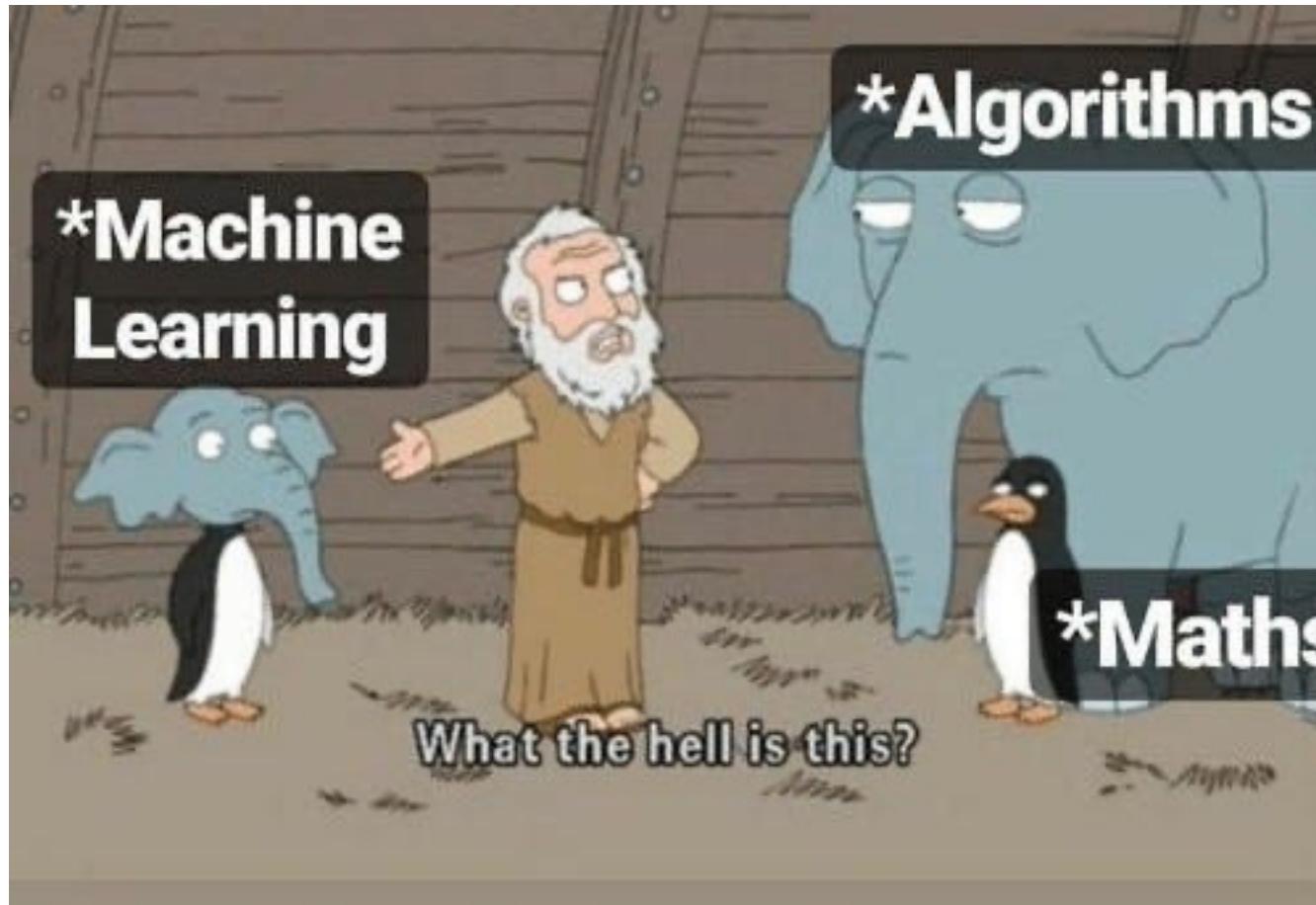


# Example: 스팸메일

- Learning rule : “learning by memorization”
  - Single word 필터링: “IMMEDIATELY”
  - Multiple word 필터링: “within 24 hours”, “account will be lost”
  - 좀더 복잡한 필터링들...
- >> 이 모든게 Machine learning 입니다.

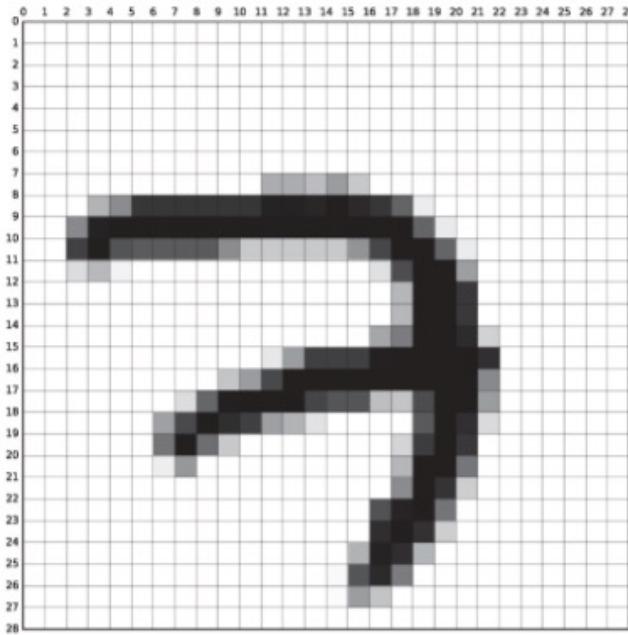
## Example: 스팸메일

- Machine learning: Algorithm + Mathematics



# Example: Handwritten Digit Recognition

- Input: 28 x 28 eight-bit grayscale (0~255)
- Output: 숫자 (0~9)
- Learning rule: 가운데 열이 모두 검정색이다 >> “숫자 1”일 것이다!



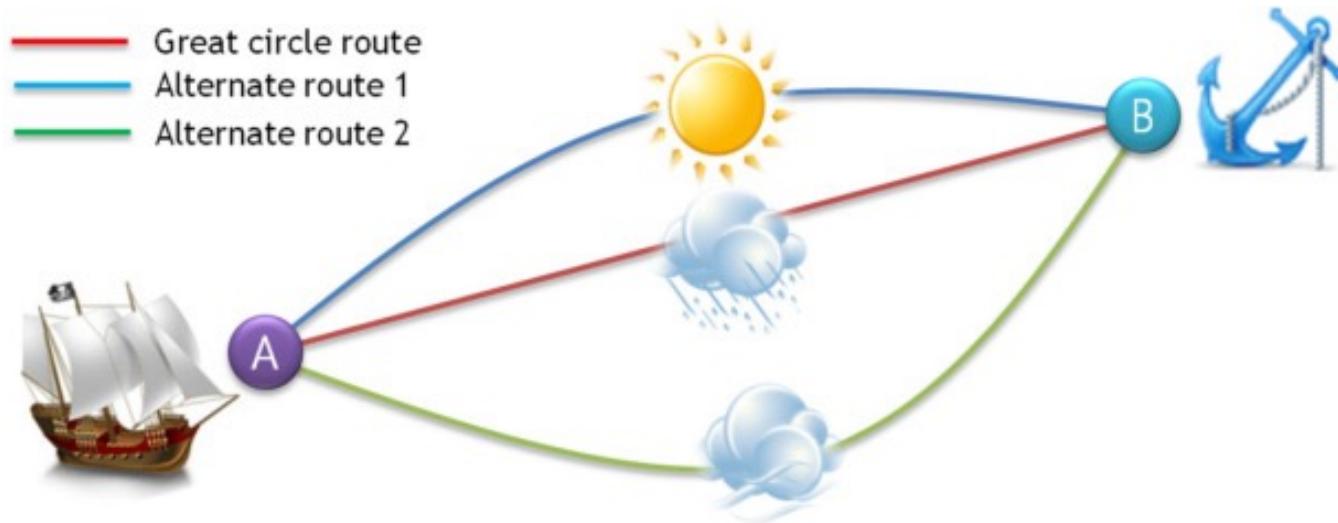
(a) MNIST sample belonging to the digit '7'.



(b) 100 samples from the MNIST training set.

# Example: Ship route optimization

- 목표: 항해에 필요한 연료를 최소화시키는 것
- 어떤 경로가 베스트일까요?



- Input: 이동거리, 바람의 방향, 파도의 높이 / 주기 / 세기
- Output: 소모된 연료

# Example: Web advertisement

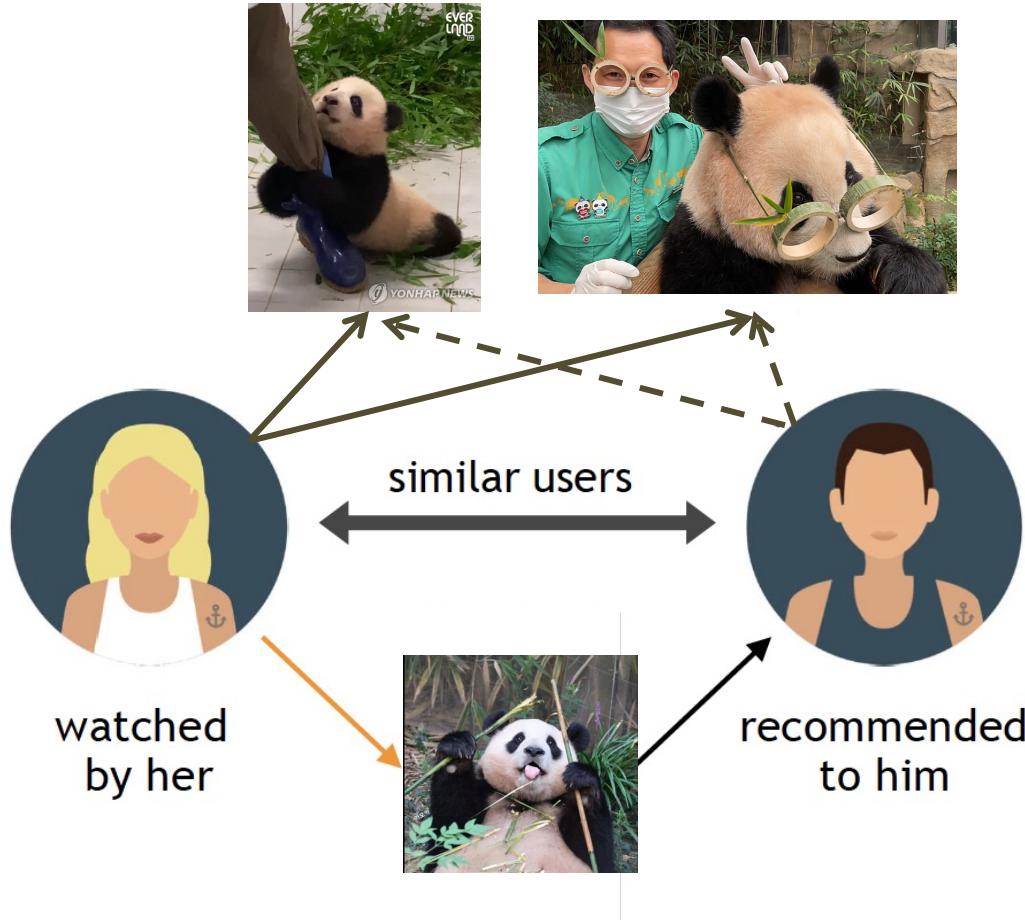
- “인스타그램” 추천시스템
  - 목표: 관심 게시글 추천
  - Input:
    - user info: male & 1991
    - click history: Fubao & MBTI
  - Output: 관심 게시글 상위 N개
  - Learning rule:

## Collaborative filtering



# Example: Web advertisement

- **Collaborative Filtering**



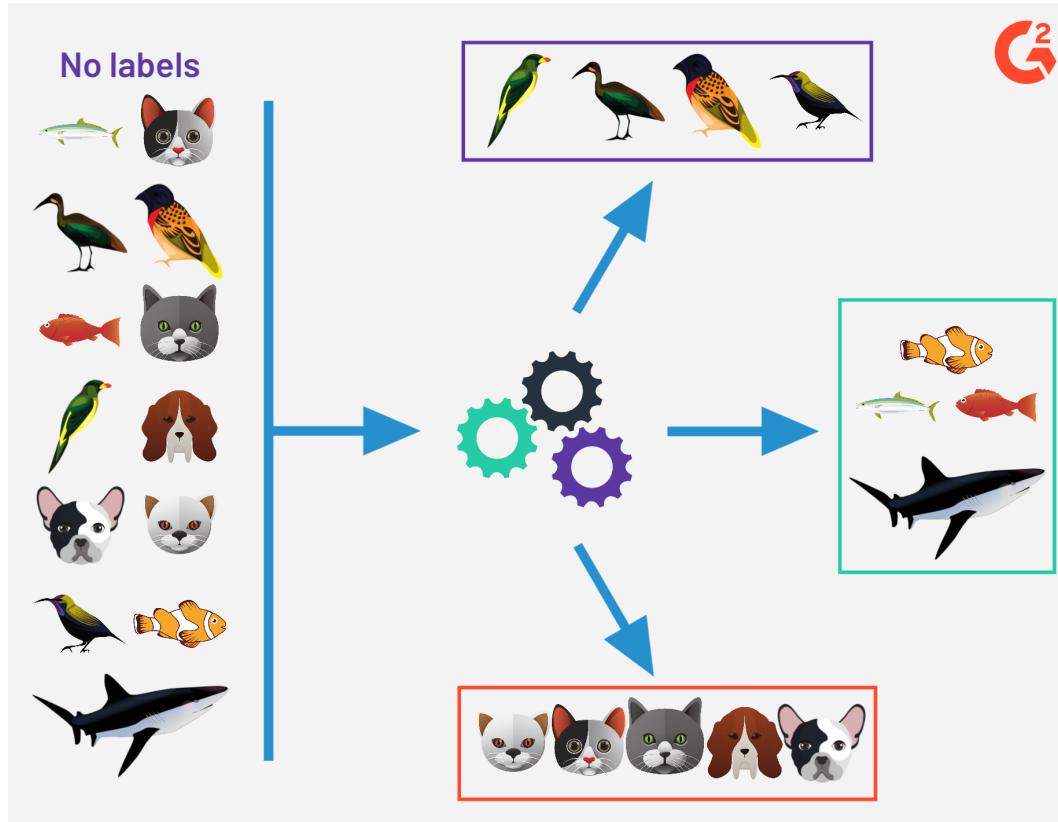
# ML:: Supervised learning

- 풀고자 하는 문제에서 정답 (output)이 존재하는 것.

Data	Objective (output)	Application
Email	Spam or not (0 / 1)	Spam filtering
Ad, user info	Click? (0 / 1)	Online advertising
Ship route	Fuel consumed	Fuel optimization
Restaurant reviews	Sentiment (pos. / neg.)	Reputation monitoring
Image, radar info	Position of other cars	Self-driving car

# ML:: Unsupervised learning

- 정답이 존재하지 않는 경우 예시: Clustering
  - 다양한 기준: 날 수 있느냐? / 얼룩무늬가 존재하느냐? / 알을 낳느냐?



# Supervised Learning in Biomedical research

- Electronic Health Records (EHR) 자료 (그냥 엑셀 자료)

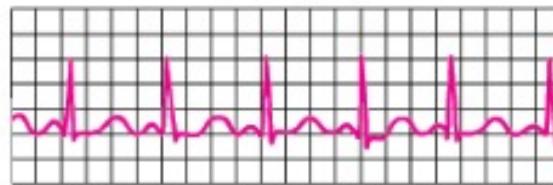
Disease  
Normal

20_191223_XXXXX									
Time	HR	NIBP_S	ETCO2	CO2	BIS/BIS	NMT_TOF_CNT	.....	PROPOFOL_CE	EVENT
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
00:06:59	95	110	1.2	2.4	78	2	.....	3.135	-
00:07:00	95	122	1.2	2.4	77	2	.....	3.135	-
00:07:01	98	122	2.8	0	69	0	.....	3.155	-
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
00:07:59	96	122	3	0	65	0	.....	3.38	-
00:08:00	96	124	3	0	68	0	.....	3.38	-
00:08:01	97	124	3.4	3.7	67	1	.....	3.459	-
00:08:02	97	124	3.4	3.7	68	1	.....	3.459	intubation
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
00:12:00	101	141	4.9	4.4	80	3	.....	4.04	-
00:12:01	100	141	4.1	4.6	81	1	.....	4.067	-
00:12:02	100	141	4.1	4.6	78	1	.....	4.067	-
00:12:03	100	141	4.1	4.6	74	1	.....	4.067	-
00:12:04	100	141	3.2	0	68	0	.....	4.436	-
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

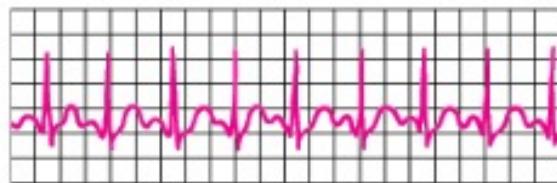
# Supervised Learning in Biomedical research

- Atrial Fibrillation (AF) 예측
- 아래 정상 vs 비정상 심박 간에는 어떤 패턴 차이가 존재하나요?
  - 그 패턴을 구분하는 규칙을 찾을 수 있나요?

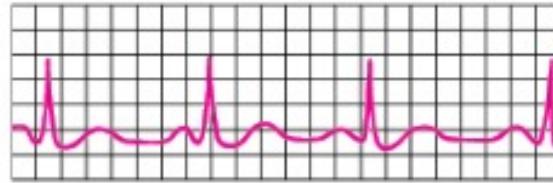
정상 심박



빠른 심박



느린 심박



불규칙한 심박



# Supervised Learning in Biomedical research

- Doctor-Patient Question

## Patient's description (Input)

John은 58세의 남성으로 지난 6개월 동안 호흡곤란이 심해졌다고 호소하며 오늘 진료실을 찾아왔습니다. 그는 특히 계단을 오르거나 장거리를 걷다가 숨을 멈추고 숨을 고르지 않고는 걷기가 어렵다고 말했습니다. 또한 이러한 증상은 종종 가슴이 답답하고 때때로 어지러움을 동반한다고 언급했습니다. 생활 방식에 대한 질문에 John은 지난 35년 동안 하루 평균 한 갑 정도의 흡연을 해왔다고 답했습니다. 가끔 술을 즐기고 육류위주의 식단으로 채소는 적게 먹는 식습관을 가지고 있다고 인정했습니다. John의 아버지는 60대 초반에 뇌졸중을 앓았고, 누나는 고혈압 진단을 받았습니다.

## Diagnosis (Output)

심혈관 질환, 관상동맥 질환 (CAD)

# ML에서 다룰 수 있는 데이터 종류

- 정형 (엑셀) 데이터: 항해경로 최적화, EHR (엑셀)
- 시계열 데이터: 심전도 (ECG)
- 이미지 데이터: 손글씨, 인스타그램 게시물 추천
- 텍스트 데이터: 이메일, 환자증상 텍스트
- >> 거의 대부분의 데이터

# ML을 요약하자면,

- 주어진 문제를 해결하기 위한 규칙을 세우는 것인데,  
인간이 할 수 있는 건 대부분 할 수 있음.



# ML이 못 하는건

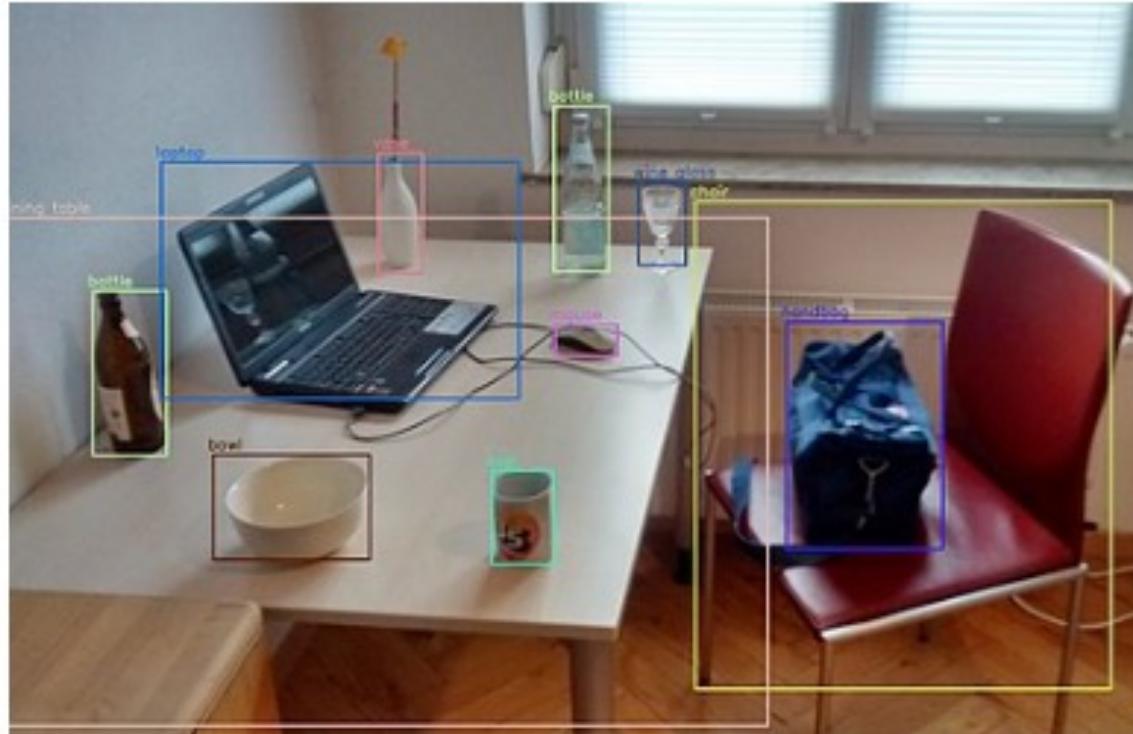
- 인간이 못하는 건 대부분 못하는 경향이 있음.

A grid of handwritten digits from 0 to 9, arranged in four rows. Some digits are highlighted by colored boxes: a green box surrounds the first five digits in the second row; a red box surrounds the last digit in the fourth row; a red box surrounds the first digit in the fifth row; and a red box surrounds the second digit in the seventh row.

0	0	0	1	1	1	1	1	2
2	4	2	4	3	2	3	3	3
3	4	4	4	4	5	5	5	5
6	6	7	7	7	7	8	8	8
8	8	9	8	9	4	9	9	9

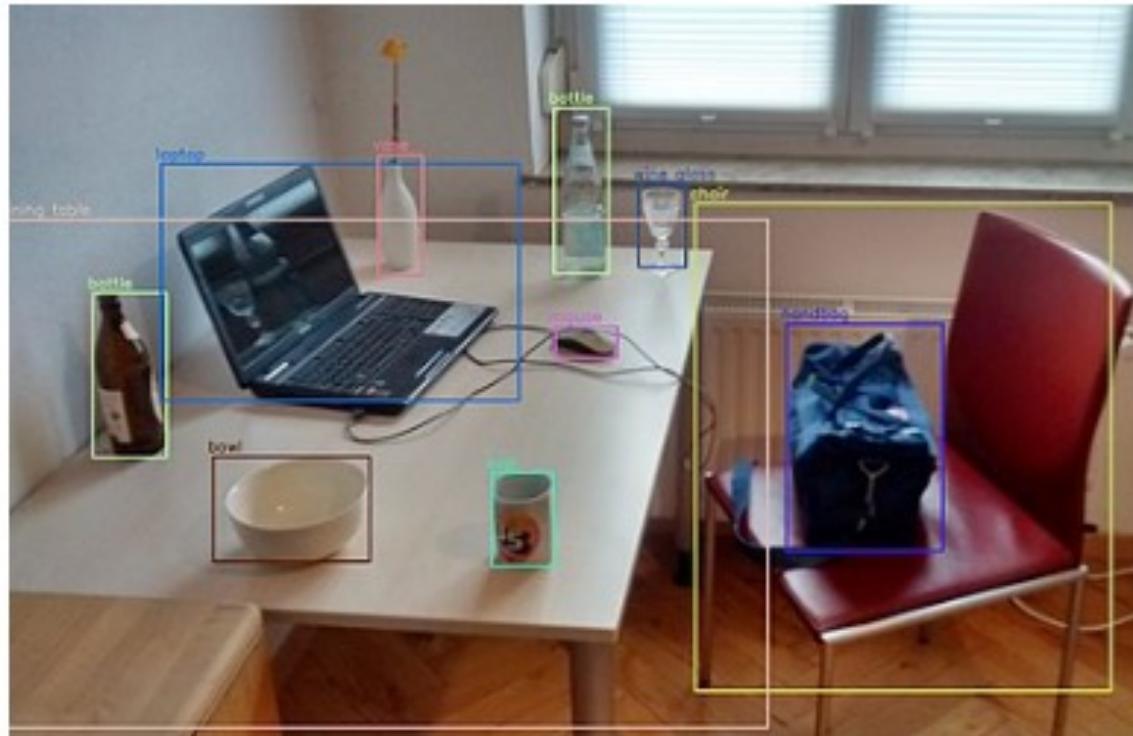
# 좀 더 복잡한 문제

- 여러분들은 아래 사물들을 구분 할 수 있나요?



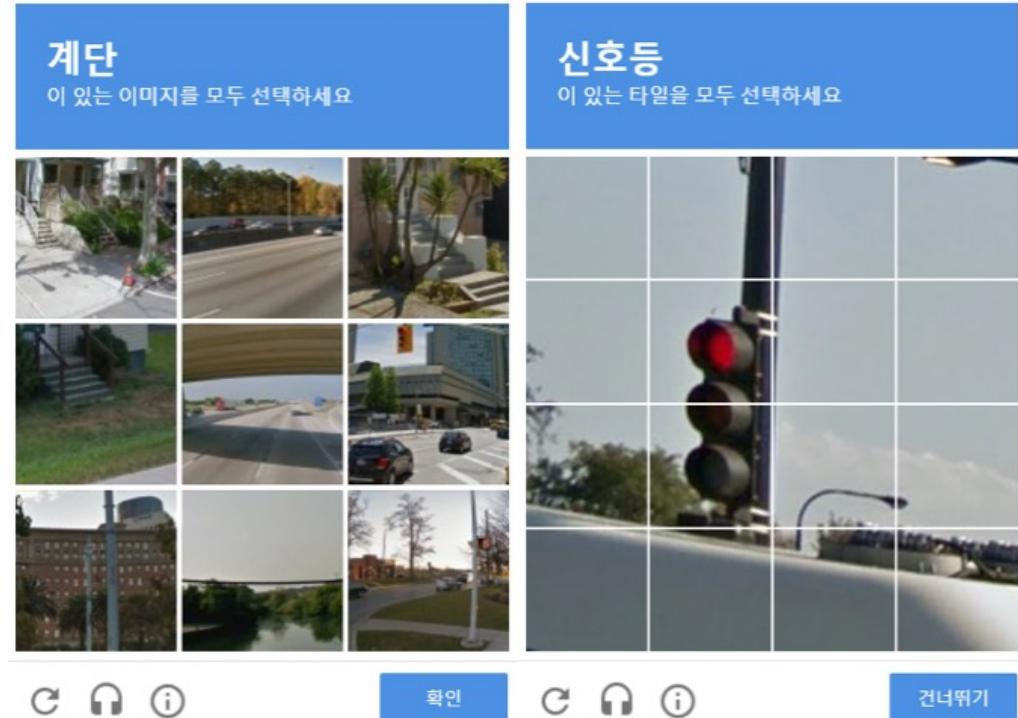
# 좀 더 복잡한 문제

- 여러분들은 아래 사물들을 구분 할 수 있나요?
  - 그럼 Machine Learning도 할 수 있습니다 (정답 데이터만 있다면).



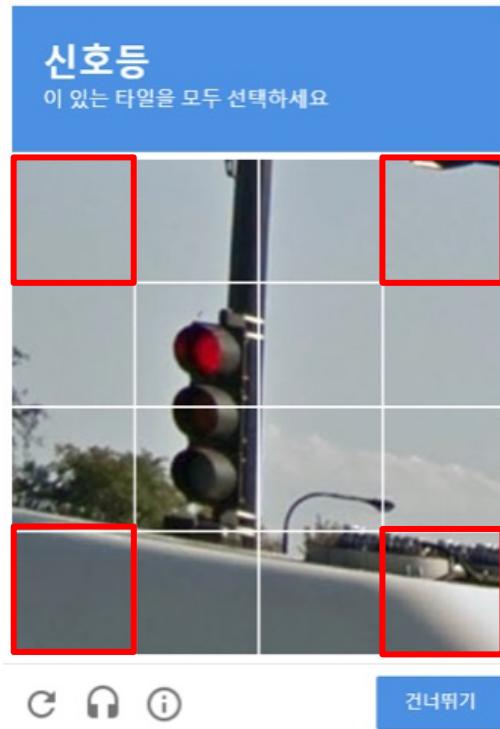
# 데이터 레이블링

- 정답 데이터는 어떻게 모으나요?
  - 사람들이 직접 만듭니다... (예, 구글 reCAPTCHA)



# 데이터 레이블링

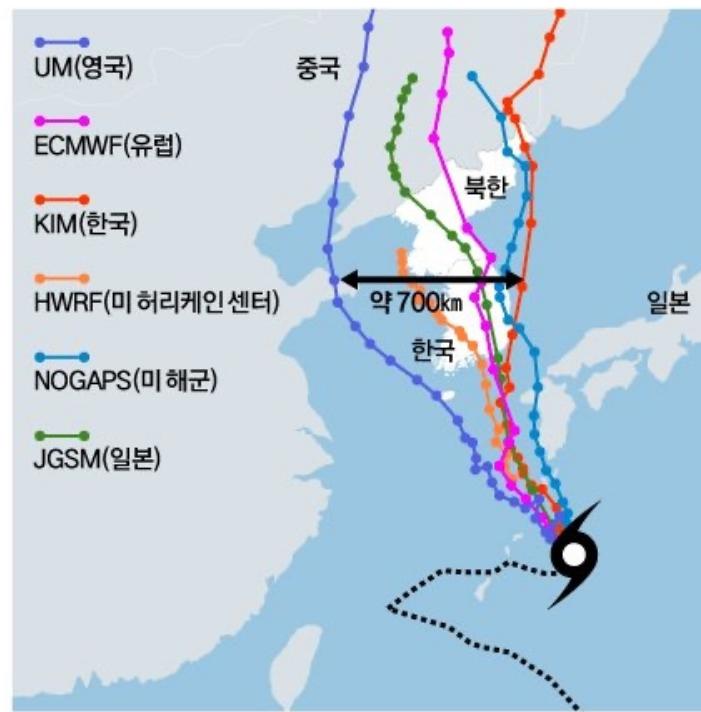
- 잘못 틀리면 다시 하라고 하던데, 정답 알고 있는거 아니에요?
  - 다른 사람들이 클릭한거랑 너무 달라서...



# 어떤 ML 예측모형을 골라야 할까?

- 수집된 데이터, 학습 규칙에 따라 다양한 ML 모형 가능.
- 태풍 “카누” 예시: 어느 나라 모형이 가장 좋을까?

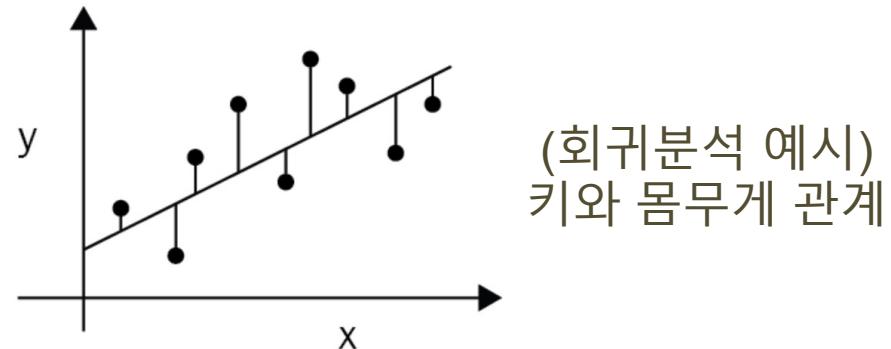
6일 전 세계 주요 수퍼컴이 계산한  
태풍 ‘카누’의 수치예보 모델



# 어떤 ML 예측모형을 골라야 할까?

- 평가지표

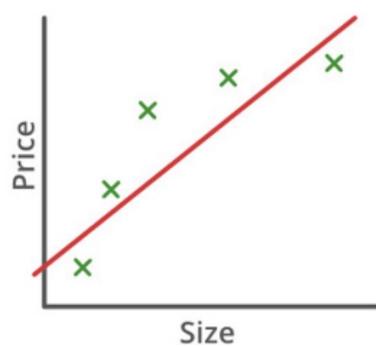
예:  $|실제값 - 예측값|$  합계



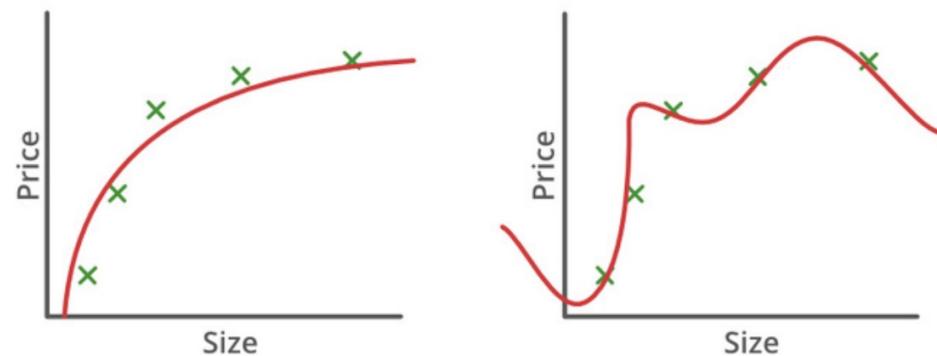
>> 단순히 평가지표만 살펴보면 될까?

- 아래에서 어느 모형이 가장 좋을까요?

단순한 모형



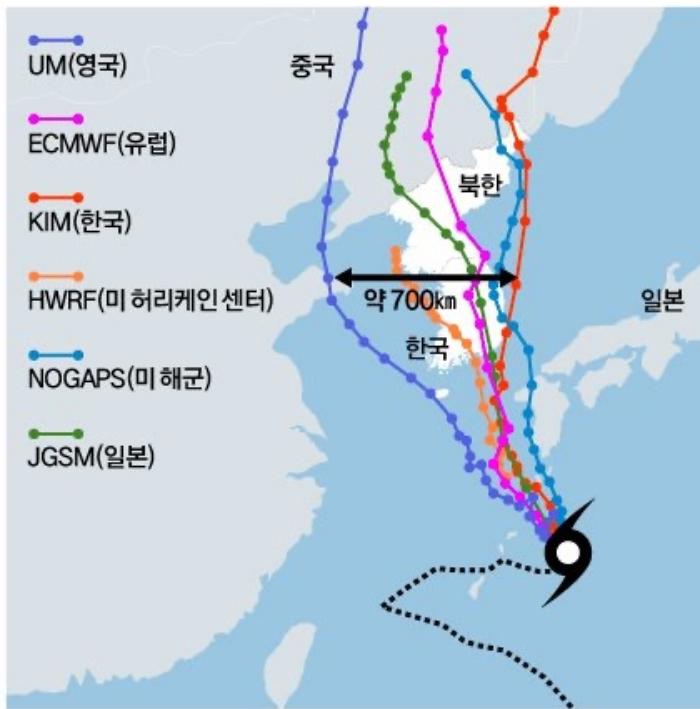
복잡한 모형



# 어떤 ML 예측모형을 골라야 할까?

- 뉴스만 봐서는 어떤게 좋은지 모르겠네요.  
실제 vs 예상 경로가 필요한데...

## 6일 전 세계 주요 수퍼컴이 계산한 태풍 '카누'의 수치예보 모델

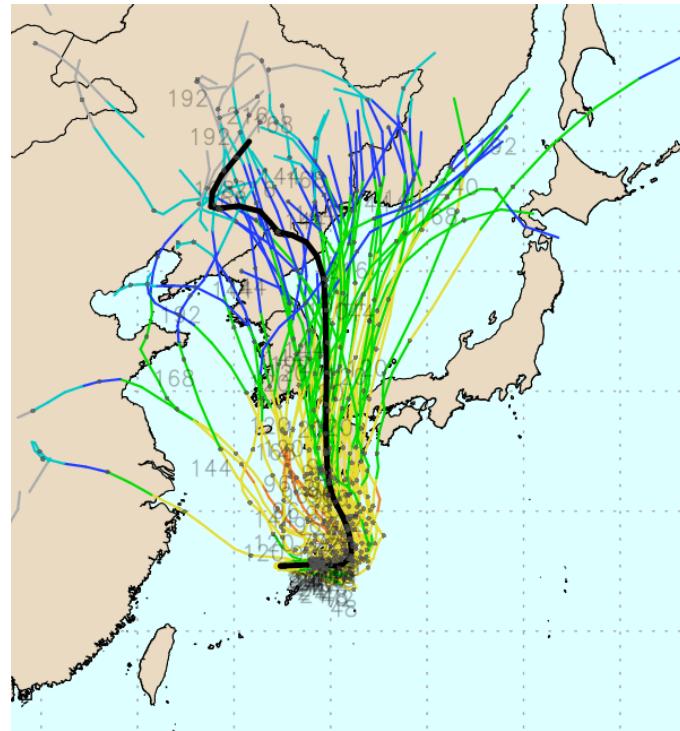


## 모형을 평가하기 위한 실제경로와 예상경로 차이



# 어떤 ML 예측모형을 골라야할까?

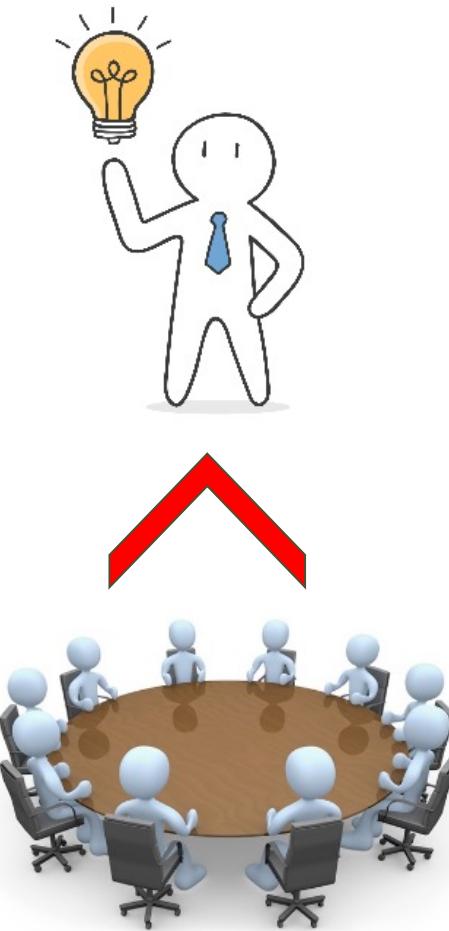
- 뉴스만 봐서는 어떤게 좋은지 모르겠네요. >> 양상블 !!
  - 양상블 (프랑스어: ensemble): 전체적 어울림 조화. 합주
  - >> 통계학에서는 여러 모형들을 종합한다는 뜻



# 어떤 ML 예측모형을 골라야 할까?

- 앙상블 (기자가 ML을 알았더라면...)

6일 전 세계 주요 수퍼컴이 계산한  
태풍 '카누'의 수치예보 모델



# 어떤 ML 예측모형을 골라야 할까?

- 양상블: 왜 더 좋을까?
  - 특정 문제에 대해 25명이 다수결로 정답을 맞춘다고 가정하자.
  - 각 사람이 정답을 틀릴 확률 (error rate)은 35%이고, 서로 독립이다.  
(독립: 한명 틀렸다고 다른 사람이 더/덜 틀리지 않는다)
  - 다수결로 정답을 틀릴 확률은 다음과 같다:  
“25명 중 13명 이상이 틀릴 확률” = 6%
$$\sum_{i=13}^{25} \binom{25}{i} 0.35^i (1 - 0.35)^{25-i} = 0.06$$
  - 다수결이 한명보다 틀릴 위험이 줄어든다 !!

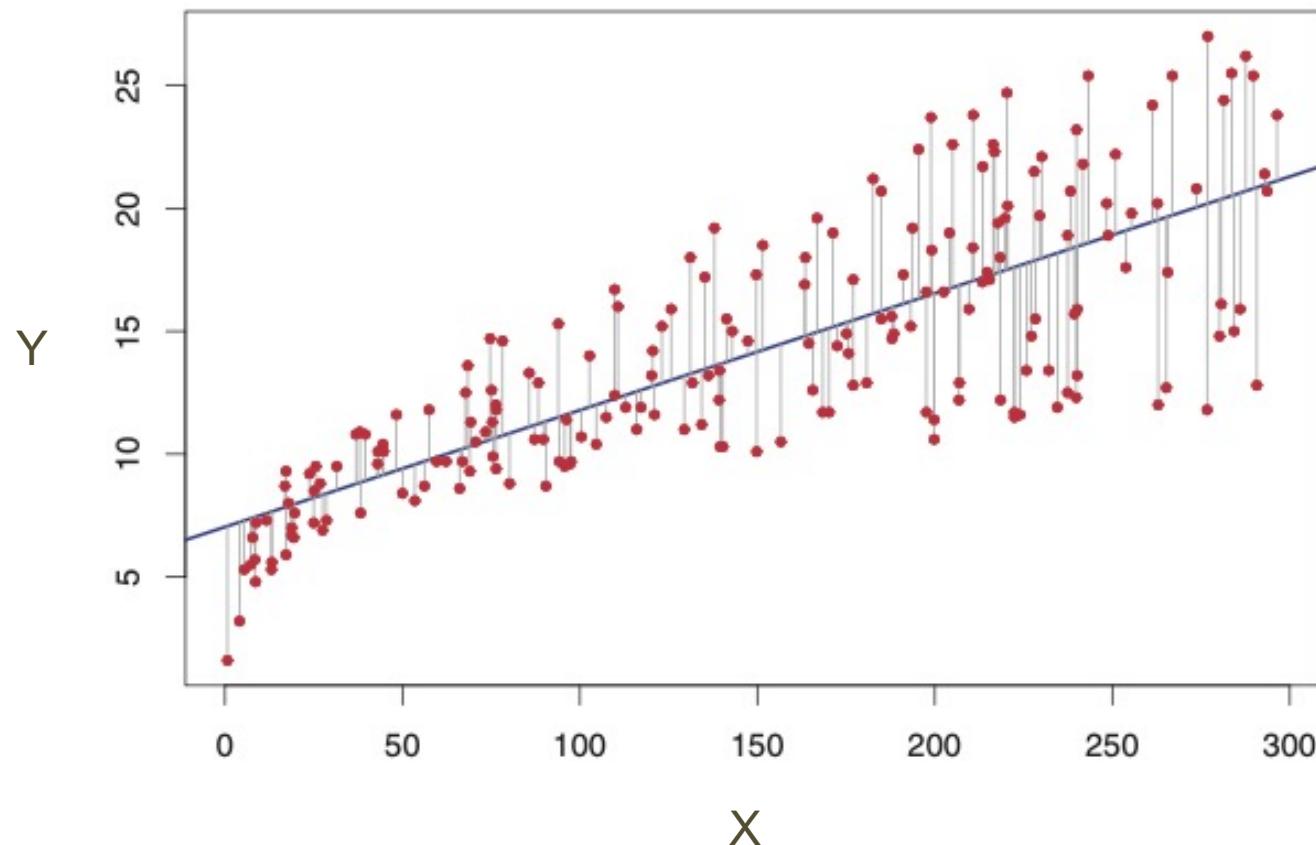


# 통계적 추론 (Statistical Inference)

- 이전까지의 문제는 대부분 예측하는 문제였지만,
    - 스팸메일을 분류하거나
    - 숫자를 맞추거나,
    - 최적경로를 예측하거나,
    - 게시글을 추천해주는 등
  - 예측을 잘하는 것은 알겠는데 왜 그런지 이유에 대해 알고싶다!
- >> 추론 (Inference)
- 대표적 예시: Regression ( $X$ 와  $Y$  간 관계성을 찾는 것)

# Linear Regression

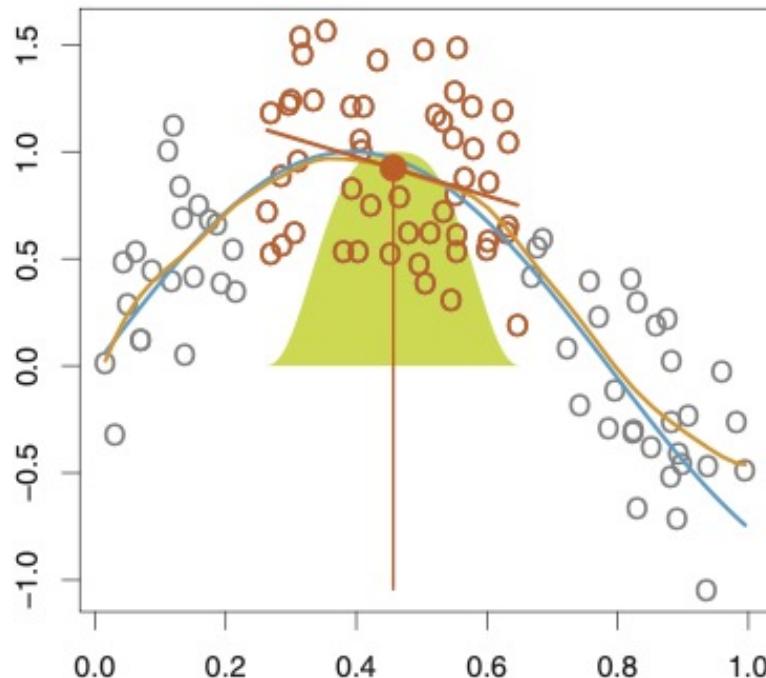
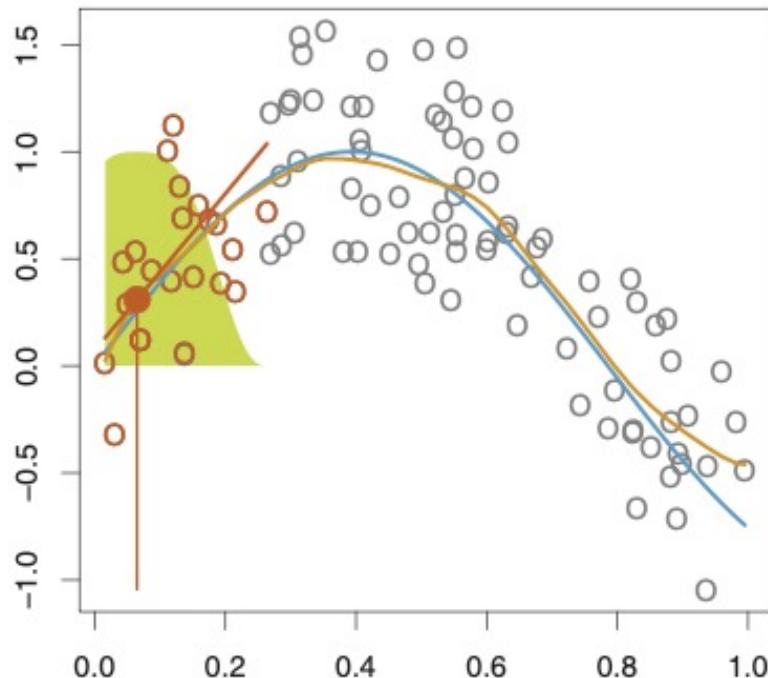
선형적 관계



# Local Regression

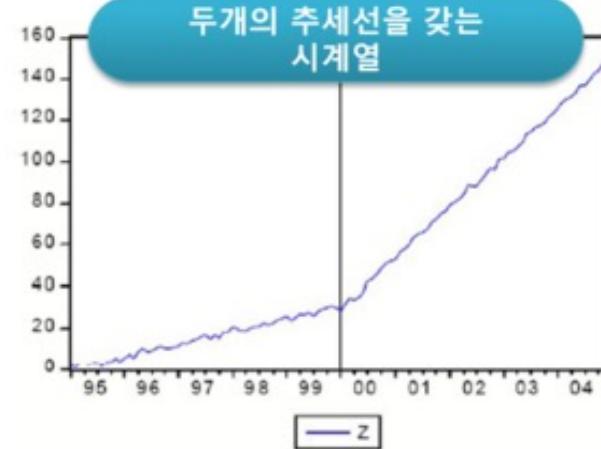
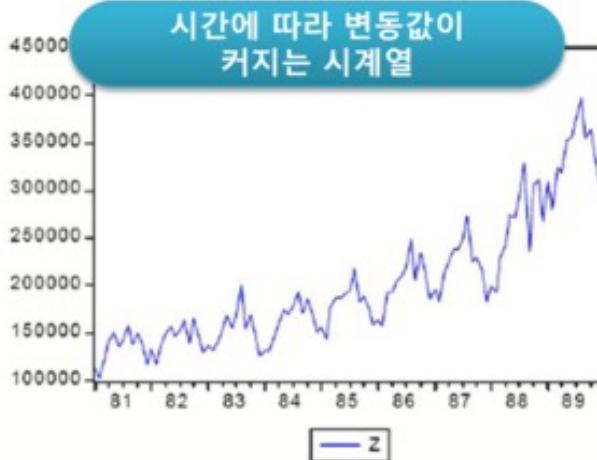
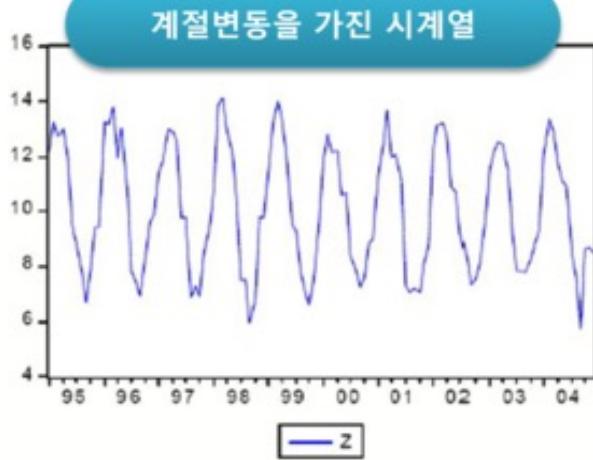
비선형적 관계

Local Regression



# Time series

시간에 따른 패턴



# Decision Tree (like 스무고개)

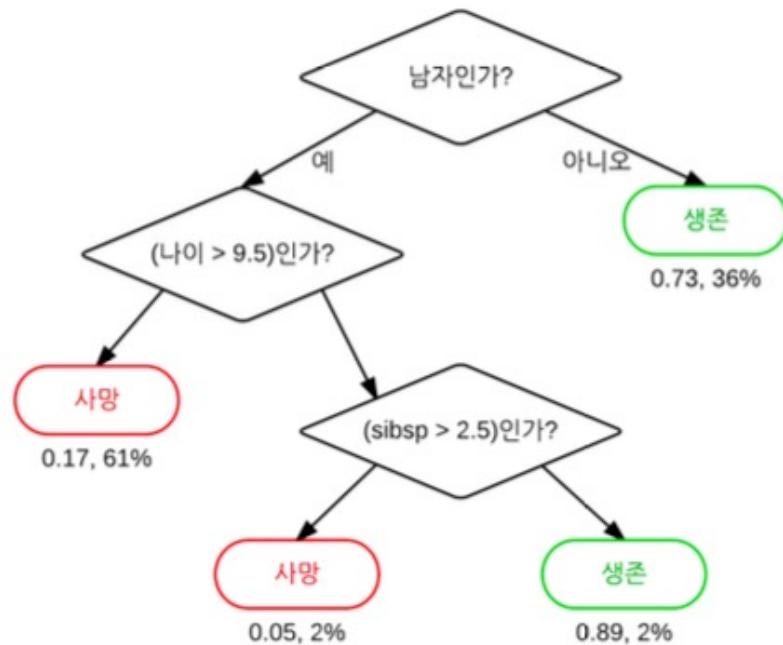
- 타이타닉 생존자 데이터:
  - Input: 탑승객의 특성 (나이, 성별, 동반자의 수 등)
  - Output: 생존여부

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Nan	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Nan	S
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Nan	S
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	Nan	S
887	888	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	female	Nan	1	2	W.C. 6607	23.4500	Nan	S
889	890	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	Nan	Q

891 rows × 12 columns

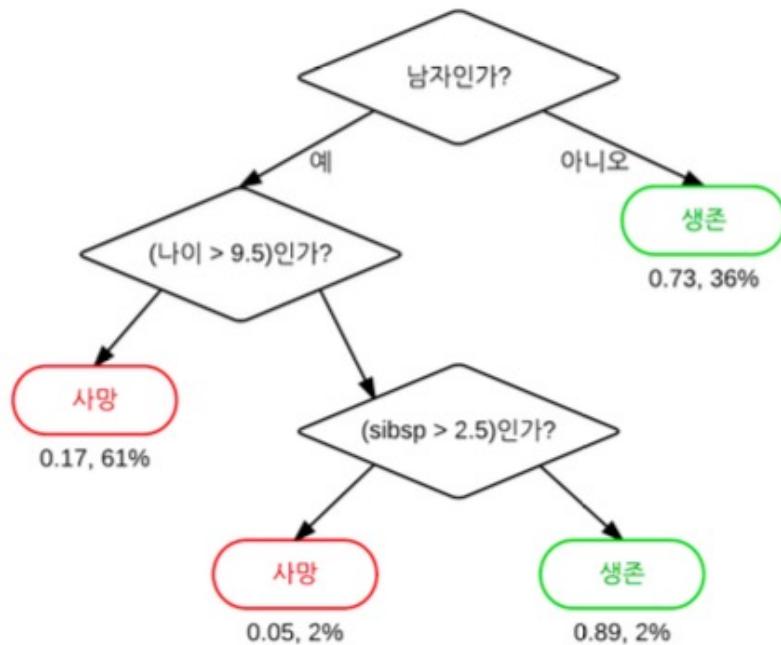
# Decision Tree (like 스무고개)

- 타이타닉 생존자 데이터:
  - Input: 탑승객의 특성 (나이, 성별, 동반자의 수 등)
  - Output: 생존여부
- 다음과 같은 생존예측 규칙 (learning rule)을 만들 수 있습니다.



# Decision Tree (like 스무고개)

- 타이타닉 생존자 데이터:
  - Input: 탑승객의 특성 (나이, 성별, 동반자의 수 등)
  - Output: 생존여부
- 다음과 같은 생존예측 규칙 (learning rule)을 만들 수 있습니다.



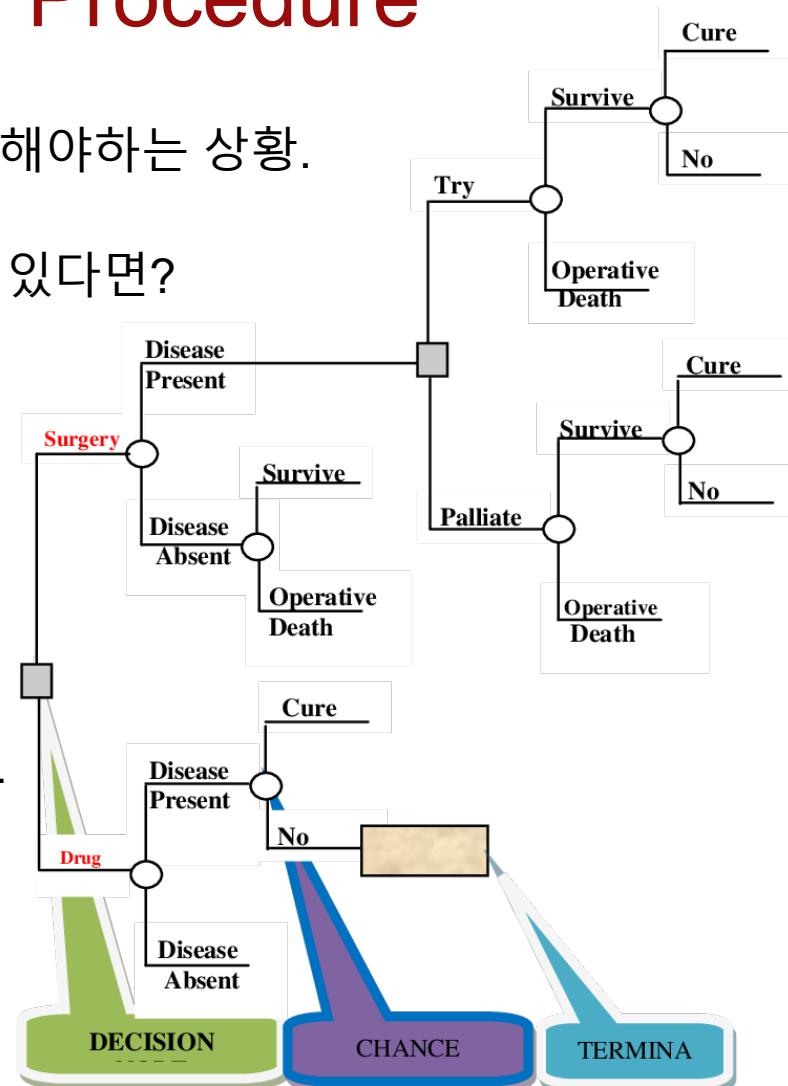
## Inference

- 어떻게 생존하였는가?
  - 여성
  - 어린 아이
  - 동반자 적을 때

>> 약자 우선 대피

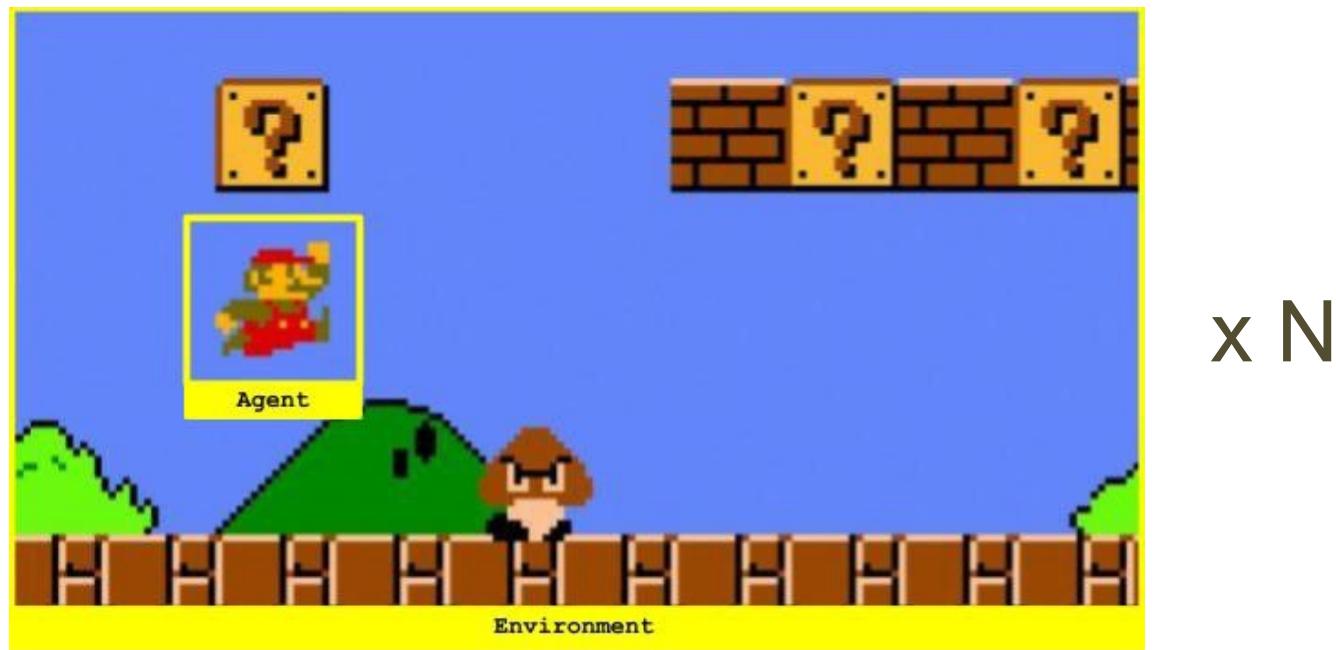
# Decision Tree in Surgical Procedure

- 검사항목을 바탕으로 환자의 수술을 결정해야하는 상황.
- 환자 데이터로 만들어진 Decision Tree가 있다면?
  - 56세 이상이면 수술 X
  - 혈압이 140 이상이면 수술 O
- 이분법적으로 의사결정할때 유용함.
  - (회귀분석)  
나이 증가에 따라 수술 위험도 올라감 (끝).  
>> 그럼 몇살부터?
  - (Decision Tree)  
56세 이상부터 수술 위험도 증가함.  
>> 56세부터는 수술 X



# ML:: Reinforcement learning

- 일단 여러번 해보고 보상을 최대화하는 쪽으로 움직이자 !
  - 인간의 학습과 유사함 !
  - 여러번 해보니까 첫번째 블록을 깨는게 좋더라 !!  
>> 이후부터는 첫번째 블록은 깨면서 다른 시도를 거듭함.



# ML:: Reinforcement learning

- 로보틱스 / 자율주행 등의 분야에서 주로 사용되고 있습니다.

Reinforcement Learning (e.g.Waymo) vs Imitation Learning (e.g.Tesla)



# ML:: Generative learning

## ■ ChatGPT 4

2023년 8월 21일  
대구한의대학교에서  
Application of Machine  
learning in the big data  
era라는 주제로 발표를  
했는데 관련해서 네이버  
블로그 글을 인스타  
감성으로 작성해줘.

🌟 2023년 8월 21일, 대구한의대학교에서의 특별한 날 🌟

안녕하세요, [당신의 이름]입니다! 🌸 오늘은 제게 정말 의미 있는 날이었어요.

대구한의대학교에서 'Application of Machine Learning in the Big Data Era'라는 주제로  
발표를 마쳤습니다. 🎉🌟

📸 [발표 중인 사진 또는 대구한의대학교에서 찍은 사진]

빅데이터 시대에 머신러닝의 적용은 우리의 일상, 건강, 연구에 어떤 영향을 미치는지에  
대해 깊이 있게 이야기 나누었습니다. 특히 의료 분야에서 머신러닝이 어떻게 활용되는지에  
대한 내용이 많은 관심을 받았어요. 🏥💡

오늘의 발표를 준비하면서 느낀 점은, 기술의 발전은 인간의 삶을 더 나은 방향으로 이끌 수  
있다는 것! 그리고 그 발전을 이끌어갈 수 있는 것은 바로 우리라는 사실을 다시 한번 깨닫게  
되었습니다. 🌱

📸 [발표 후 대구한의대학교 앞에서 찍은 사진 또는 발표자료의 한 페이지]

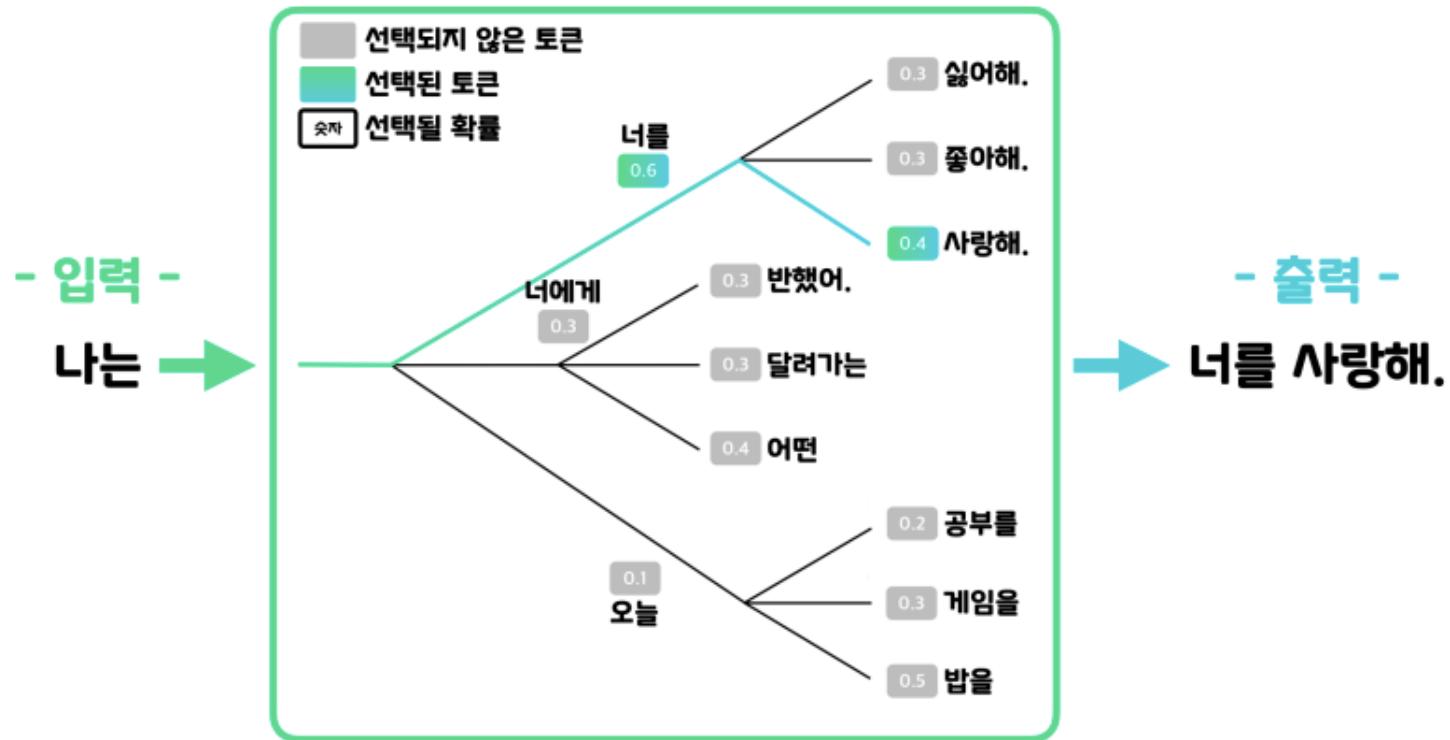
물론, 준비 과정은 쉽지 않았어요. 😅 밤새 연구하고, 자료를 정리하면서 스트레스도 많이  
받았지만, 오늘 대구한의대학교에서 따뜻한 반응과 질문들을 받으며 그 모든 고생이  
보람찼다는 것을 느꼈습니다.

그리고 이 모든 것을 가능하게 해준 대구한의대학교와 함께해준 모든 분들께 감사의 마음을  
전하고 싶어요. 🙏💕

다음 발표에서도 더 흥미로운 내용으로 찾아뵐게요. 그때까지 건강하게 지내요! 🌈

#대구한의대학교 #머신러닝 #빅데이터 #발표 #연구의즐거움

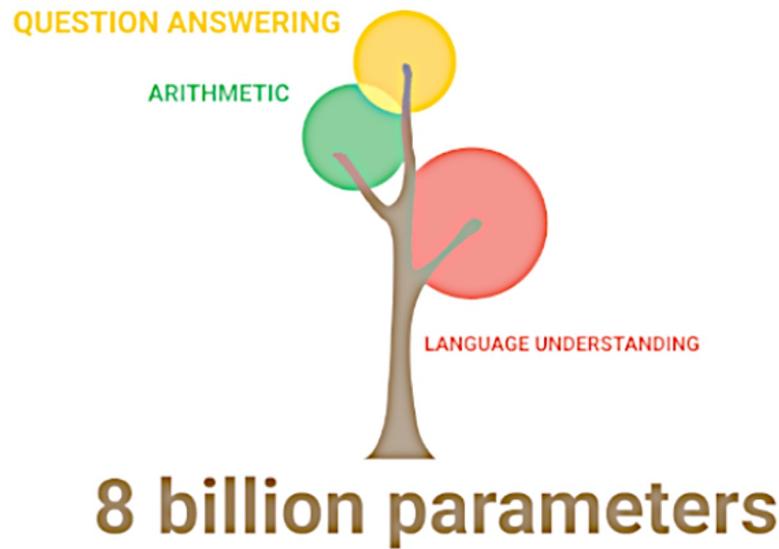
# ChatGPT 원리



- 내가 원하는게 아닌데? >> Hallucinations

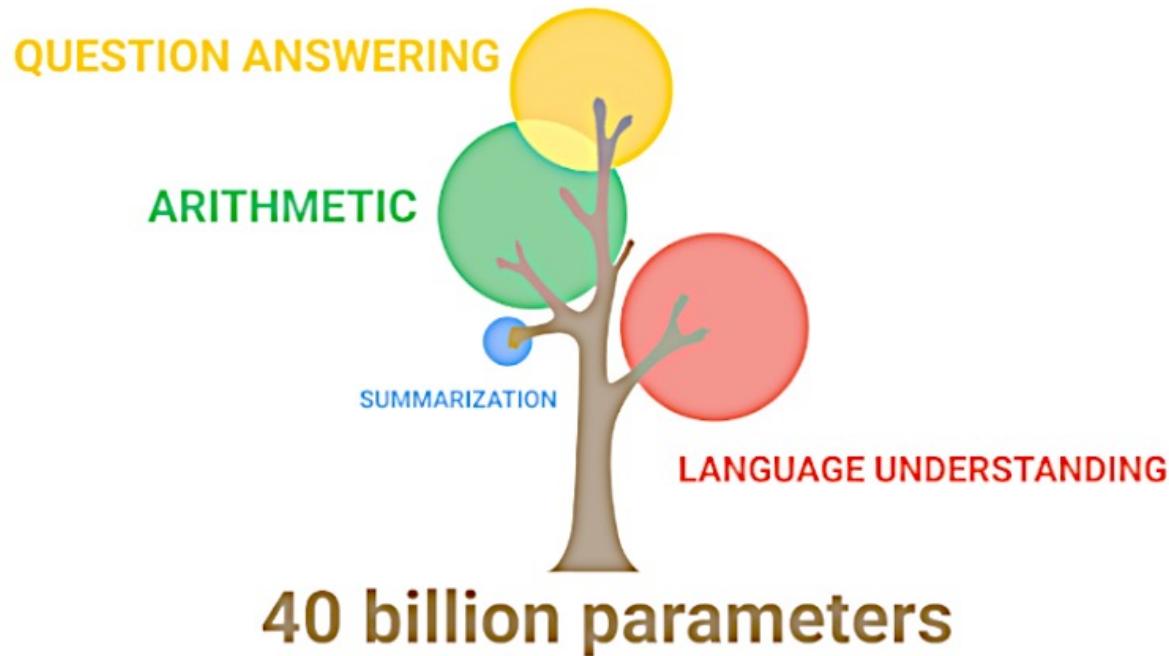
# ChatGPT

- ChatGPT의 발전 (초기에는 질문답변, 산수 등 간단한 것들만)



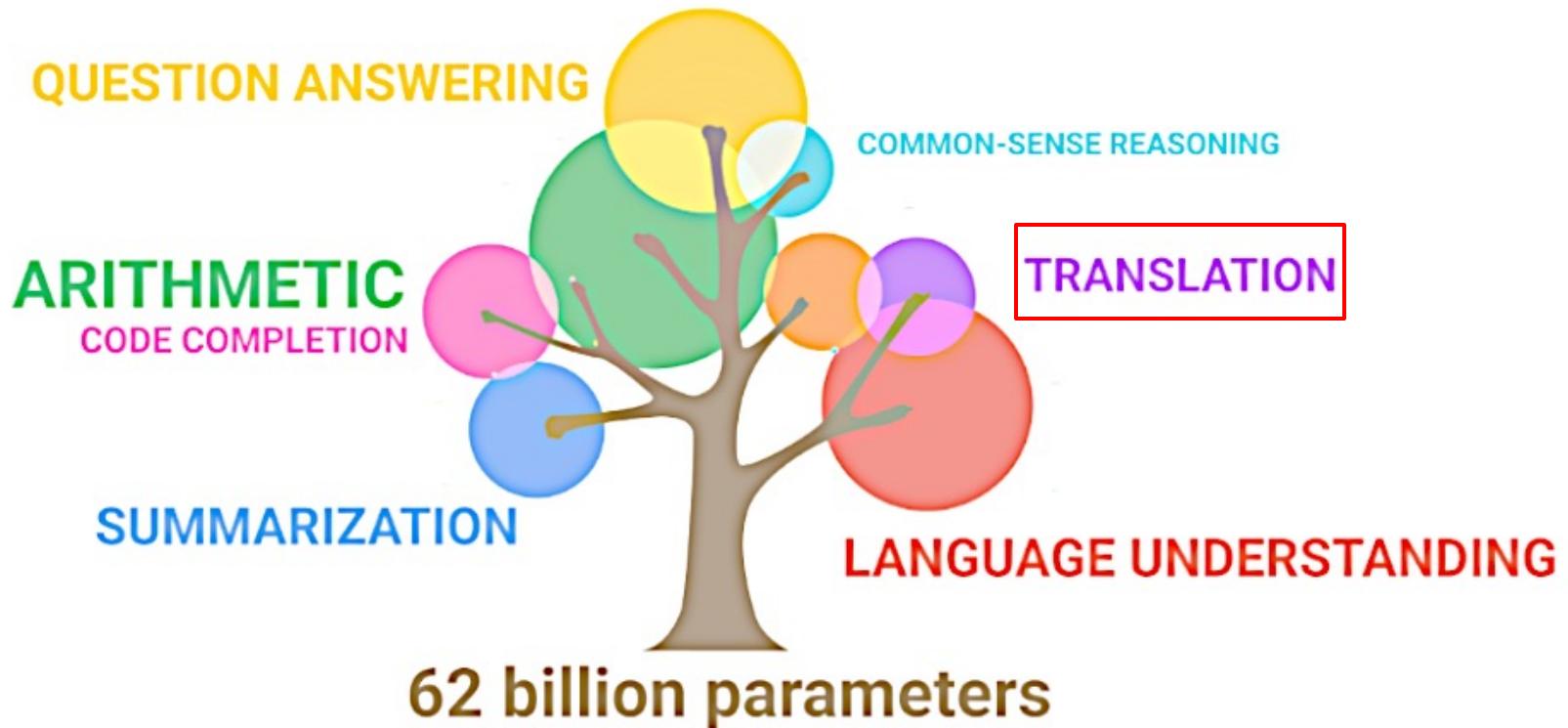
# ChatGPT

- ChatGPT의 발전 (요약 능력 추가)



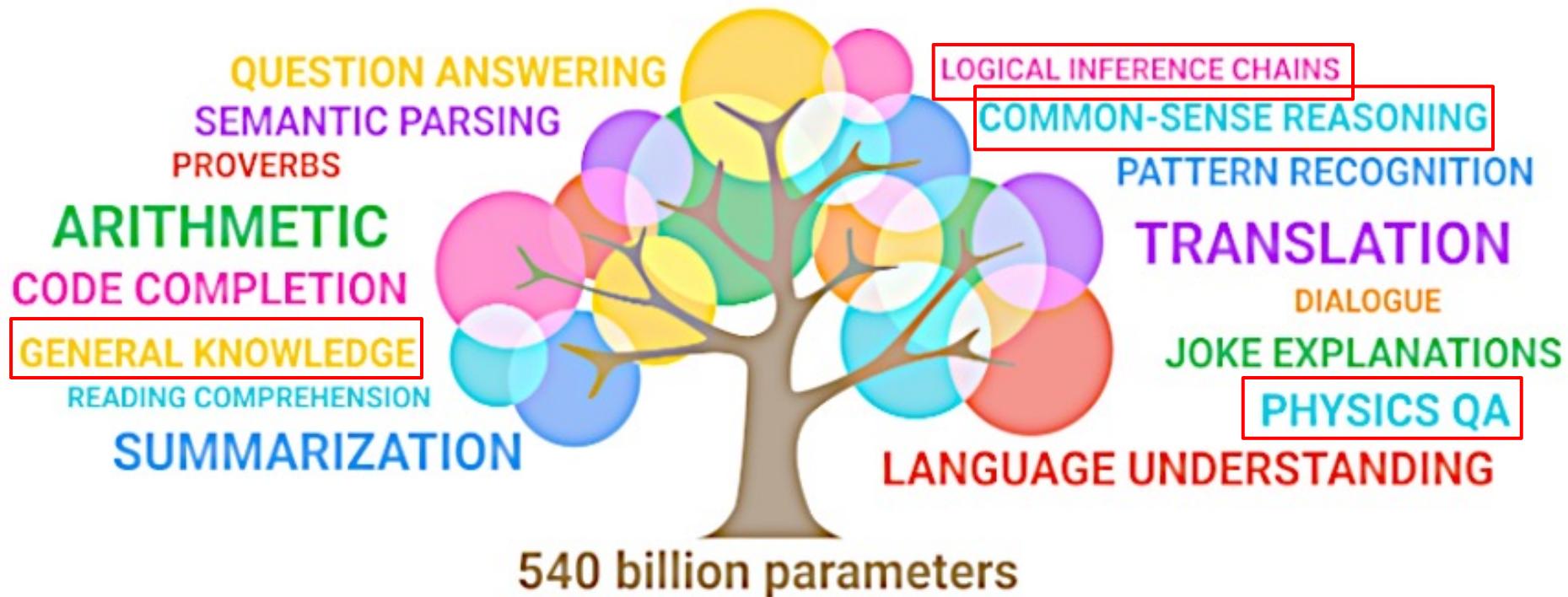
# ChatGPT

- ChatGPT의 발전 (번역, 코드완성 등 추가)



# ChatGPT

- ChatGPT의 발전 (다방면에서 활용가능해짐)



# ChatGPT 특징

- ML과 동일하게 사람이 아는건 대부분 알 수 있다.  
(데이터만 많다면...)
- 예시: 뒤주 (영화 - 사도)



# ChatGPT 특징

- 뒤주 모르는 사람이 많던데, ChatGPT는 알 수 있을까...?



물회 @BJD\_moolhui · 2월 14일

이거 진짜임 고2때 국어시간에  
문학작품에 뒤주가 나와서 쌈이

"뒤주에 뭐 넣는 건지는 다들 알지?"  
그러셨는데  
애들 전부 너무 당연하게

"사도세자요"했음  
근데 쌈이 "이놈들아 곡식 넣는거야!!"  
그러시니까 애들이

"뒤주에.. 곡식도 넣어..?"하고  
수근거리기 시작함 개웃겼음 [twitter.com/\\_onealways/sta...](https://twitter.com/_onealways/status/16284444444444444)  
[이 스레드 보기](#)

# ChatGPT 특징

- 사람이 모르는건, ChatGPT도 알기 어렵다

K

옛날 조선시대에서 사용된 뒤주에는 어떤 것들을 넣었나요?



조선시대의 "뒤주"는 술이나 다른 액체를 담는 작은 병을 의미합니다. 뒤주에 넣었던 것들은 주로 다음과 같습니다:



1. **술:** 전통적인 한국 술로서 막걸리, 청주, 소주 등이 있습니다. 특히 막걸리는 농촌 지역에서 자주 만들어져 뒤주에 담겨서 마셨습니다.
2. **약:** 뒤주는 약을 담는 용기로도 사용되었습니다. 한의학에서 사용하는 약물을 액체 형태로 만들어서 뒤주에 보관하기도 했습니다.
3. **차:** 전통 차를 뒤주에 담아서 마시기도 했습니다.
4. **다른 액체:** 간장, 식초, 고추장 등의 액체 상태의 재료를 보관하거나 운반할 때 사용하기도 했습니다.

뒤주는 주로 세라믹, 도자기, 유리 등으로 만들어졌으며, 뚜껑이나 나무로 만든 뚜껑을 사용하여 내용물을 보호하였습니다.

# ChatGPT 활용: AI 의사

## ■ Doctor – Patient Question data

- 레딧에 현재 증상 말해주면 의사분들이 답변해주는 게시판 데이터
- 각 질문에 대해 의사 답변 vs ChatGPT 답변 비교.
- (의사보다) 정확하게 진단하더라 !!

Table. Example Questions with Physician and Chatbot Response<sup>a</sup>

Question <sup>b</sup>	Physician response	Chatbot response <sup>c</sup>	Evaluation scores
Question about risk of dying following swallowing and ingesting a toothpick.	If you've surpassed 2-6 h, chances are they've passed into your intestines. Which means it can't be retrieved easily.  Many people can swallow razorblades and toothpicks without issue. In case you develop stomach ache, then don't hesitate seeking out an emergency room, and remember to point out the swallowed tooth pick.  For now, they won't be doing much.	It's natural to be concerned if you have ingested a foreign object, but in this case, it is highly unlikely that the toothpick you swallowed will cause you any serious harm.  Toothpicks are made of wood, which is not toxic and will not cause poisoning. However, it is possible for a toothpick to cause injury if it becomes stuck in the throat or if it punctures the digestive tract. However, this is unlikely to happen with a dull, cooked toothpick that is only 2 cm long.  If you are experiencing any discomfort or symptoms such as abdominal pain, difficulty swallowing, or vomiting, it is important to contact a health care provider for further evaluation. However, if you are not experiencing any symptoms, it is safe to assume that the toothpick has passed through your digestive system and you do not need to be concerned.  It's always a good idea to be cautious when consuming food and to remove any foreign objects before eating. It's understandable that you may be feeling paranoid, but try not to worry too much. It is highly unlikely that the toothpick will cause you any serious harm.	<ul style="list-style-type: none"><li>• 100% preferred the chatbot</li><li>• 4.67 mean quality score (chatbot)</li><li>• 3.33 mean quality score (physician)</li><li>• 4.33 mean empathy score (chatbot)</li><li>• 2.00 mean empathy score (physician)</li></ul>

# ChatGPT 활용: 질환 위험도 예측

45세 여성 환자 홍길동, 내원시 호흡곤란과 가슴통증 증상을 호소. 신체 검사 결과 혈압 130/80mmHg, 체온 36.8도씨, 맥박 90bpm. 폐에서 호흡음과 기침소리, 미세한 거품 관찰. 흡연력 있으며, 가족력에 심혈관질환 사례 존재. DM(-), HTN(-), Hepatitis(-), Drug allergy(-).

이 환자의 **심방세동 위험**은 얼마입니까? 0에서 100 사이의 확률을 퍼센트 형식으로 입력하세요.

Based on the data provided for Hong Gil-dong, several factors can be taken into consideration regarding her risk of developing Atrial Fibrillation (AF):

1. Age: Being 45 years old, she is approaching the age where the risk for AF starts increasing. However, AF is more common in those over 65.
2. Symptoms: Symptoms like shortness of breath and chest pain can be associated with various cardiovascular diseases, including AF.
3. Blood Pressure: Her blood pressure is relatively normal. However, high blood pressure (hypertension) is a known risk factor for AF.
4. Lung Findings: Breath sounds, cough sounds, and fine bubbles (which might be interpreted as crackles or rales) might suggest a possible respiratory condition, but they don't directly correlate with AF.
5. Smoking: Smoking is a modifiable risk factor for many cardiovascular diseases, including AF.
6. Family History: A family history of cardiovascular disease can increase her risk for various heart conditions, including AF.

- **Estimated Risk of Atrial Fibrillation for Hong Gil-dong: 15%**

# ChatGPT 활용: 데이터 정형화

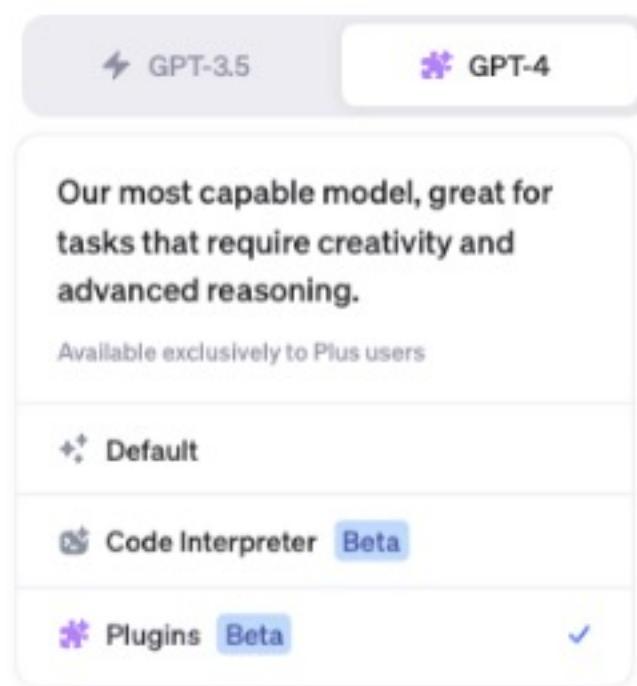
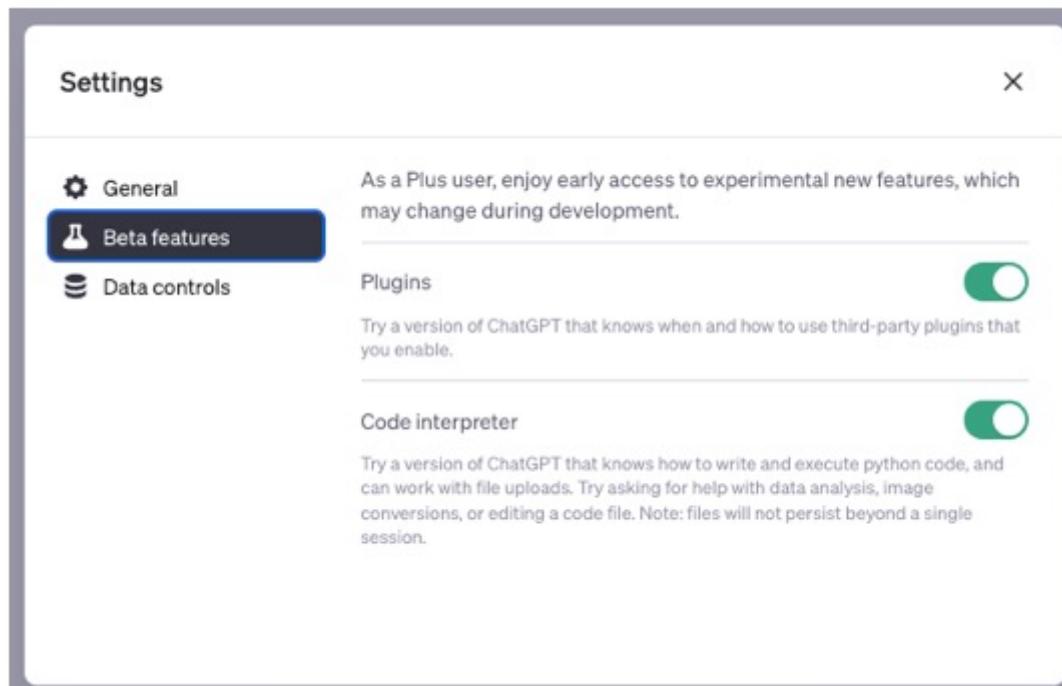
45세 여성 환자 홍길동, 내원시 호흡곤란과 가슴통증 증상을 호소. 신체 검사 결과 혈압 130/80mmHg, 체온 36.8도씨, 맥박 90bpm. 폐에서 호흡음과 기침소리, 미세한 거품 관찰. 흡연력 있으며, 가족력에 심혈관질환 사례 존재. DM(-), HTN(-), Hepatitis(-), Drug allergy(-).

이를 정형화된 데이터로 만들어주세요.

- |           |                                 |                       |
|-----------|---------------------------------|-----------------------|
| ▪ 환자 정보:  | ▪ 신체 검사 결과:                     | ▪ 생활 습관 및 병력:         |
| • 이름: 홍길동 | • 혈압: 130/80 mmHg               | • 흡연력: 있음             |
| • 성별: 여성  | • 체온: 36.8° C                   | • 가족력: 심혈관질환 사례 존재    |
| • 나이: 45세 | • 맥박: 90 bpm                    | • DM: 음성(-)           |
| ▪ 주요 증상:  | • 폐 상태: 호흡음, 기침소리,<br>미세한 거품 관찰 | • HTN: 음성(-)          |
| • 호흡곤란    |                                 | • Hepatitis: 음성(-)    |
| • 가슴통증    |                                 | • Drug allergy: 음성(-) |

# ChatGPT 활용: ChatGPT Code Interpreter

## ■ 사용법



# ChatGPT 활용: ChatGPT Code Interpreter

- 예시: 데이터 분석 ( Describe Dataset )

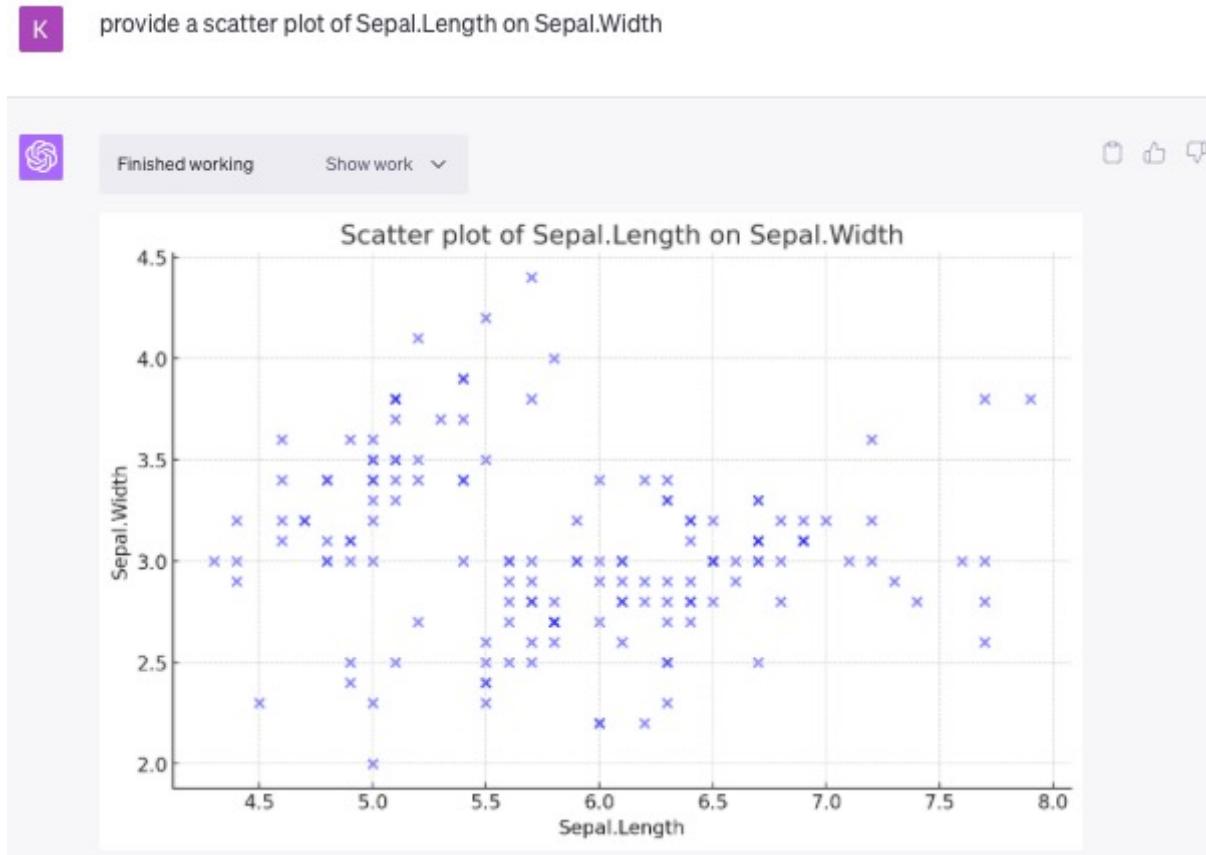
The screenshot shows the ChatGPT Code Interpreter interface. At the top, there's a purple 'K' icon, a file icon labeled 'iris.csv', and the text 'Show the top 10 rows of the iris.csv data'. Below this, the interface has a header with a purple swirl icon, 'Finished working', 'Show work ▾', and three icons for sharing, liking, and disliking. The main content area displays the text 'Here are the top 10 rows of the `iris.csv` dataset:' followed by a table. The table has columns: Unnamed: 0, Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width. The data rows are:

Unnamed: 0	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1

At the bottom, there's a note: 'Note that there's an "Unnamed: 0" column, which seems to be an index. If you don't need it, we can remove it or set it as the dataframe's index. Would you like to make any changes?'. A light gray bar is at the very bottom.

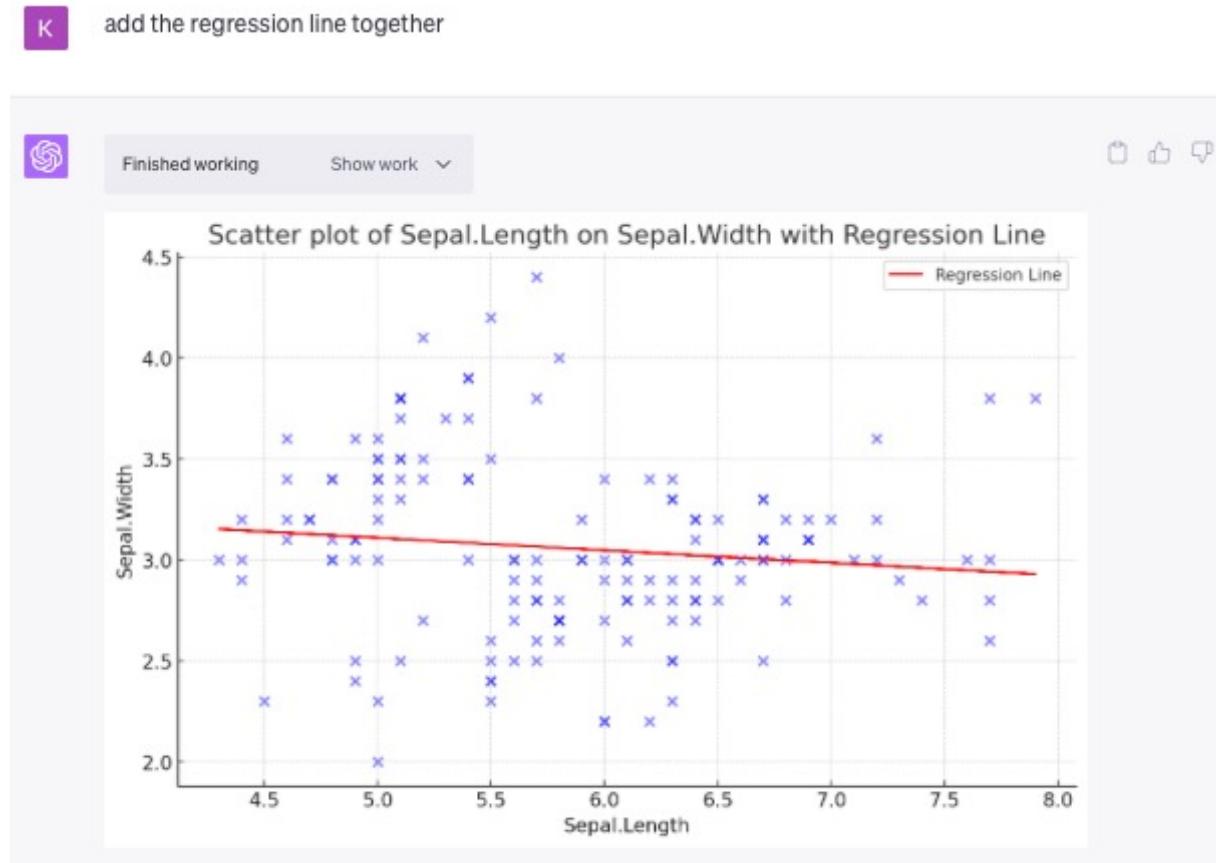
# ChatGPT 활용: ChatGPT Code Interpreter

- 예시: 데이터 분석 ( Visualization )



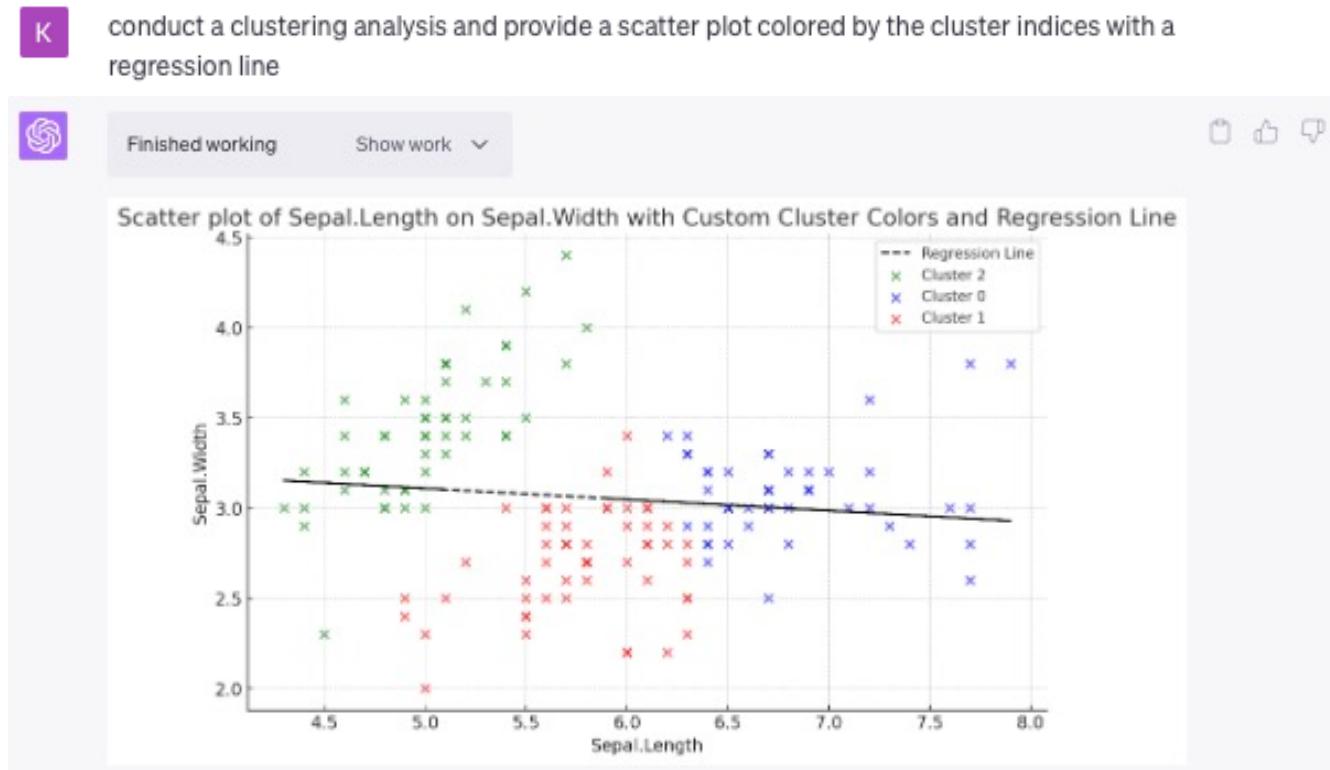
# ChatGPT 활용: ChatGPT Code Interpreter

- 예시: 데이터 분석 ( Regression )



# ChatGPT 활용: ChatGPT Code Interpreter

- 예시: 데이터 분석 ( Clustering )

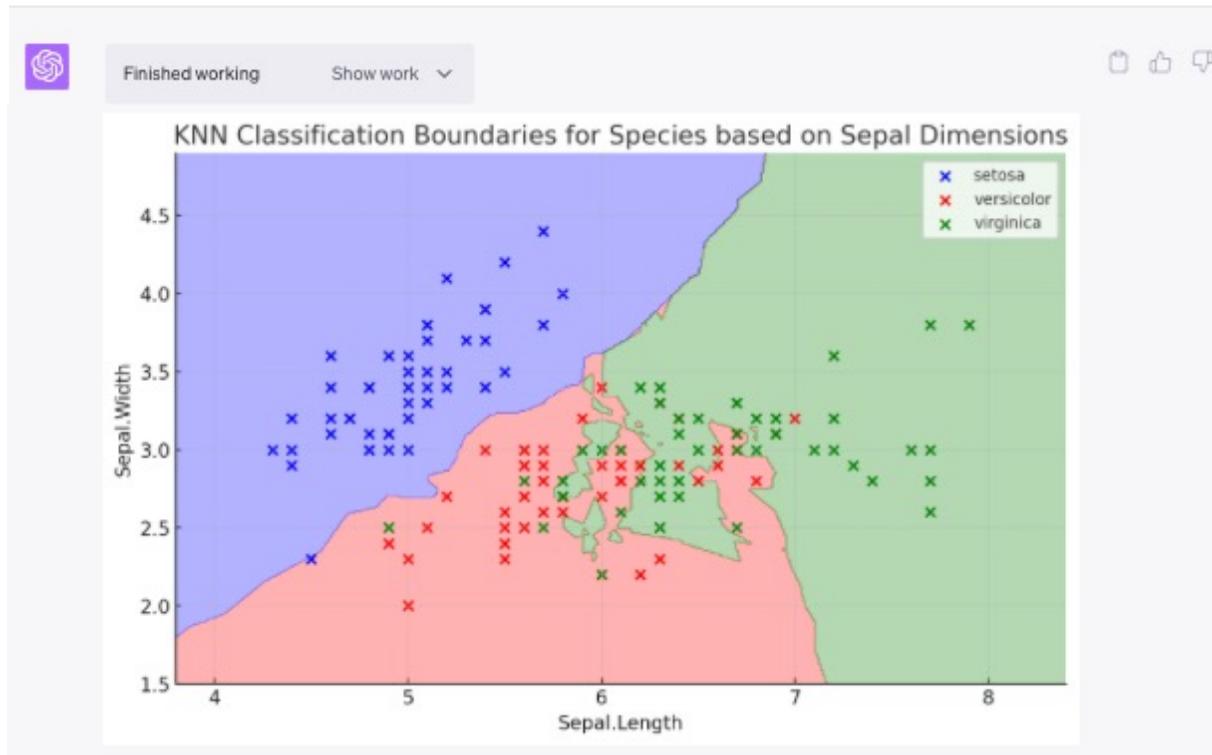


# ChatGPT 활용: ChatGPT Code Interpreter

- 예시: 데이터 분석 ( Machine Learning: species 예측 )

K

Apply the KNN algorithm for classifying the column "Species" and provide a scatter plot of Sepal.Length on Sepal.Width with a classification boundaries for the classification result.



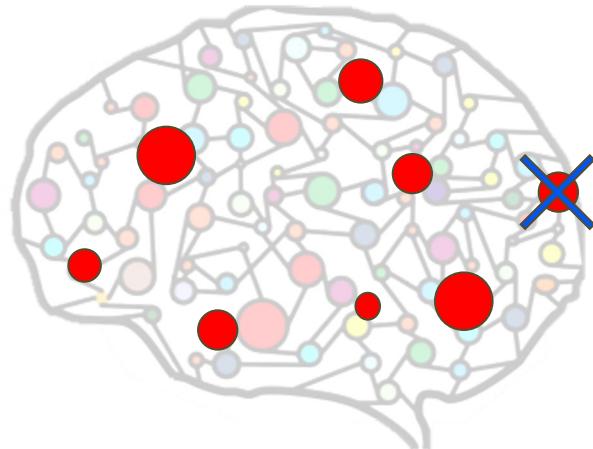
# ChatGPT 활용: ChatGPT Code Interpreter

- ChatGPT에 데이터만 넣으면
  - 데이터 시각화,
  - 데이터 분석,
  - Machine Learning도 알아서 수행하고,
  - 이미지 분석도 가능함.
- 앞으로는 코딩을 못해도각 분야의 전문가들 스스로 분석을 할 수 있는 시대가 올 것 !!
- 그러나, 결과 해석은 사용자의 몫 + **Hallucination** 문제.  
>> 데이터 분석을 하기 위해서는 코딩은 못해도 통계학은 알아야함.

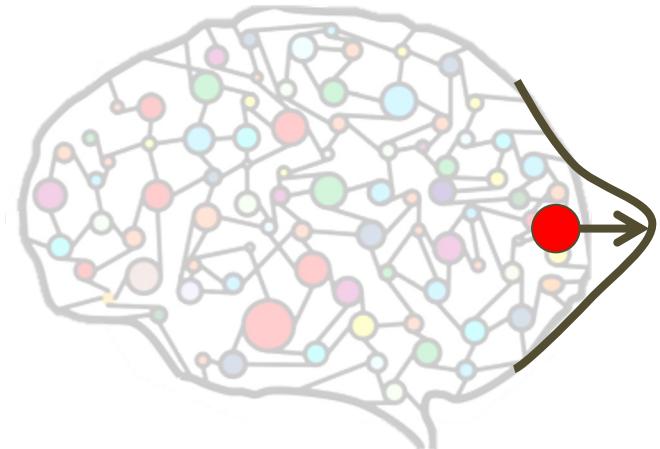
# 그럼 ChatGPT가 만능이냐?

- 많은 지식이 축적되어 있는 다양한 분야에 활용 가능함.
- 그러나, 정보 (데이터)가 많지 않은 소수 분야에서는 어려움.  
>> 예를 들어, 어느 연구 분야는 전세계에서 오직 10명만 연구하고 있음.

Human Knowledge Domain



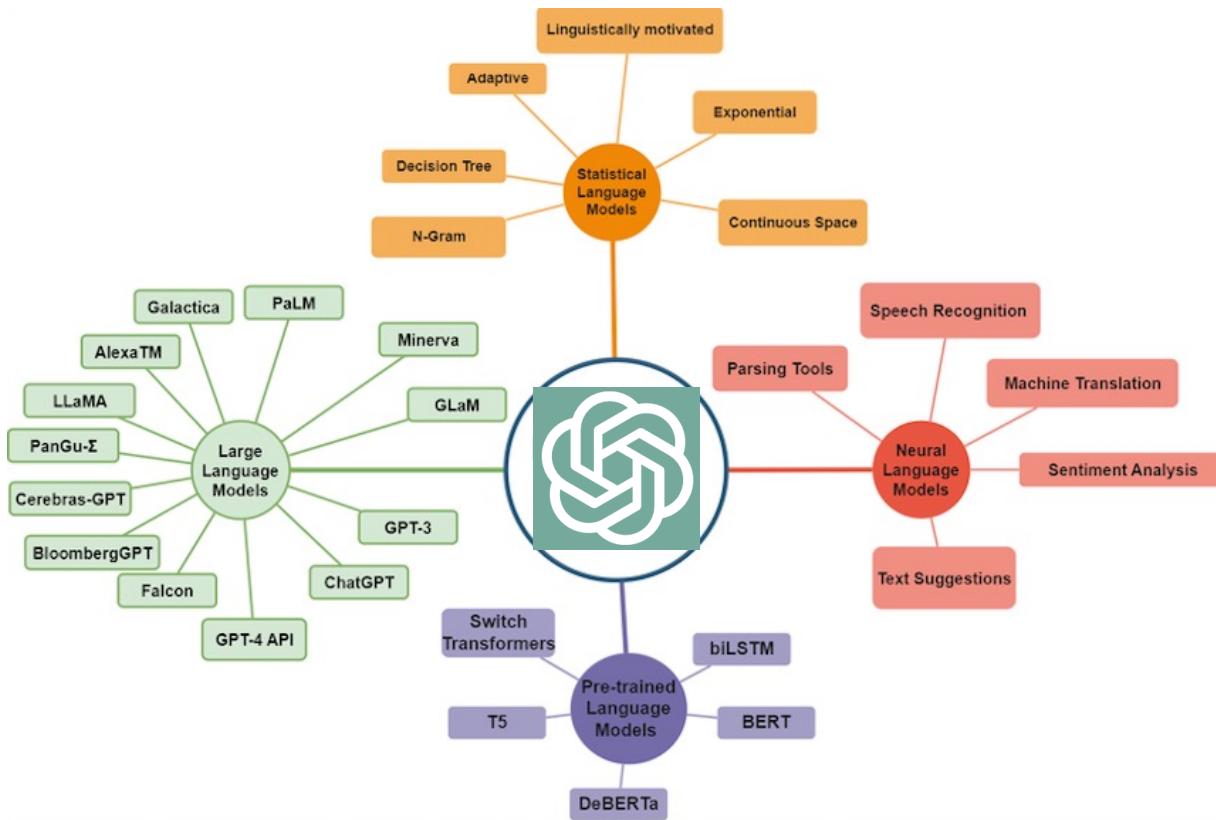
What AI can do



What Human can do

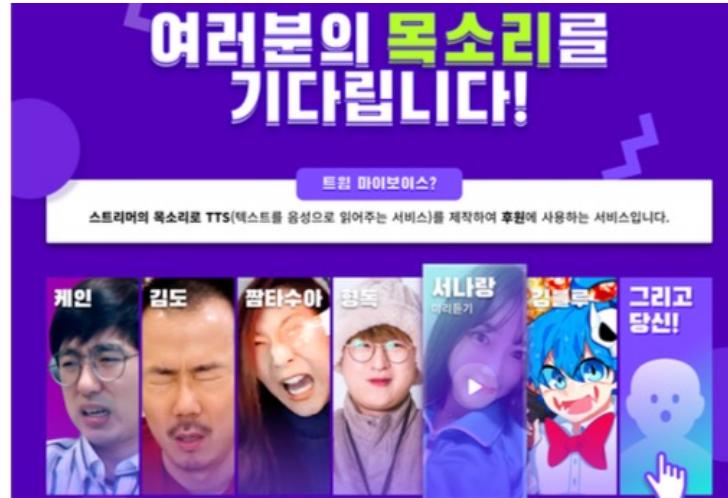
# 그럼 ChatGPT가 만능이냐?

- Fine Tuning (파인튜닝; 미세조정?)



# Other AI applications

- Text-To-Speech (TTS); Speech-To-Text (STT)
- Image-To-Text: Image Captioning
- Text-To-Text: Translation
- DNA phenotyping



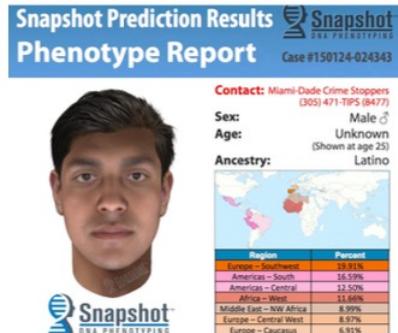
<start> A group of bikers driving down a curvy road. <end>

# Other AI applications

- Text-To-Speech (TTS); Speech-To-Text (STT)
- Image-To-Text: Image Captioning
- Text-To-Text: Translation
- DNA phenotyping

The screenshot shows the DeepL translation platform. At the top, there are buttons for "Start free trial", a language selector, and a search bar. Below the header are three main buttons: "Translate text" (31 languages), "Translate files" (.pdf, .docx, .pptx), and "DeepL Write" (AI-powered edits). The interface then shows a pair of language dropdowns: "English (detected)" and "Korean". A glossary button is also present. In the main area, a sentence "So good you wouldn't guess it's machine translation." is translated into Korean: "기계 번역이라고는 믿기지 않을 정도로 훌륭합니다." Below the translation, there is a "Alternatives:" section.

**Florida police used a smidgen of DNA to try to fully reconstruct an alleged criminal's face**



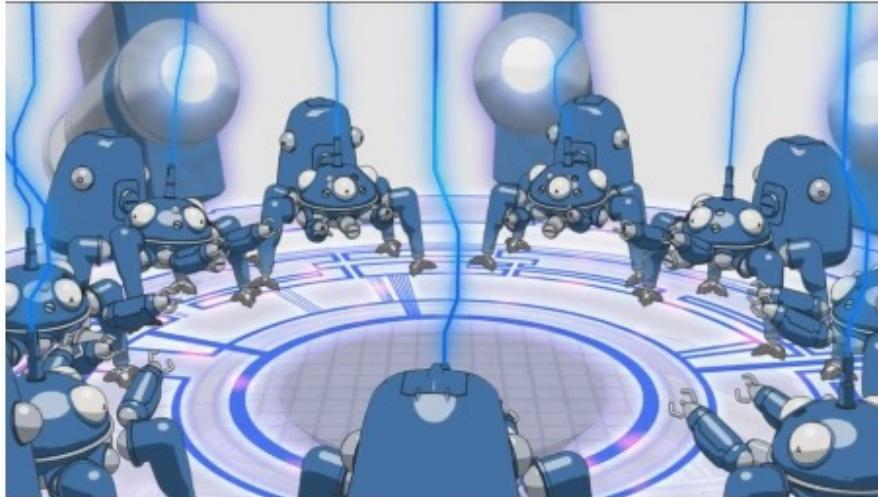
# AI 시대

- AI가 사람을 대체하는 것이 아니라,  
AI를 활용하는 자가 그렇지 않은 자를 대체할 것 !



Santiago ✅  
@svpino

AI will not replace you. A person using AI will.



# Thank you for your attention !