

AI와 머신러닝

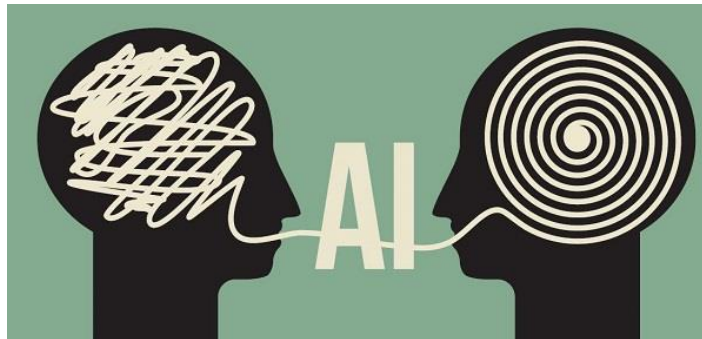


목차

1. 인공지능, 머신러닝, 딥러닝의 이해
2. 인공지능과 주요 응용 분야
3. 인공지능의 현재

인공지능의 정의

- 인공지능이란 인간이 행동하는 지능적인 작업을 자동으로 하기 위한 연구분야로,
 - ✓ 사람의 생각과 관련된 활동을 자동화하는 것을 연구
 - ✓ 인간처럼 사고하고 행동하는 기계장치를 연구하는 분야
 - ✓ 예를 들어, 로봇이 인간의 언어를 해석하고, 나아가 인간의 감정까지 파악하는 수준을 지향한다.

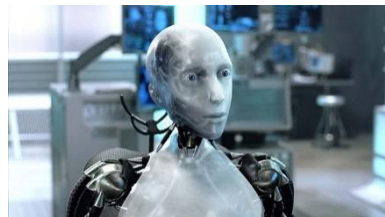


<https://www.techm.kr/news/articleView.html?idxno=2209>

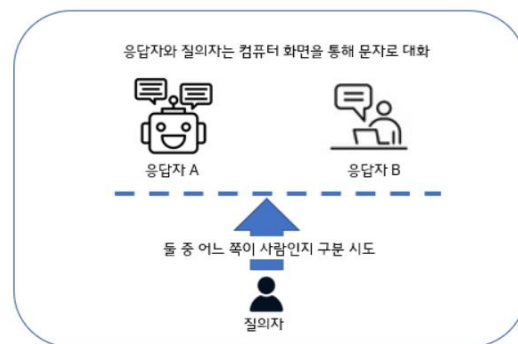
강한인공지능 vs 약한인공지능

- 인공지능은 인공지능의 [지적인 수준]에 따라 강한인공지능과 약한인공지능으로 분류

- 강한인공지능



- ✓ 모든 영역에서 인간과 유사한 지적 수준을 갖는 인공지능
 - 사람과 같이 자아의식, 감정, 지혜, 양심 등을 갖는 인공지능을 의미한다.
 - 예) 기계가 얼마나 인간과 비슷하게 대화할 수 있는가로 기계의 지능 유무를 판단하는 튜링테스트



<튜링 테스트 개요>

- 약한인공지능

- ✓ 특정 문제의 해결에 집중되는 지적 수준을 갖는 인공지능
 - 대부분의 인공지능이 인간의 지능을 모방하는 수준으로 구성된다.
 - 예) 자율주행, 스팸메일 필터링, 외국어 번역, 유튜브 영상 추천



인간을 상대한 인공지능

1997년, IBM 딥블루
(Deepblue)
세계 체스 챔피언을 상대로 승리



2011년, IBM 왓슨(Watson)
미국 퀴즈 방송 제퍼디에서 전
챔피언 브래드 루터와 켄 제닝
을 상대로 승리

2016년, 구글 딥마인드
알파고(AlphaGo)
이세돌과의 바둑대결에서
4승 1패로 승리



인공지능 알파고	vs	인간
2015. 10		vs 판후이
2016. 03		vs 이세돌
2017. 07		vs 커제

[이미지 출처]
<https://www.yna.co.kr/view/AKR20160311132000009>
<https://ppss.kr/archives/188405>
<https://m.dongascience.com/news.php?idx=10919>

우리 생활에서 인공지능
기술이 적용됐다고 생각하는
것은 무엇이 있나요?

■ 음성지능

✓ <https://speech.selvasai.com/speech-recognition-demonstration>

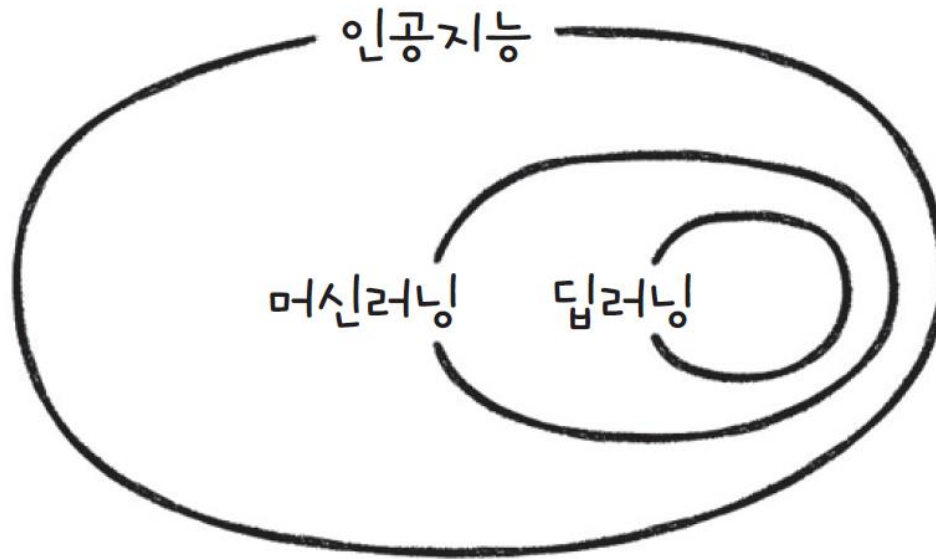
- 음성을 텍스트로 생성해 줌

✓ <https://speech.selvasai.com/text-to-speech-demonstration>

- 텍스트를 음성으로 합성해줌

머신러닝

- 머신러닝과 딥러닝은 약한 인공지능의 범주에 속한다.
- 딥러닝은 머신러닝의 다양한 방법론 중 하나이다.

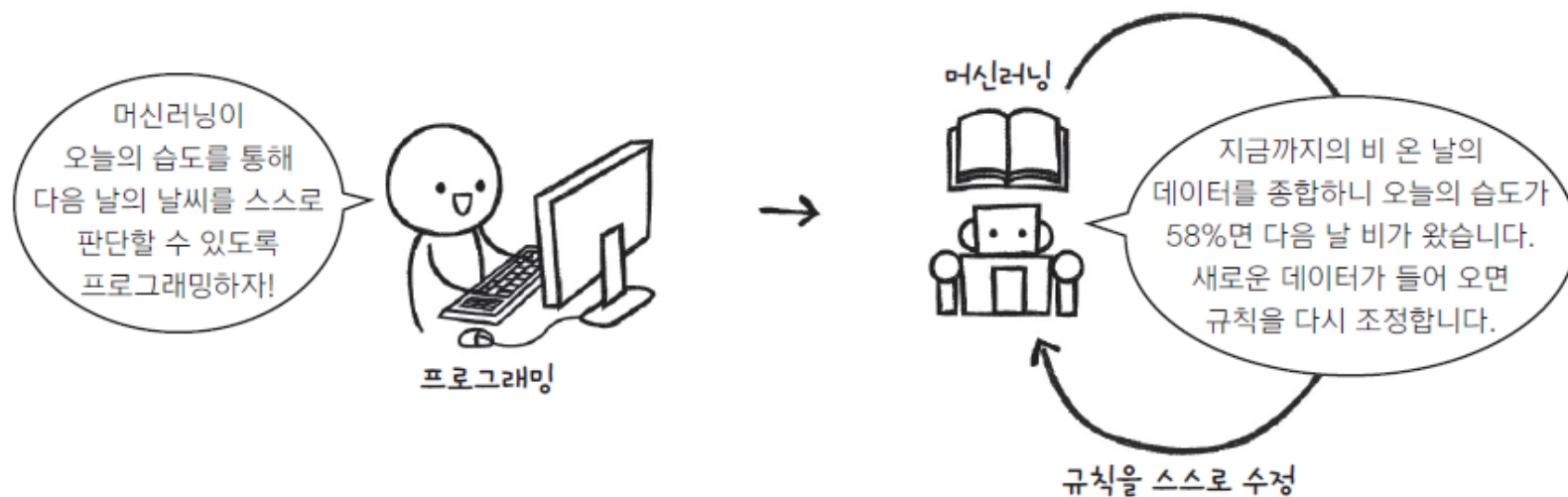


- 전통적인 프로그램은 전문가가 규칙을 수정하여 반영시킨다.



머신러닝

- 머신러닝은 사람이 개입하지 않고, 스스로 규칙을 수정해 나간다.



머신러닝(Machine Learning)

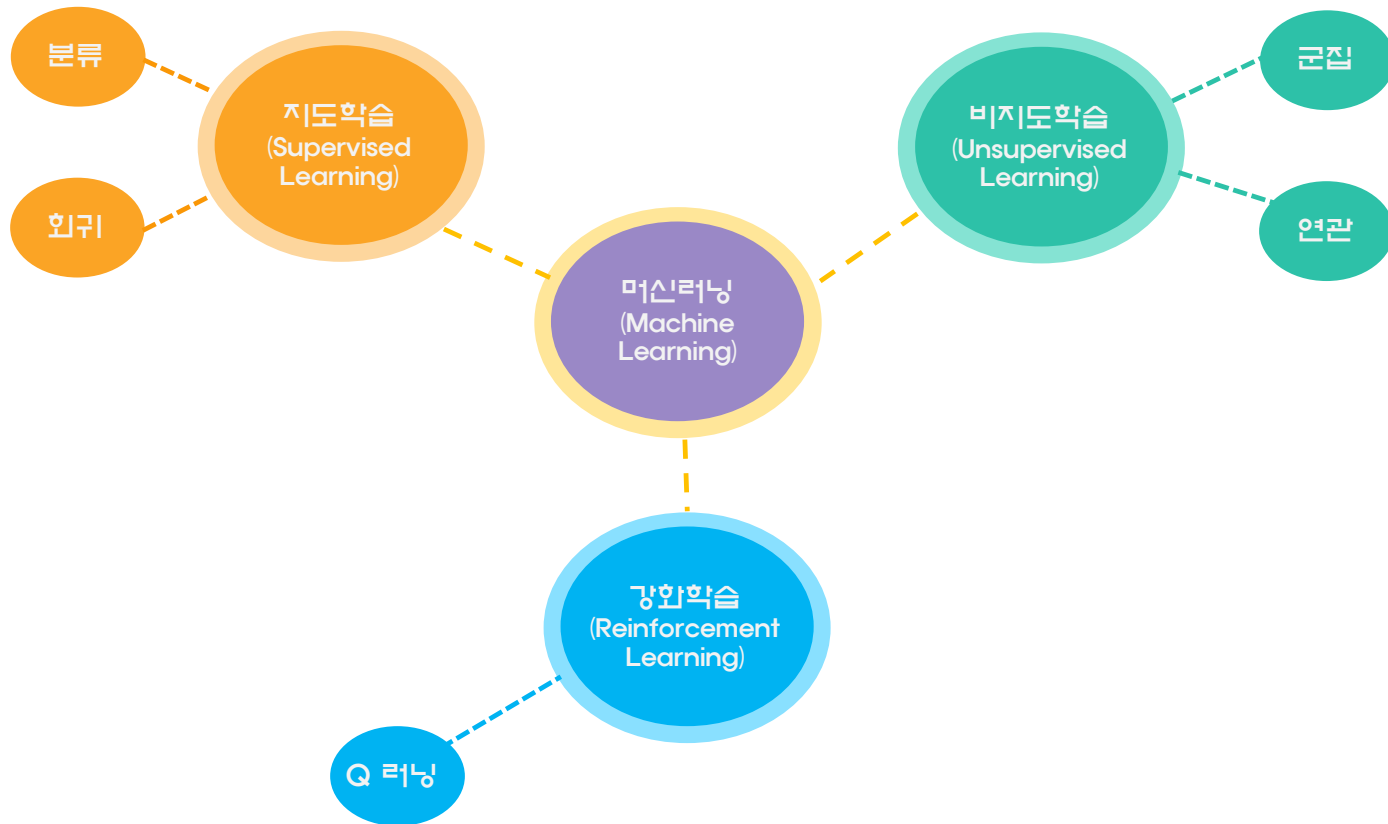
■ 머신러닝이란?

- ✓ 대량의 데이터와 알고리즘을 기반으로 경험을 통해 학습하여 데이터의 특징과 패턴을 발견
- ✓ 프로그래머가 명시적 코딩이 아닌 자기 개선과 예제를 통해 학습
- ✓ 학습으로 얻어진 정보를 기반으로 **미래의 임의의 데이터에 대한 결과(값, 분포)를 예측**

■ 머신러닝에 필요한 주요 요소

- ✓ 다양한 형식의 데이터(동영상, 이미지, 텍스트, 로그, html, xml, xlsx 등)
- ✓ 알고리즘 : 컴퓨터에게 지능(생각)을 심어주는 부분
 - 수학적 통계 모델, 의사결정트리, 서포트벡터머신, 신경망 등

머신러닝의 학습 방법



■ 분류(Classification)

✓ 개와 고양이 분류 예

학습데이터 셋 (training data set)

이미지 데이터	독립변수		종속변수
	눈모양	귀모양	정답(label/ Class)
	1	2	Dog
	2	3	Cat
	1	4	Dog
	2	5	Cat
	3	6	Dog

① 입력



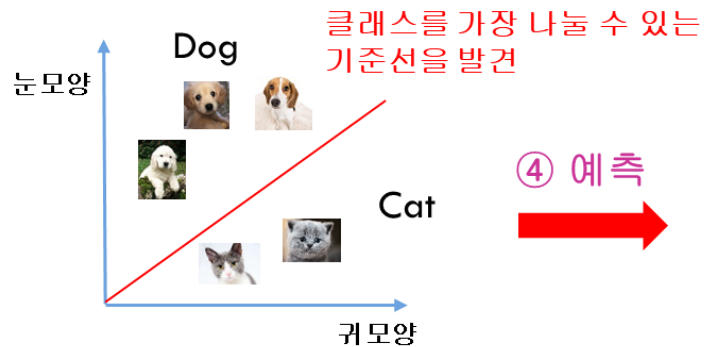
② 학습



③ 이것은 개? 고양이?



학습되지 않는
새로운 데이터



④ 예측



Dog

■ 회귀(Regression)

✓ 일별 온도와 레모네이드 판매량 예측 예

학습데이터 셋 (training data set)

독립변수 종속변수

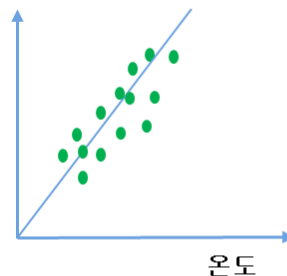
날짜	온도	판매량
2021.1.3	20	40
2021.1.4	21	42
2021.1.5	22	44
2021.1.6	23	46
2021.1.7	24	48

① 입력

② 학습



판매량



④ 예측

날짜	온도	판매량
2021.1.8	26	52
2021.1.9	27	54

③ 1.8과 1.9의 판매량은?

날짜	온도	판매량
2021.1.8	26	?
2021.1.9	27	?

학습되지 않는
새로운 데이터

지도학습의 예

■ 분류

- ✓ 가지고 있는 데이터에 독립/종속 변수가 있고 종속변수가 이름(범주형 데이터)일 때 사용

독립변수	종속변수
공부시간	합격여부 (합격/불합격)
엑스선 사진과 영상 속 종양의 크기, 두께	악성 종양 여부(양성/음성)
품종, 산도, 당도, 지역, 연도	와인의 등급
키, 몸무게, 시력, 지병	현역, 공익, 면제
메일발신인, 제목, 본문내용(사용 단어, 이모티콘 등)	스팸 메일 여부
고기의 지방함량, 지방색, 성숙도, 육색	소고기 등급

■ 회귀(Regression)

- ✓ 가지고 있는 데이터에 독립/종속 변수가 있고 종속변수가 숫자(양적 데이터)일 때 사용

독립변수	종속변수
공부시간	시험점수
온도	레모네이드 판매량
역세권, 조망 등	집값
온실 기체량	기온 변화량
자동차 속도	충돌 시 사망 확률
나이	키

■ 비지도학습(Unsupervised Learning)

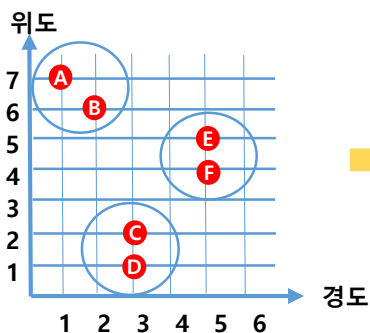
- ✓ 학습할 데이터에 **정답이 없이** 명확한 해를 계산하는 대신 **입력 데이터만으로** 숨겨진 특징이나 패턴을 학습을 통해 발견하는 방법
- ✓ 군집(Clustering)
 - 사전 정보가 없는 주어진 데이터들에서 공통적인 요소들을 찾아 유사성이 높은 군들끼리 다른 그룹으로 나누는 방법
- ✓ 연관규칙(Association rule)
 - 주어진 데이터에서 어떤 항목이 어떤 항목을 동반하여 등장하는지, 항목들 간에 연관규칙을 찾는 방법

■ 군집(Clustering)

- ✓ 전국적으로 분포하는 사용자에게 최적의 배달을 위한 배달 본부 배치
예

사용자 위치 데이터

사용자 명	위도	경도
A	7	1
B	6	2
C	2	3
D	1	3
E	5	5
F	4	5
...



사용자 명	위도	경도	그룹
A	7	1	1
B	6	2	
C	2	3	2
D	1	3	
E	5	5	3
F	4	5	
...	

■ 연관규칙(Association rule)

✓ 음악추천 예

음악 구매 내역 (구매한 가수 1, 구매하지 않는 가수 0)

	사용자1	사용자2	사용자3	사용자4
BTS	1	0	0	1
브루노마스	1	0	0	0
아이유	1	0	0	1
마룬5	0	1	1	0
퀸	0	1	0	0



마룬5와 다른 가수의 유사도 계산

	BTS	브루노마스	아이유	마룬5	퀸
마룬5	0	0	0	1	0.71



'마룬 5'를 구매한 사용자 3에게 '퀸'의 노래 추천

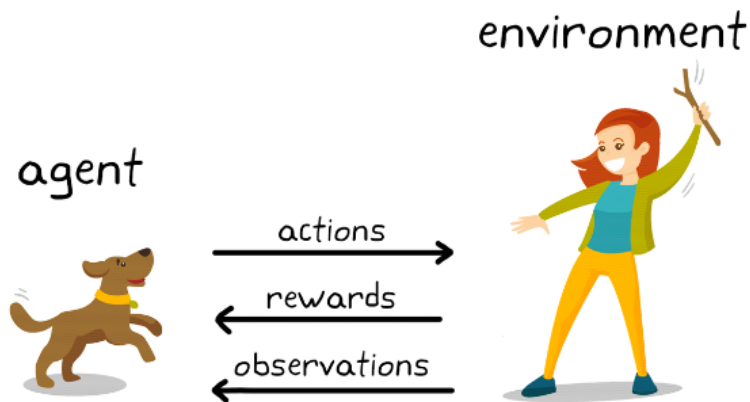
※ 마룬5와 다른 가수의 유사도 계산

$$\text{유사도} = \cos \theta = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

$$\text{코사인유사도} = \frac{0 \times 0 + 1 \times 1 + 1 \times 0 + 0 \times 0}{\sqrt{0^2 + 1^2 + 1^2 + 0^2} \times \sqrt{0^2 + 1^2 + 0^2 + 0^2}} = \frac{1}{\sqrt{2} \times \sqrt{1}} = 0.71$$

■ 강화학습 (Reinforcement learning)

- ✓ 학습의 주체가 어떤 환경에서의 시행착오의 경험을 통해 학습
- ✓ 지도학습과의 차이는 사람으로부터 학습을 받는 것이 아니라, 변화되는 환경으로부터 보상과 벌을 반복하며 학습한다는 차이를 가짐
 - 알파고 - 바둑의 기본 규칙과 자체 경기를 통해 습득한 3,000만 개의 기보를 학습한 후 스스로 대국하며 훈련하는 강화학습 알고리즘을 사용해 개발됨
 - 자율 주행 자동차와 드론 분야 등에서 활용



✓ Q-러닝

- 주어진 환경(state)에서 주체(agent)가 현재의 상태(state)를 관찰하며 선택할 수 있는 행동(action)들 중 가장 최대의 보상(reward)을 가져다주는 행동이 무엇인지 정책 데이터를 지속적으로 업데이트하는 알고리즘

■ 로봇에게 강화학습을 시키는 예

- ✓ 목표 : 출발점(S)에서 시작하여 도착점(G)까지 구멍(H)에 빠지지 않고 도착

현재상태 (State) : 16개의 격자 중 하나

학습주체 (Agent) : 로봇

행동 (Action) : 상하좌우 이동, 제자리

보상 (Reward) : 도착점에 도달 시 +1점

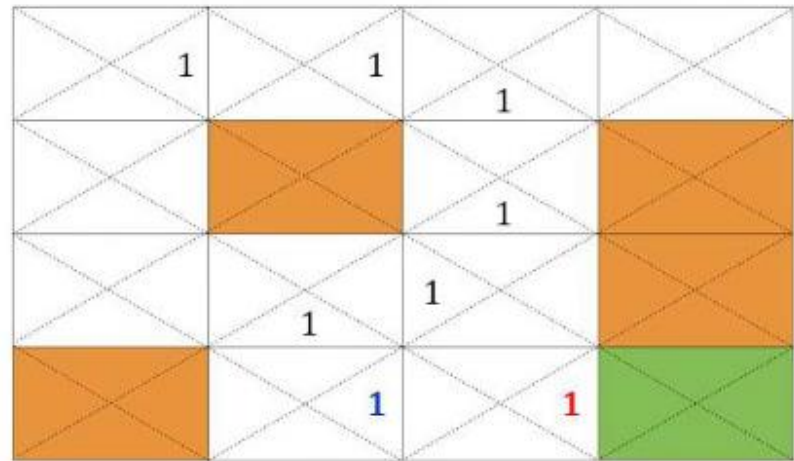


S	F	F	F
F	H	F	H
F	F	F	H
H	F	F	G

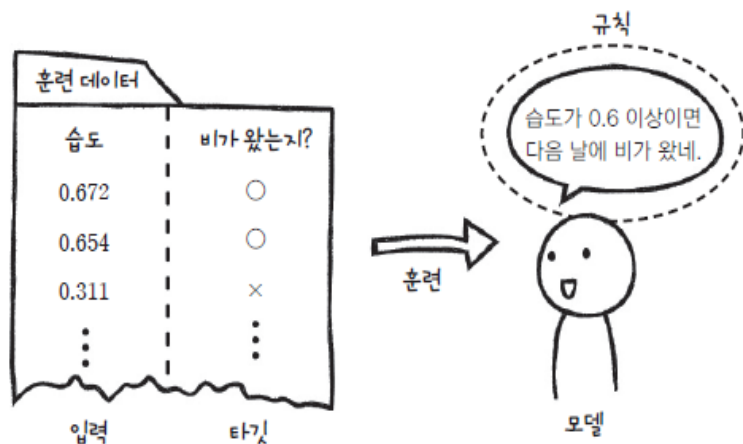
■ 로봇에게 강화학습을 시키는 예

✓ 보상과 벌점의 반복을 통해 경로를 학습

- 녹색박스 바로 앞에서 녹색격자로 이동한 마지막 행위에 +1
- 빨간 +1점의 격자로 들어서는 경우 +1점



- 기계학습에서 규칙이란 가중치와 절편을 의미한다.



$$1.5 \times x + 0.1 = y$$

(y 가 1 이상이면 다음 날 비가 온다고 예측)

가중치 입력 타겟 절편

- 손실함수를 활용하여, 모델의 규칙을 점진적으로 조정해 나간다.

훈련 데이터	
습도	비가 왔는지?
0.672	○
0.654	○
0.311	×
0.472	○

새로운 입력과 출력

입력 x → 타겟 y

$$1.5 \times x + 0.1 = y \quad (y \text{가 } 1 \text{ 이상이면 다음 날 비가 온다고 예측})$$

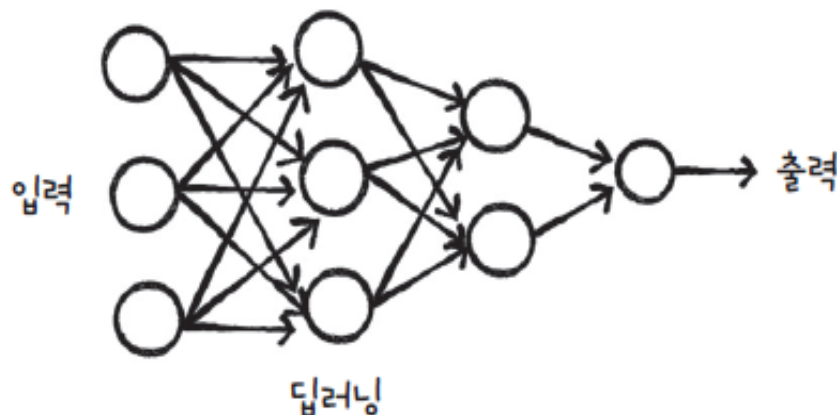
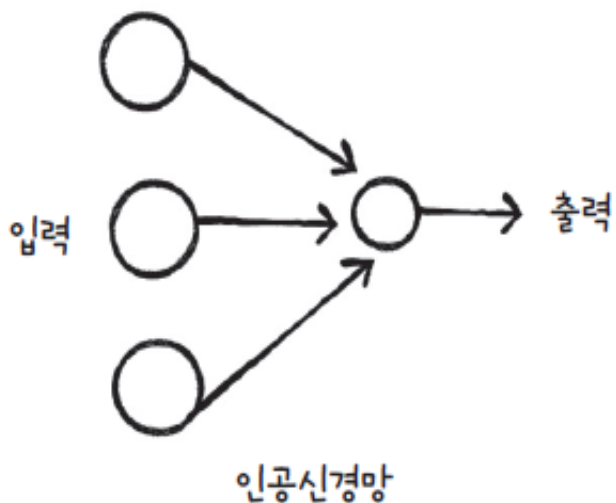
가중치 1.5 절편 0.1

$1.5 * 0.472 + 0.1 = 0.808$ 로 실제 타겟값인 1과 비교시, 맞지 않는다. 이 경우 가중치와 절편을 수정한다. 이때 규칙을 수정하는데 기준이 되는 함수를 손실함수라 부른다.

- 손실함수는 머신러닝이나 딥러닝 모델이 예측한 값과 실제 값 사이의 차이를 측정하는 함수
- 이를 통해 모델의 성능을 평가하고, 모델이 어떤 방향으로 개선되어야 할지 알려주는 역할 수행
- 손실함수의 값을 최소화하는 것이 모델 학습의 목표이다.

딥러닝

- 신경망(Neural Network)을 통해 인공지능을 만드는 머신러닝의 한 종류이다.
- 머신러닝의 수많은 방법 중 선형회귀, 로지스틱 회귀와 같은 하나의 방법론으로 존재하는 것이 딥러닝이다.
- 인공신경망 구조를 여러겹으로 쌓으면, 딥러닝 모델이다.



- 딥러닝은 머신러닝이 쉽게 분류, 예측하지 못하는 데이터를 상대적으로 쉽게 처리한다.(비정형 데이터)

딥러닝에 잘 맞는 데이터



이미지/영상, 음성/소리, 텍스트/번역
등의 비정형 데이터

머신러닝에 잘 맞는 데이터



데이터베이스, 레코드 파일, 엑셀/CSV
등에 담긴 정형 데이터

딥러닝은 주로 인지하는 것과 연관된 문제를 잘 풀이하는 것으로 알려져 있다.

■ 데이터마이닝

- ✓ 데이터마이닝은 대규모 데이터에서 의미 있는 정보를 채굴하는 것을 의미한다.
 - 광산에서 금을 캐는 것과 같이 빅데이터라는 산에서 의미 있는 데이터를 채굴하는 것
 - 기계학습, 통계학기법, 연관규칙, 분류패턴, 군집화패턴, 텍스트마이닝, 그래프마이닝, 추천, 시각화 등



■ 자연어 처리

✓ 사람이 사용하는 언어로 된 문서를 처리하고 이해하는 분야에 활용

- 주요 연구분야

- 자연어의 형태소분석 / 구문분석 / 품사태깅 / 의미분석

- 챗봇 등

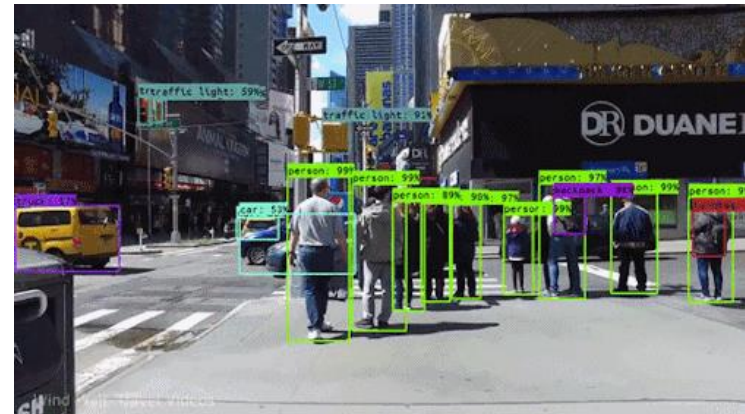
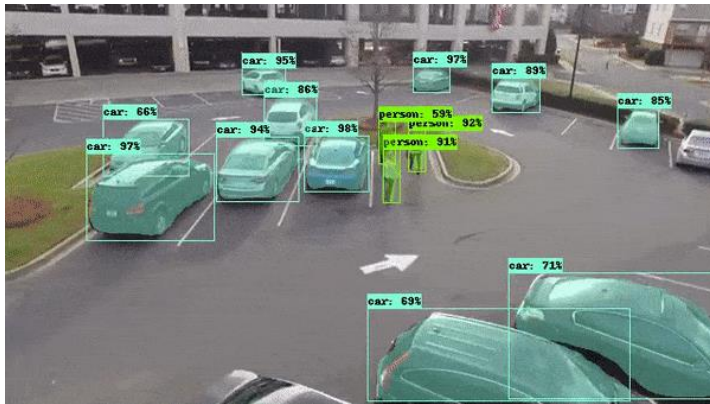
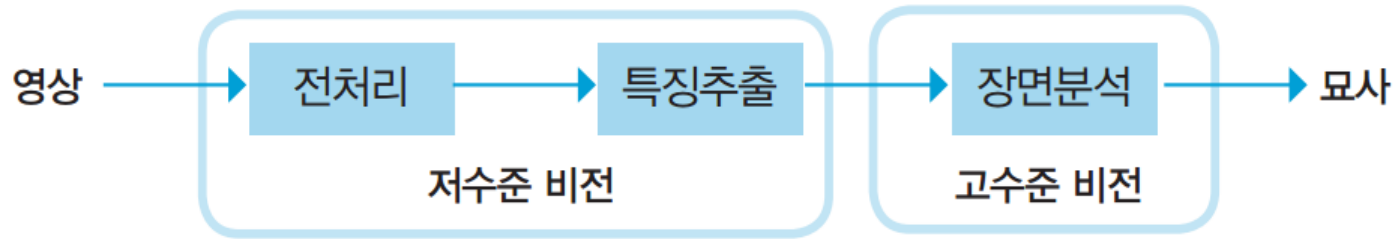


Ex) [이순신]_{인명}은 [1545]_{날짜}년부터[1598]_{날짜}년까지 [조선]_{지명} 중기의 무신이었다.

주요 응용 분야

■ 컴퓨터 비전(vision)

✓ 컴퓨터를 이용해 시각적인 기능을 갖는 기계장치를 만드는 분야



■ 로보틱스

- ✓ 인간과의 interaction이 가능한 로봇은 지능형 로봇공학과 연관이 깊다.

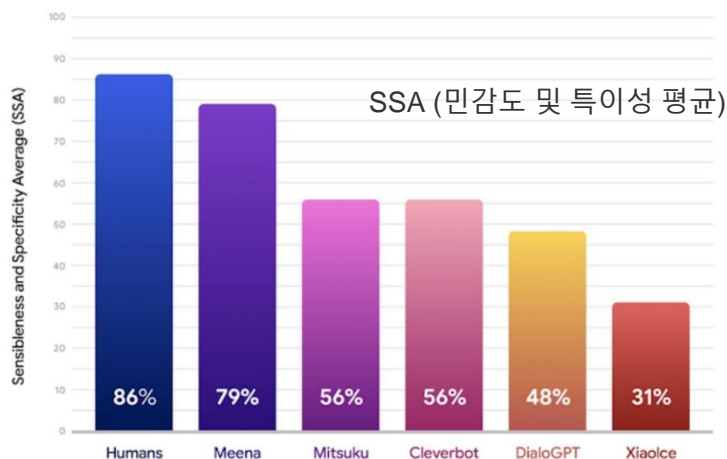


인공지능의 현재

▶ 인공지능 챗봇

▶ 구글 미나(Meena, 2020)

- ▶ 2020년 초 구글에서 발표한 멀티턴 오픈 도메인 챗봇 모델
- ▶ 공개된 소셜데이터들을 수집하여 학습데이터로 사용한 일상대화 딥러닝 모델



Conversation 2

Human: Hi!
Meena: Hi! How are you this fine evening?
Human: I'm doing well. it's morning my time!
Meena: Good morning! What are your plans for today?
Human: Get my work done then watch a movie at home in the night
Meena: Sounds fun! What movie?
Human: I haven't decided yet, perhaps an Oscar-nominated one. any suggestion?
Meena: I'm not too familiar with recent Oscar winners, but I do know I wanted to see The Grand Budapest Hotel.
Human: ah, tell me about that movie!
Meena: It's from the same director and writer as Moonrise Kingdom. Wes Anderson is pretty well known for making funny, colorful films. I've heard it's pretty good.
Human: ha! I've searched on the internet and both movies are very colorful!
Meena: I'm looking forward to it. I love colorful movies.

[이미지 출처] <http://aidev.co.kr/chatbotdeeplearning/8881>

인공지능의 현재

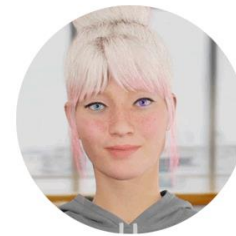
▶ 인공지능 챗봇

- ▶ 페이스북 블렌더(Blender, 2020)
 - ▶ facebook AI Research에서 발표한 오픈도메인 챗봇 모델
 - ▶ 대화의 일관성 유지, 호응도, 사람다움에 초점
- ▶ Kuki (Mitsuku, 2020)
 - ▶ 판도라 보츠(Pandora Bots)에서 발표한 18살 여성 챗봇
 - ▶ 뢰브너 상(Loebner Prize)을 5번 수상
 - ▶ <https://www.kuki.ai/>



블렌더와 쿠키의 봇 배틀

[영상출처] <https://youtu.be/Lxneek9fGbQ>



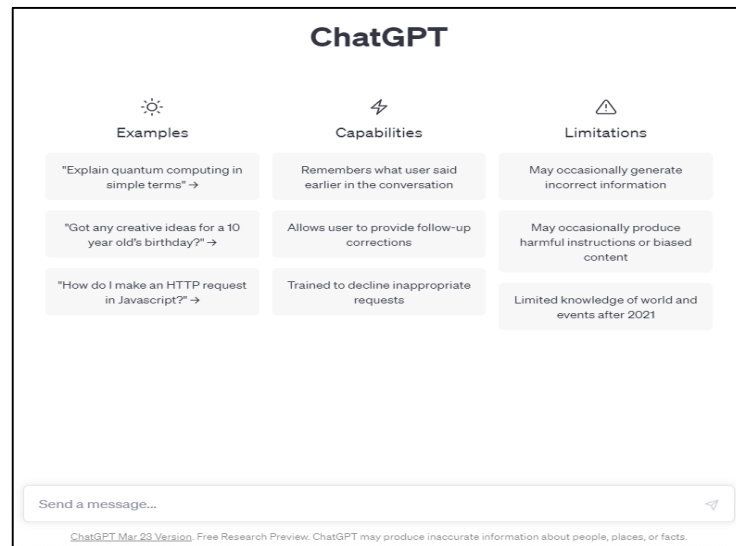
@kuki_ai

인공지능의 현재

▶ 인공지능 챗봇

▶ 챗 GPT(Open AI, 2022)

- ▶ 오픈에이아이(OpenAI, openai.com)가 개발한 대화 전문 인공지능 챗봇
- ▶ 사람의 피드백을 활용한 강화학습을 사용해 인간과 자연스러운 대화를 나누고 질문에 대한 답변도 제공
- ▶ 대화의 주제는 지식정보 전달은 물론 창의적 아이디어에 대한 답변 및 기술적 문제의 해결방안 제시
- ▶ 대화의 숨은 맥락을 이해하거나 이전의 질문 내용이나 대화까지 기억해 답변에 활용하는 등 기존의 챗봇과는 확연한 차이



인공지능의 현재

▶ 일상에서의 AI 기반 서비스



인공지능의 현재

▶ 휴머노이드 로봇 '소피아(Sophia)'

- ▶ 2016년 한슨 로보틱스(Hanson Robotics)가 개발한 휴머노이드 로봇



2017년 UN 무대에 선 소피아

<https://www.youtube.com/watch?v=qNoTjrgMUcs>

▶ 로봇 개 '스팟(Spot)'

- ▶ 보스턴 다이내믹스(Boston Dynamics) 개발, 2020. 6월 상용화



<https://youtu.be/fn3KWM1kuAw>

인공지능의 현재

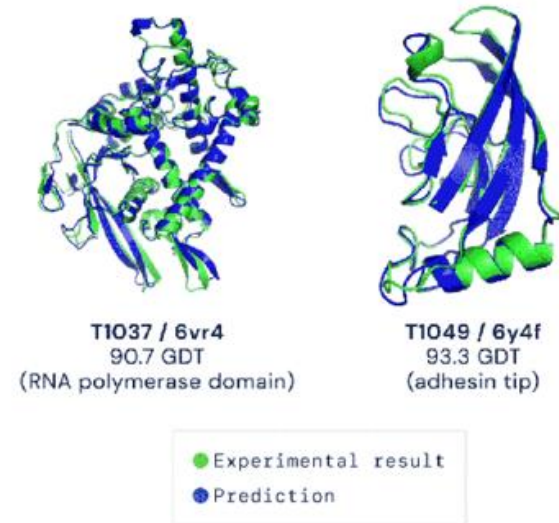
▶ AI 의사



https://www.youtube.com/watch?v=qTqI9_ix_k&t=2s

▶ 알파 폴드

- ▶ 구글 딥마인드의 단백질 구조 예측 AI
- ▶ 알파폴드2가 코로나 바이러스 단백질



[이미지 출처] <http://www.aitimes.com/news/articleView.html?idxno=134544>

인공지능의 현재

▶ 모창 AI



<https://youtu.be/7JR2ehHia04>

▶ 이미지 동영상 합성 AI



<https://youtu.be/hqZhH9Qr4B0>

인공지능의 현재

▶ AI 판사

- ▶ 에스토니아 민사 재판에 AI판사 시범도입



https://www.youtube.com/watch?v=3-Gx32wsf_g&t=1s

▶ AI 변호사

- ▶ 인공지능 '알파로' 변호사와 법률 자문대결 승리



https://youtu.be/qY_5y_zAzTg

인공지능의 현재

▶ 범죄 예방 AI

▶ 안면인식 기술

- 2017년 중국이 AI 기술을 활용한 범죄 용의자 식별 기술 도입



<https://www.youtube.com/watch?v=DrPpSOE4pVo>

▶ 오디오 몽타주 AI

▶ 음성 인식을 통한 이미지 생성



<https://www.youtube.com/watch?v=5x1VvVDyvHE>

인공지능의 현재

▶ 뉴럴링크

- ▶ 인간의 뇌와 인공지능을 연결시켜 인간지능을 증강시키기 위한 기술



https://www.youtube.com/watch?v=feie00oQyo_k

▶ 자율주행 자동차

- ▶ 미국 아마존 자율주행 택시



<https://www.youtube.com/watch?v=8NlnX6350Xw>

인공지능 기술이 우리의 직
업에 어떻게 활용이 될 수
있을까요?

■ AI 이미지 생성

✓ <https://dream.ai/create>

- 텍스트를 입력하고 아트 스타일을 선택하면 이미지를 생성해 준다.

✓ <https://aigreem.com/>

- 텍스트를 입력하면 AI 이미지를 생성해준다.