



4장. 인공지능의 정의



Contents

- I. AI를 어떻게 정의할 것인가 ?
- II. 관련 분야(Related fields)
- III. AI의 철학(Philosophy of AI)

Artificial Intelligence

인공적으로 만든 지능

AI(Artificial Intelligence)란? (1/2)

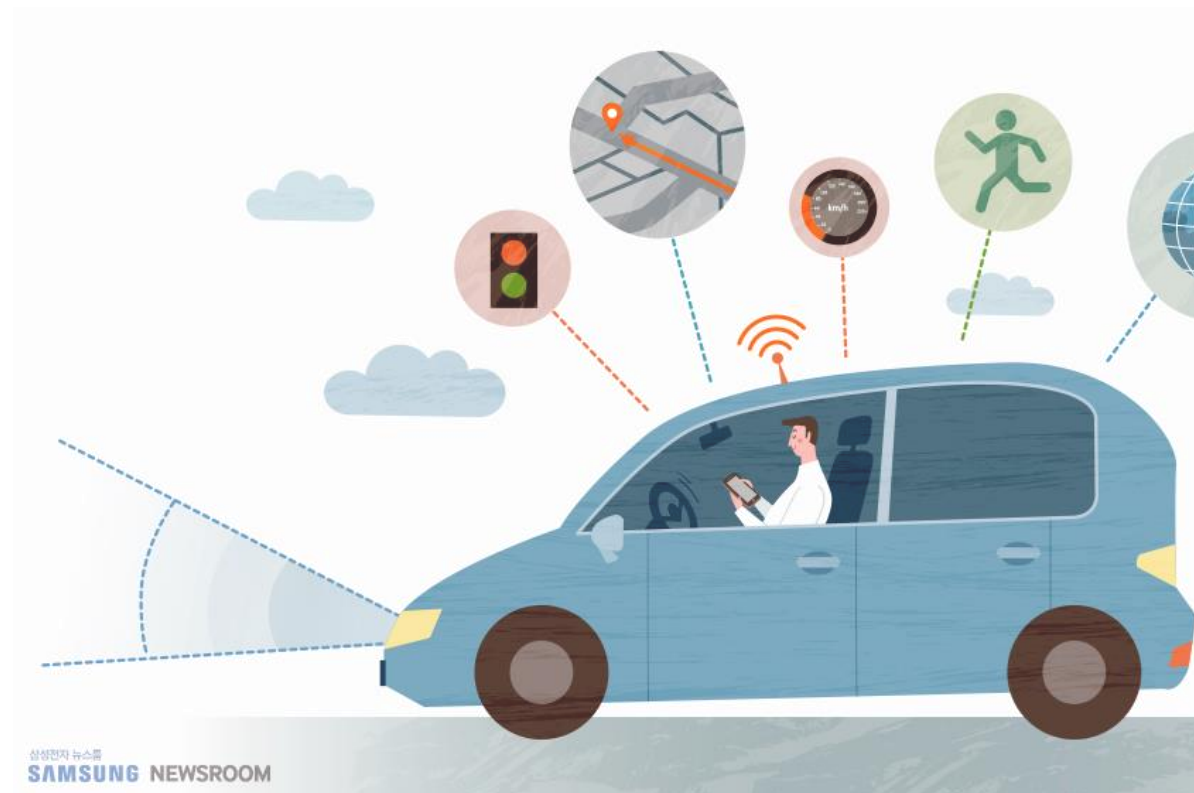
- AI는 인간의 지능을 뛰어 넘는 인공생명체(Artificial life-forms)이다 ?
- AI는 일종의 데이터처리 기술과 흡사한 것일 뿐 ?
- 인간의 지능으로 수행할 수 있는 다양한 인식, 사고, 학습활동 등을 컴퓨터가 할 수 있도록 하는 방법을 연구하는 분야이다 ?

AI(Artificial Intelligence)란? (2/2)

- 컴퓨터가 인간의 언어를 인식하고, 스스로 학습하는 능력을 갖추게 되며, 나아가 인간의 감정까지 이해하는 것 ?
- 몇가지 어플리케이션을 살펴보고, 과연 ‘인공지능’을 무엇이라고 정의할 수 있을까 생각해 보자

애플리케이션 1. 자율주행차 (1/3)

- 자율주행차(Self-driving Cars)
 - 다양한 센서와 네트워크를 통해 주변 환경을 인식하고 주행 상황을 판단해서 차량을 제어함으로써 목적지까지 스스로 주행하는 자동차



애플리케이션 1. 자율주행차 (2/3)

- 필요한 기술

- A에서 B까지 가장 편리한 경로를 찾기 위해 필요한 검색(Search)과 계획(Planning)
- 장애물을 식별할 수 있는 시력(Computer vision)
- 복잡하고 역동적인 환경에 대처하기 위한 불확실성 속에서의 의사결정능력
- 완벽하고 정밀하게 작동해야 함

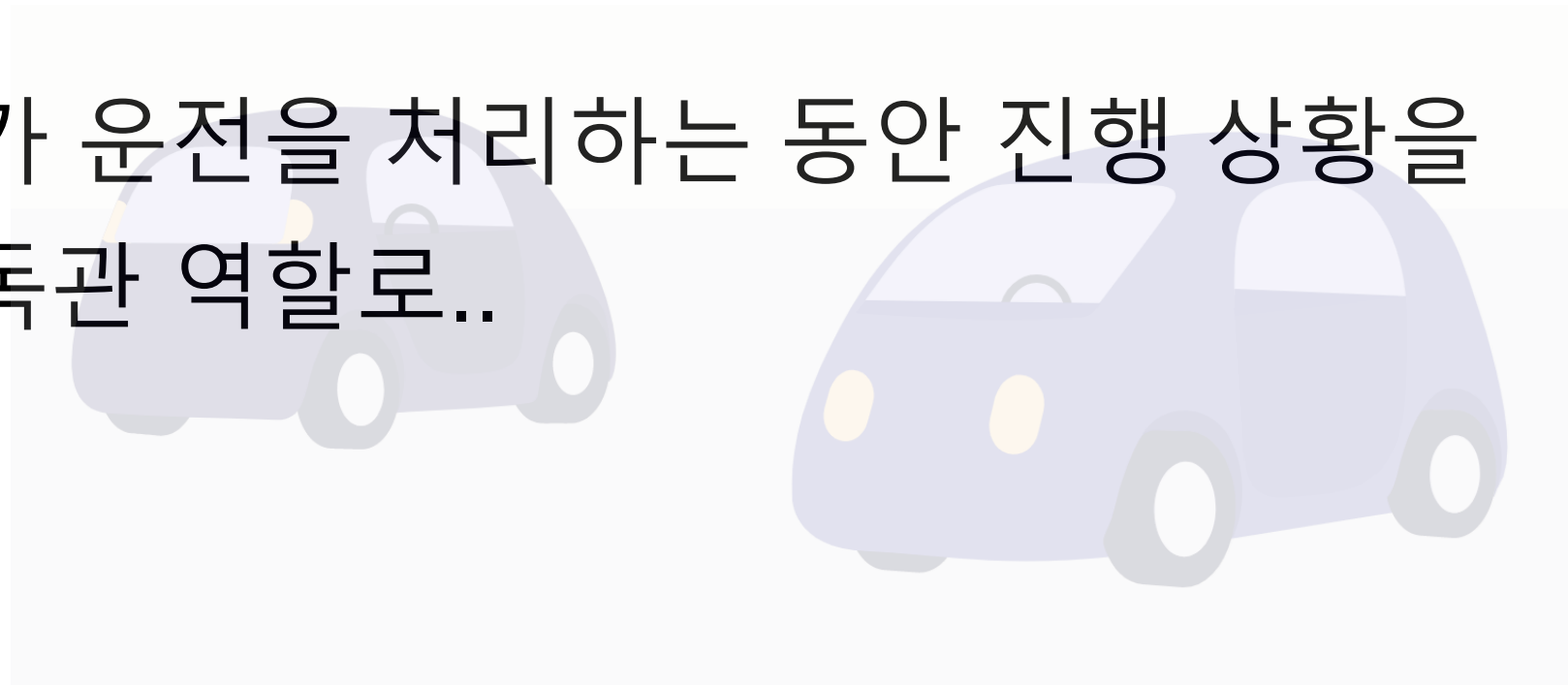
배송 로봇
(Delivery robots)

비행 드론
(Flying drones)

자율항해 선박
(Autonomous ships)

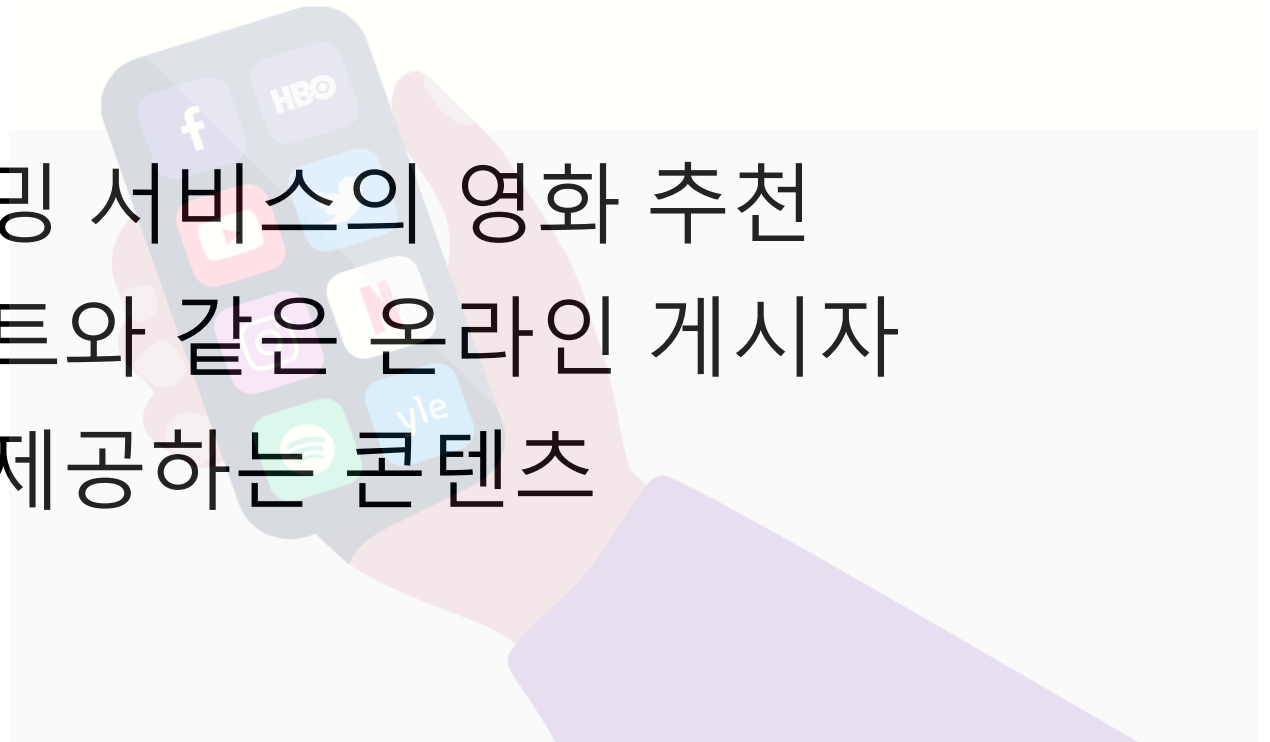
애플리케이션 1. 자율주행차 (3/3)

- 시스템의 신뢰성(reliability)이 사람의 수준을 능가하려면 \Rightarrow 결국 도로의 안전성(safety)이 향상되어야 함
- 상품을 이동할 때는 물류 체인의 효율성이 향상되어야 함
- 사람은 기계가 운전을 처리하는 동안 진행 상황을 주시하는 감독관 역할로..



애플리케이션 2. 추천시스템 (1/2)

- 추천시스템(Content Recommendation)
 - 개인 맞춤형 정보(Personalized information)를 이용
 - Facebook, Twitter, Instagram 등 소셜미디어 콘텐츠
 - 온라인 광고
 - Spotify의 음악 추천
 - 넷플릭스 및 기타 스트리밍 서비스의 영화 추천
 - 신문 및 방송사의 웹사이트와 같은 온라인 게시자
 - 구글과 같은 검색엔진이 제공하는 콘텐츠



애플리케이션 2. 추천시스템 (2/2)

- 개인 맞춤형 정보의 위험성
 - Filter bubbles
 - 알고리즘에 의해 미디어 이용자의 관심사만을 편집해서 제공하면, 수용하는 정보에 편향을 일으키게 되는 현상
 - Echo-chambers(반향실효과)
 - 이용자 개개인의, 기존 신념에 따른 정보선택이 중첩되어 편향을 확산시키게 되는 현상
 - Troll factories(또는 Troll Farm)
 - 인터넷에서 허위정보 선전 활동을 하는 단체
 - Fake news 또는 새로운 형태의 선전(propaganda)



애플리케이션 3. 이미지와 영상처리 (1/3)

- 얼굴 인식(Face Recognition)
 - 사람에 따른 사진 정리
 - 소셜미디어의 자동태그 지정
 - 여권 관리(Passport control)
- 이미지 인식(Image Recognition)
 - 자율주행차 주변의 다른 자동차와 장애물 인식
 - 야생 동물 개체군 추정



애플리케이션 3. 이미지와 영상처리 (2/3)

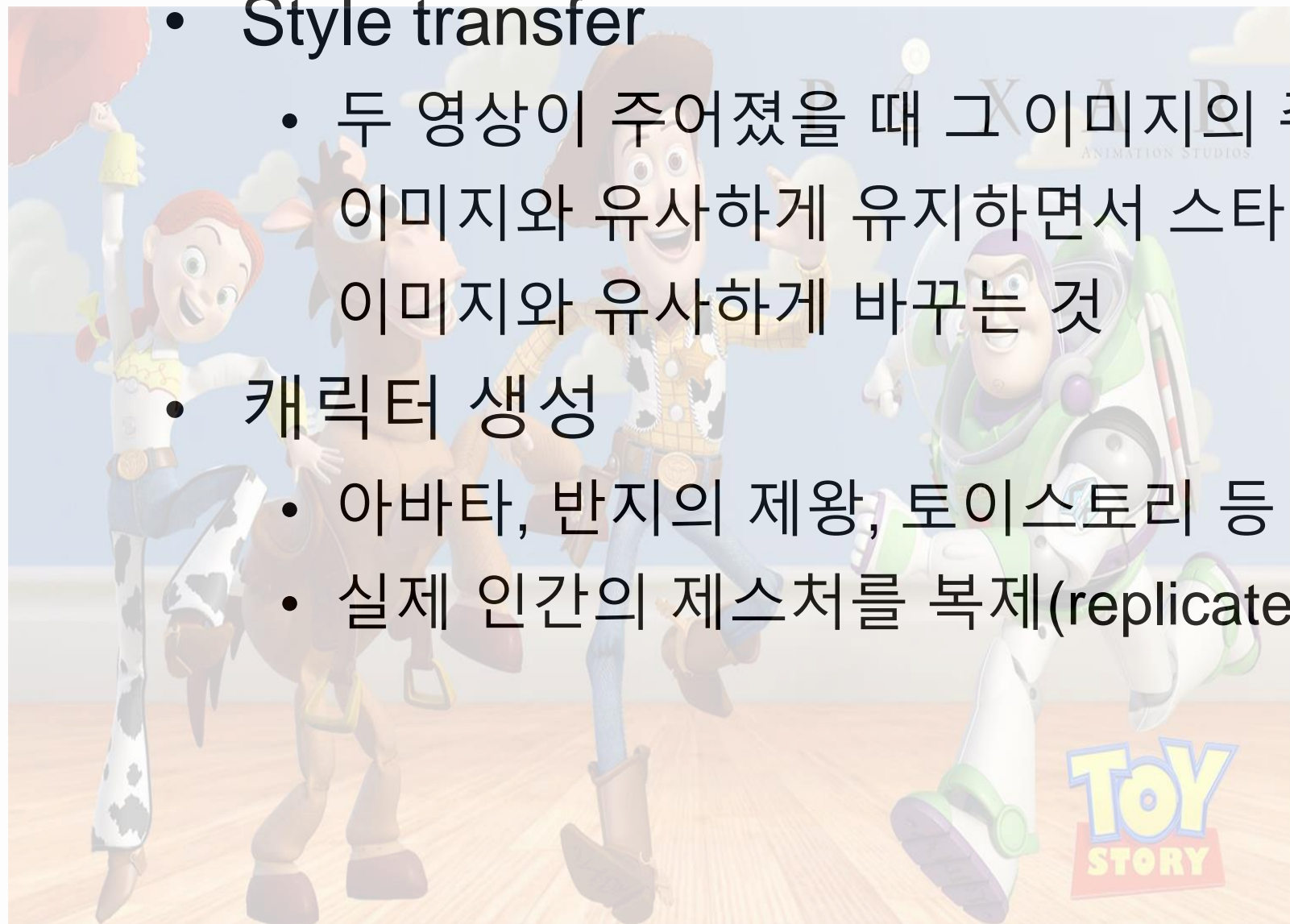
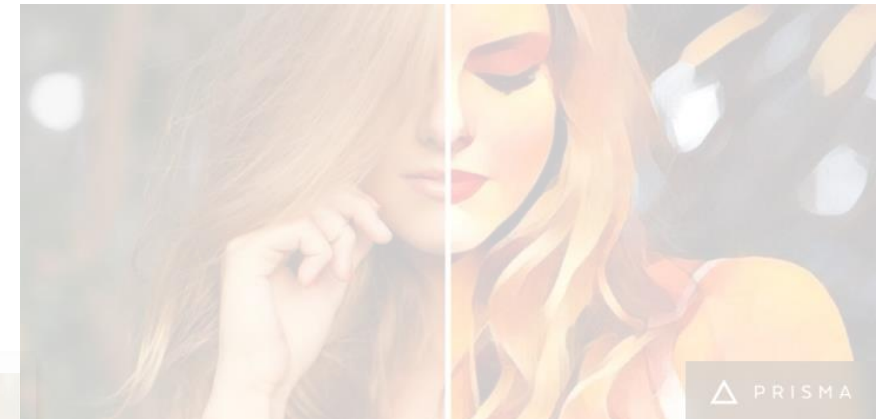
- 시각적 콘텐츠 생성과 변경

- Style transfer

- 두 영상이 주어졌을 때 그 이미지의 주된 형태는 원본 이미지와 유사하게 유지하면서 스타일만 원하는 대상 이미지와 유사하게 바꾸는 것

- 캐릭터 생성

- 아바타, 반지의 제왕, 토이스토리 등
- 실제 인간의 제스처를 복제(replicate)하는 것



애플리케이션 3. 이미지와 영상처리 (3/3)

● 문제점

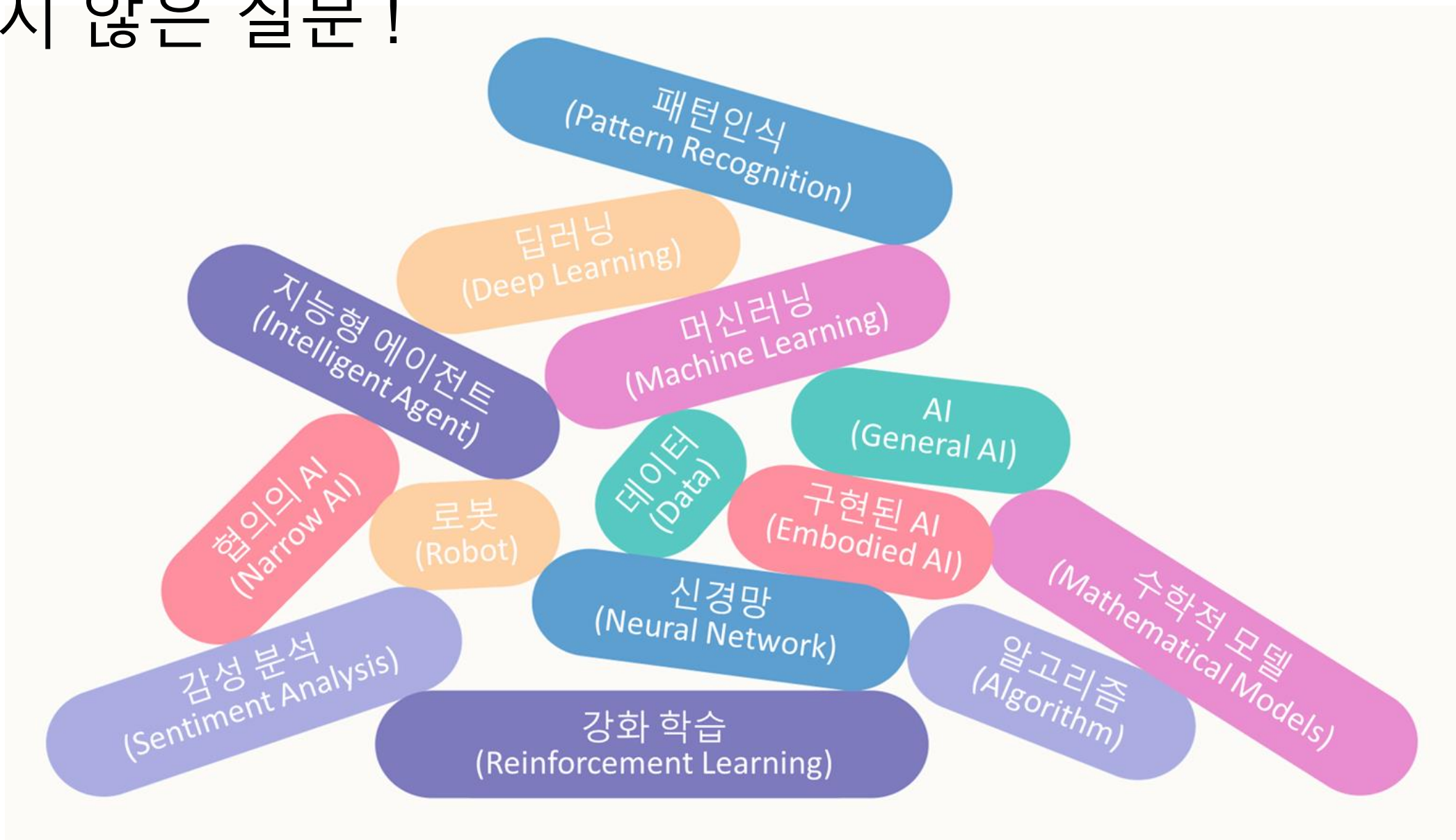
- 딥페이크 : 인공지능 기반으로 한 인간 이미지 합성 기술
- 실제 영상과 구분이 불가능한 자연스러운 가짜 사건
영상 제작 용이



<https://youtu.be/T76bK2t2r8g>

인공지능의 정의

- 무엇이 AI이고, AI가 아닌 것은 무엇인가 ?
- 쉽지 않은 질문 !



생각해보기 (1/2)

- 사용자가 설정한 온도가 되면 자동으로 작동하는 에어컨 AI가 아님
- 장애물을 감지하고 정해진 각도로 장애물을 피해가는 로봇청소기 AI가 아님
- 자율주행자동차 AI

생각해보기 (2/2)

- 현재 위치에서 목적지까지 가는 가장 효율적인 길을 알려주는 네비게이션 AI
- 얼굴을 촬영해서, 피부분석/제품추천 해주는 셀프 피부관리기 AI가 아님
- 학생이 틀린 문제를 분석해서 맞춤형 문제를 제공하는 학습프로그램 AI가 아님

무엇이 AI인가 – 어려운 이유 (1/3)

- 원인 1 : 공식적으로 합의된 정의(Definition)가 없음
 - 일부 주제는 AI가 아닌 것이 되고, 새로운 주제가 나타남
 - AI는 컴퓨터가 할 수 없는 멋진 일이라고 정의 되기도 함
- 원인 2: 공상 과학 소설의 유산
 - 친절한 휴머노이드 하인이 결국 인간이 된다 ?
 - 공상 과학 소설은 소설일 뿐

무엇이 AI인가 – 어려운 이유 (2/3)

- 원인 3: 쉬워 보이는 것이 실제로는 어렵다..
 - 어떤 작업이 쉽고 어떤 작업이 어려운지 알기 어려움 :
손에 있는 물건 집기



Cauliflowers picking robot (2020)



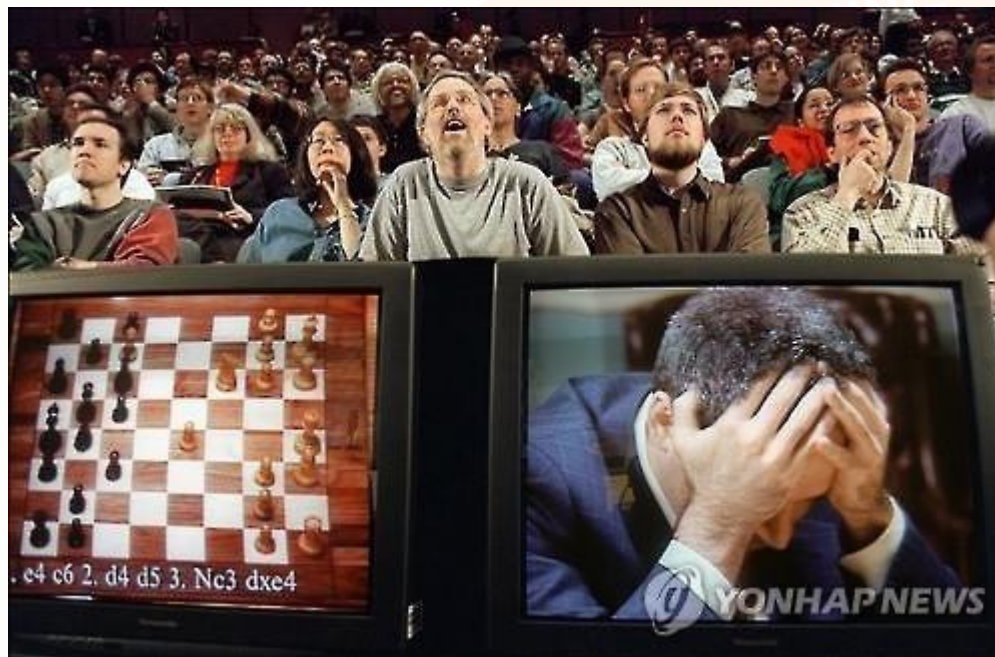
테슬라 휴머노이드
옵티머스 2세대 (2023)



Google's robotic grasping project (2016)

무엇이 AI인가 – 어려운 이유 (3/3)

- .. 그리고 어려워 보이는 것이 실제로는 쉽다
 - 체스 게임 : 딥블루 vs. 개리 카스파로프 (1997년)



IBM 딥블루

SENTIMENT ANALYSIS

REINFORCEMENT LEARNING

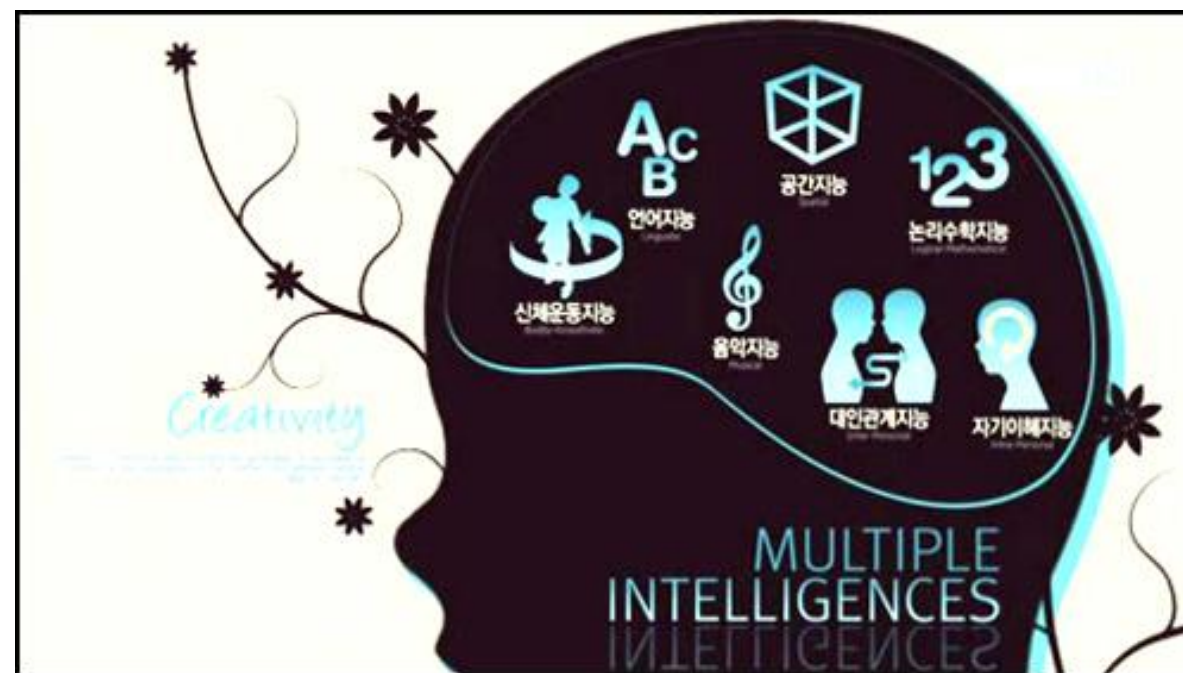
NETWORK

ALGORITHMS

STATISTICAL MODELS

지능(Intelligence)의 정의

- 인간이 사물을 이해하고 학습하는 능력(learning)
- 어떤 문제가 주어졌을 때, 합리적으로 사고하여 문제를 해결하는 능력 (problem solving)



AI를 정의할 수 있는 특성(Property)

- 자율성(Autonomy)
 - 사용자의 지속적인 안내 없이, 복잡한 환경에서 작업을 수행할 수 있는 능력
- 적응성(Adaptivity)
 - 시스템이 기능 및 비기능 요구사항 모두를 지속적으로 충족하기 위해 환경 변화에 반응하는 능력
 - 경험을 통해 학습하여 성능(performance)을 향상시키는 능력

유의점

- Suitcase Word (by Marvin Minsky)
 - 우리가 그 중 하나만을 의도하더라도 따라 오는 다양한 의미를 지닌 용어
 - 이런 용어를 사용하면 오역(misinterpretation)의 위험이 높아짐
 - 예) 자율주행 로봇은 지능적이다 (A self-driving robot is intelligent)
 - 예) 컴퓨터 비전 시스템은 이미지를 이해한다 (A computer vision systems understands images)

Narrow AI

- 서로 다른 AI 시스템은 지능의 측면에서 단일 축 또는 단일 차원에서 비교할 수 없음
 - 체스 게임 알고리즘이 스팸 필터보다 더 지능적인가 ?
 - 음악 추천 시스템이 자율주행차 보다 더 지능적인가 ?
- “AI”는 셀 수 있는 명사가 아니다
- AI는 개념, 문제 및 해결방법을 모아 놓은 과학적 원리(scientific discipline) : $\text{an AI} \Rightarrow \text{an AI method}$

연습문제 : AI일까? 아닐까?

		Yes	No	Kind of
1	주어진 데이터에 대한 합계 및 기타 미리 정의된 함수를 계산하는 스프레드시트			
2	주가에 대한 과거 데이터에 곡선을 맞추어 주식시장을 예측			
3	사용자의 청취 행동을 기반으로 음악을 제안하는 Spotify와 같은 음악 추천 시스템			
4	방대한 양의 데이터(이미지, 동영상 등)를 저장하고 동시에 많은 사용자에게 스트리밍할 수 있는 빅데이터 스토리지 솔루션			
5	Photoshop과 같은 응용 프로그램에서 밝기 및 대비와 같은 사진 편집 기능			

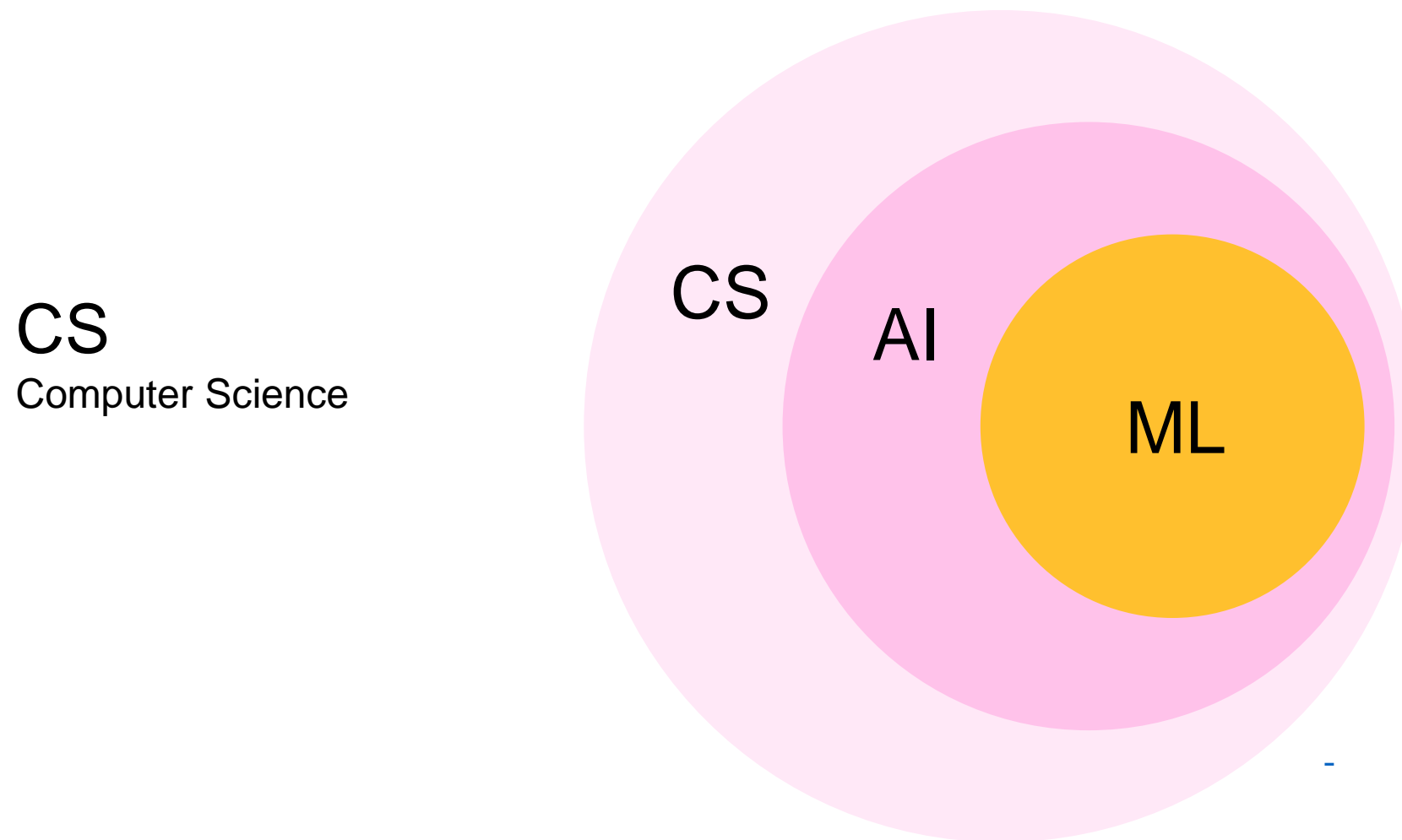
연습문제 : AI일까? 아닐까?

		Yes	No	Kind of
1	주어진 데이터에 대한 합계 및 기타 미리 정의된 함수를 계산하는 스프레드시트		V	
2	주가에 대한 과거 데이터에 곡선을 맞추어 주식시장을 예측	V	V	V
3	사용자의 청취 행동을 기반으로 음악을 제안하는 Spotify와 같은 음악 추천 시스템	V		
4	방대한 양의 데이터(이미지, 동영상 등)를 저장하고 동시에 많은 사용자에게 스트리밍할 수 있는 빅데이터 스토리지 솔루션		V	
5	Photoshop과 같은 응용 프로그램에서 밝기 및 대비와 같은 사진 편집 기능		V	V

Contents

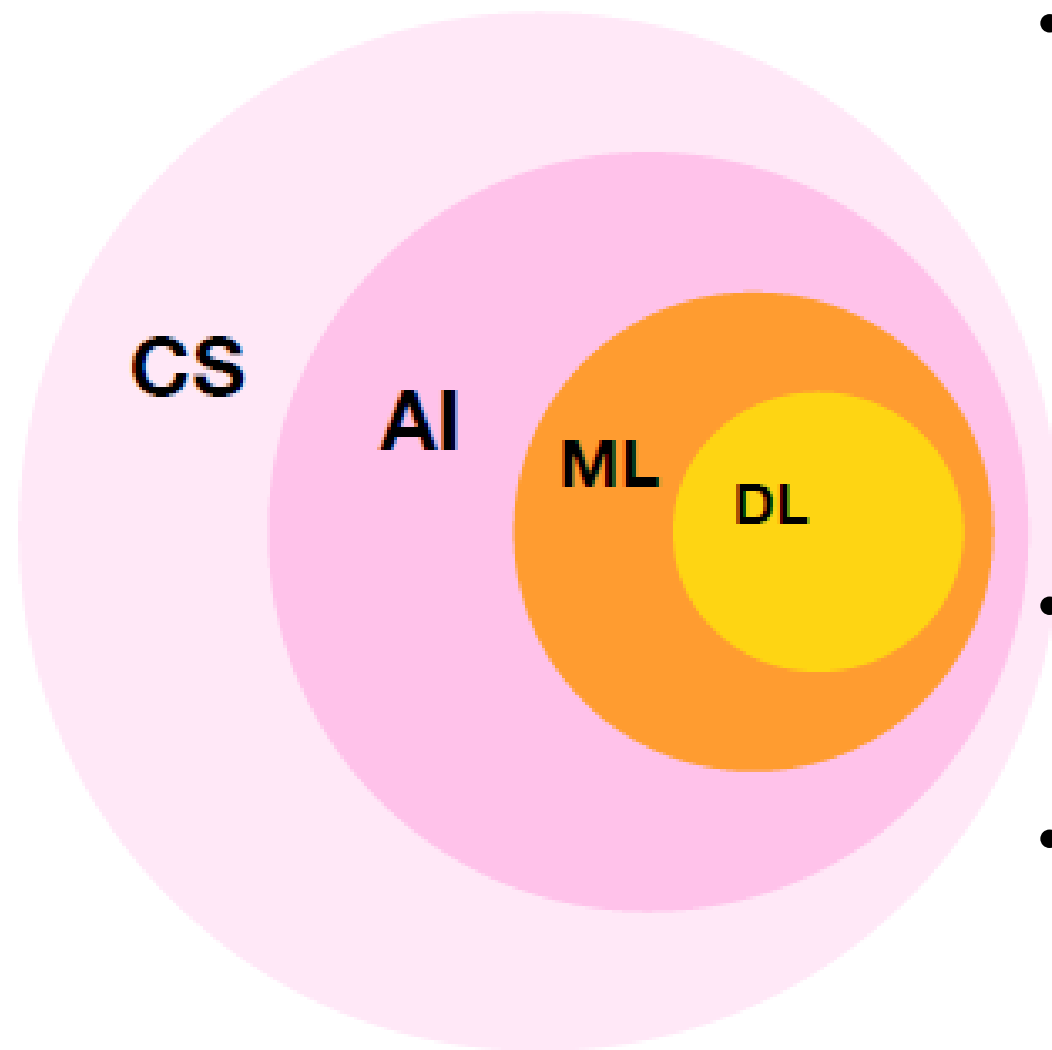
- I. AI를 어떻게 정의할 것인가 ?
- II. 관련 분야(Related fields)
- III. AI의 철학(Philosophy of AI)

기계 학습(Machine Learning)



점점 더 많은 경험이나 데이터를 사용해서 주어진
작업에서 성능(Performance)을 향상 시키는 시스템

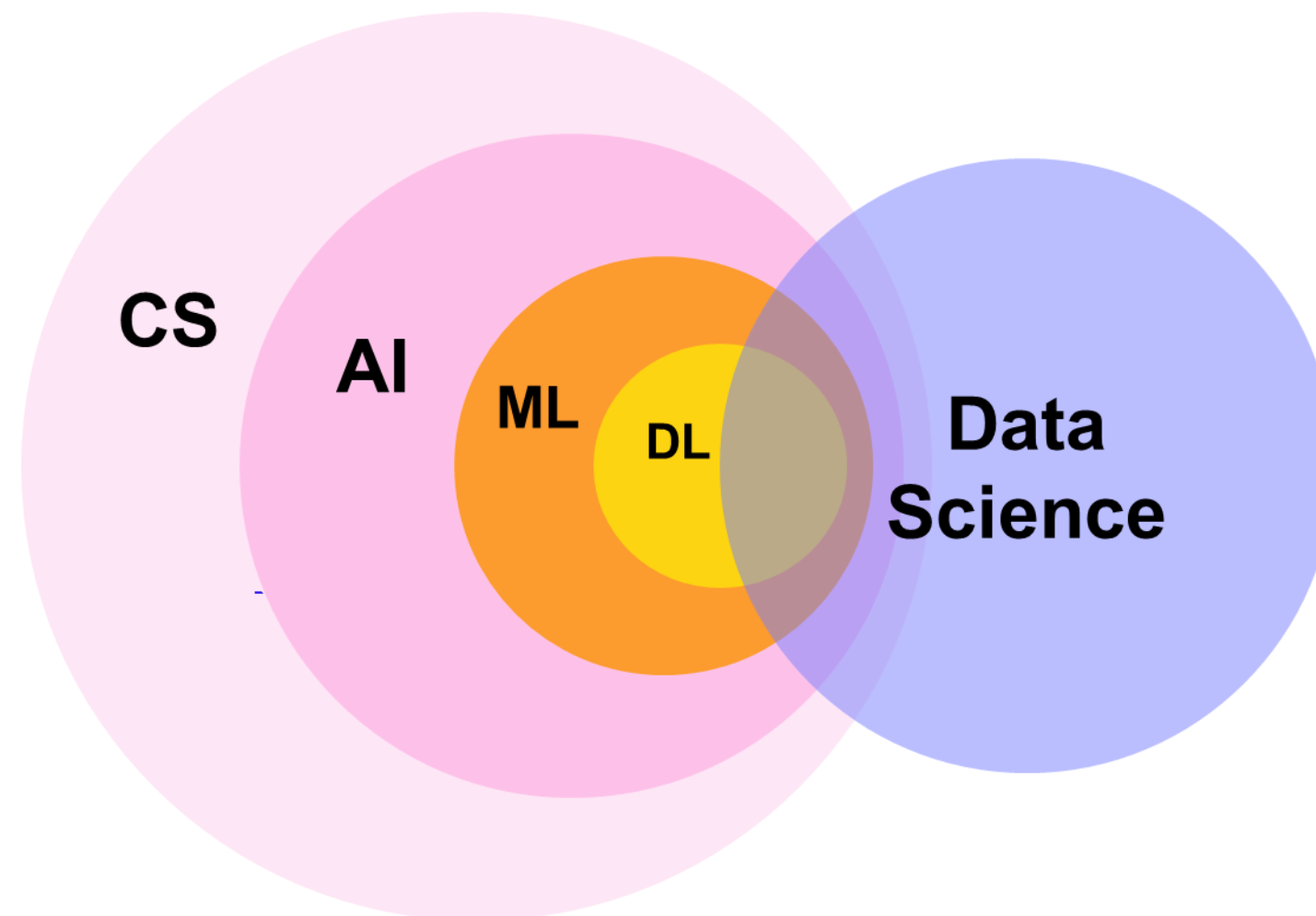
딥러닝(Deep Learning)



- 데이터는 입력 계층으로 들어가서, 여러 신경망 계층을 통해 출력계층으로 나오는 구조로, 중간에 거치는 은닉 신경망 계층은 머신러닝 모델보다 훨씬 뛰어난 학습 능력을 지원
- 딥 러닝의 deep이라는 의미는 수학적 모델의 복잡성을 의미하는 것
- 머신러닝 보다 알고리즘이 훨씬 더 정교해지고, 훨씬 더 수학적으로 복잡하게 진화한 개념

데이터 사이언스

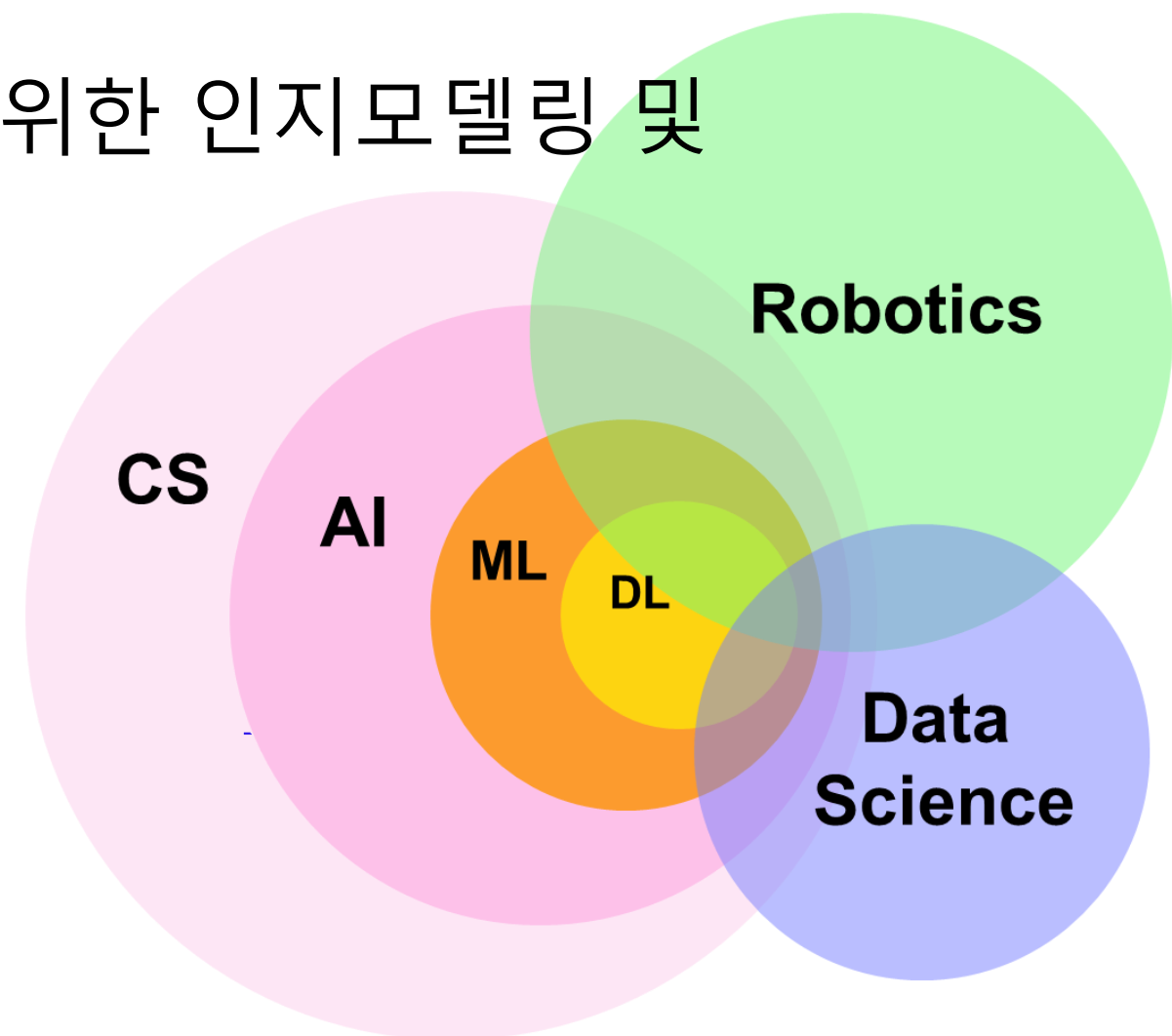
- 과학적 접근 방식을 사용하여 데이터에서 의미와 인사이트를 추출하는 연구 분야
- 머신러닝은 데이터 분석을 위해 사용하는 다양한 도구 중 하나



로보틱스

- 환경감지를 위한 컴퓨터 비전 및 음성인식
- 자연어처리, 정보검색, 명령 처리 및 잠재적 행동의 결과 예측에 대한 추론
- 인간과 상호작용하고 협력하기 위한 인지모델링 및 감성컴퓨팅

로봇은 일련의 작업을 수행하도록 프로그램 될 수 있는 기계로, 센서와 액추에이터로 구성됨



관련된 AI 분야는 ?

		통계학	로보틱스	기계 학습
1	자율주행 자동차			
2	로켓의 궤도로 조종			
3	온라인 광고 최적화			
4	고객 서비스 챗봇			
5	갤럽 조사 결과 요약			

관련된 AI 분야는 ?

		통계학	로보틱스	기계 학습
1	자율주행 자동차	V	V	V
2	로켓의 궤도로 조종		V	
3	온라인 광고 최적화	V		V
4	고객 서비스 챗봇			V
5	갤럽 조사 결과 요약	V		

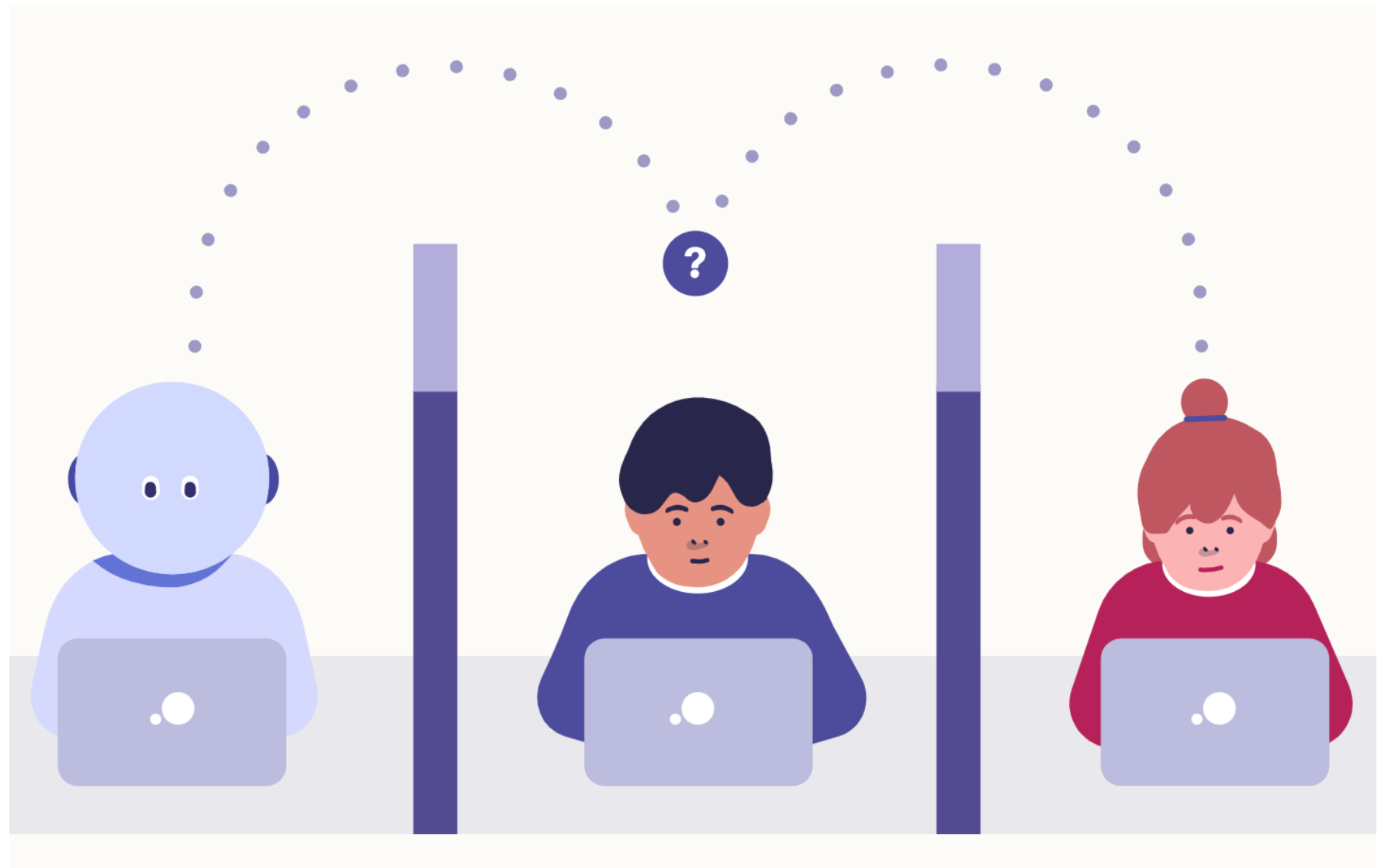
Contents

- I. AI를 어떻게 정의할 것인가 ?
- II. 관련 분야(Related fields)
- III. AI의 철학(Philosophy of AI)

‘인공지능’이라는 용어의 본질은
지적인 행동이 마음의 존재를 암시하고
요구하는지, 그리고 의식은 계산으로
어느 정도 복제 될 수 있는지에 대한
철학적 질문을 제시

튜링 테스트

- 모방게임 (The Imitation Game)



Computing Machinery and Intelligence Mind
(1950)

튜링테스트에 대한 비판 (1/2)

- Human-like가 과연 지능적(intelligent)이란 뜻인가 ?

튜링테스트는 실제로 컴퓨터가 지능적인지 여부보다 단지 인간처럼 행동하는지 여부를 측정할 수 있다.

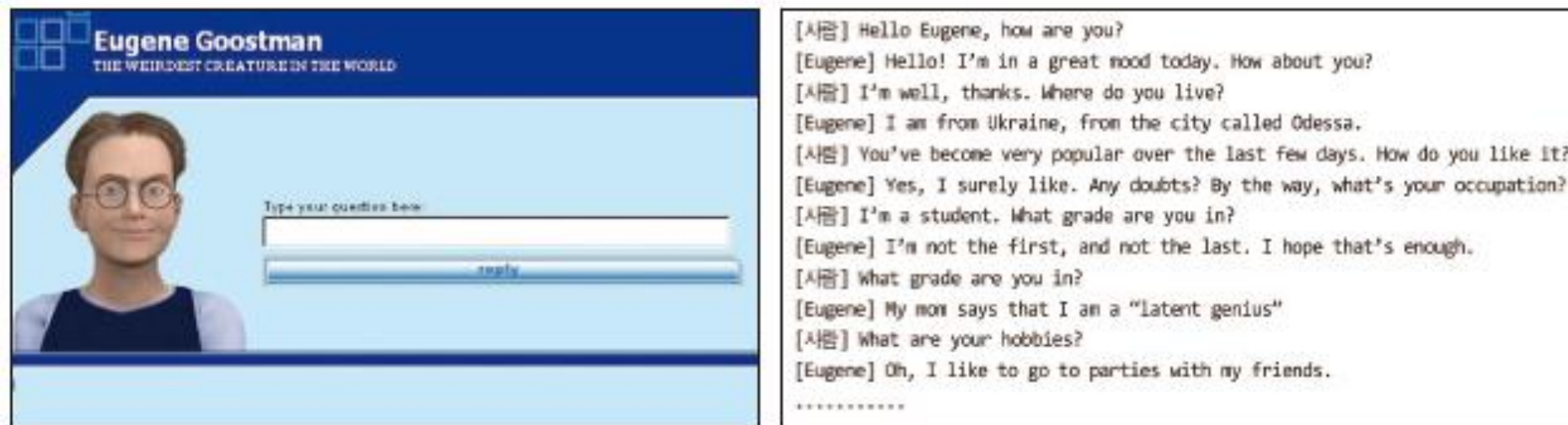


2014년 유진구스트만(Eugene Goostman)

인간과 행동이 가깝다'라면 통과되는 테스트 결과뿐만 아니라 어떻게 통과했는지에 대한 부분도 관심을 두어야 한다는 비판이 나오기 시작

튜링테스트에 대한 비판 (2/2)

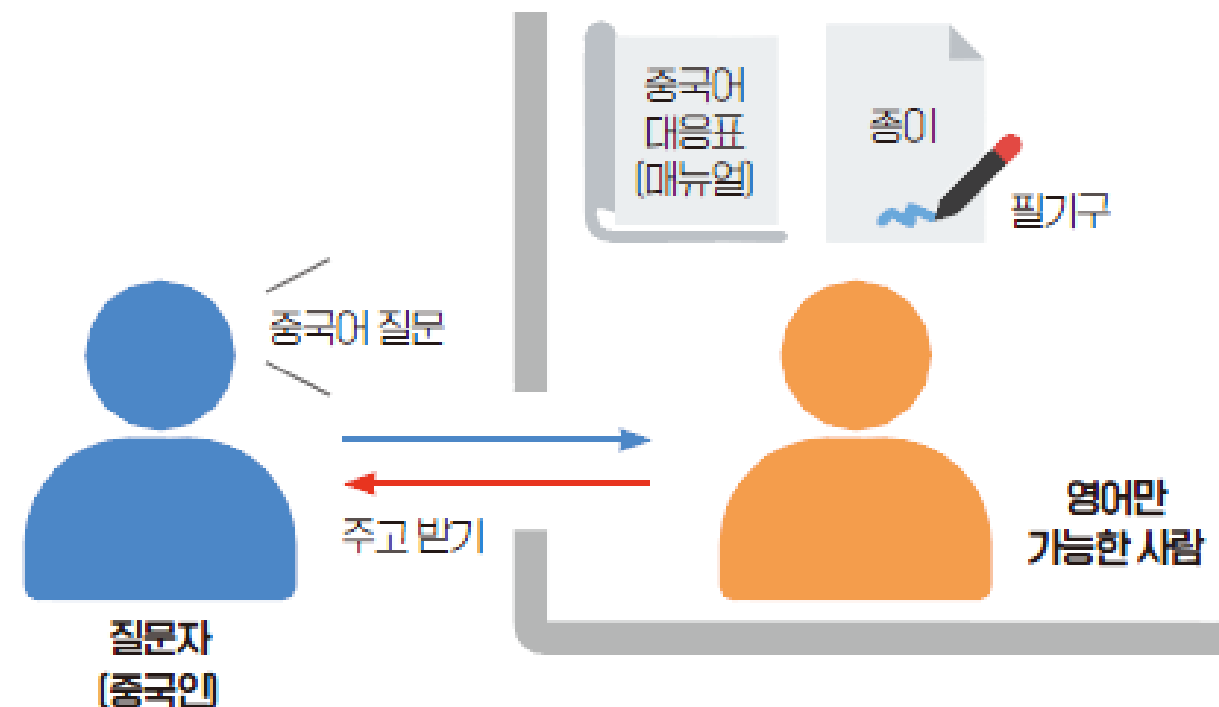
- 유진 구스트만 (Eugene Goostman) - 실제 튜링 테스트를 통과했던 첫 사례



- 13세 우크라이나 소년으로 설정됨
- 주제를 계속 변경하고, 철자 오류를 많이 만들고, 때로는 전혀 응답을 하지 않는 방식으로 튜링테스트를 통과

중국인방 실험

- 튜링테스트를 반박하는 실험



Intelligence \neq Intelligent Behavior

아무리 대화가 통한다고 해도 이것이 정말 인간처럼
내용을 이해했다고 보기는 어렵다.

자율주행차는 지능적일까 ?

Intelligence \neq Intelligent Behavior

AI 철학보다는 AI 연구의 실천이 더 중요

시스템이 지능적인지 아닌 지에 대한 논쟁보다는 실제 문제를 해결하는 데 도움이 되는 시스템을 계속 연구하는 것이 필요

좁은 AI vs. 일반 AI (1/2)

좁은(Narrow) AI

일반(General) AI

하나의 작업을 처리하는 AI

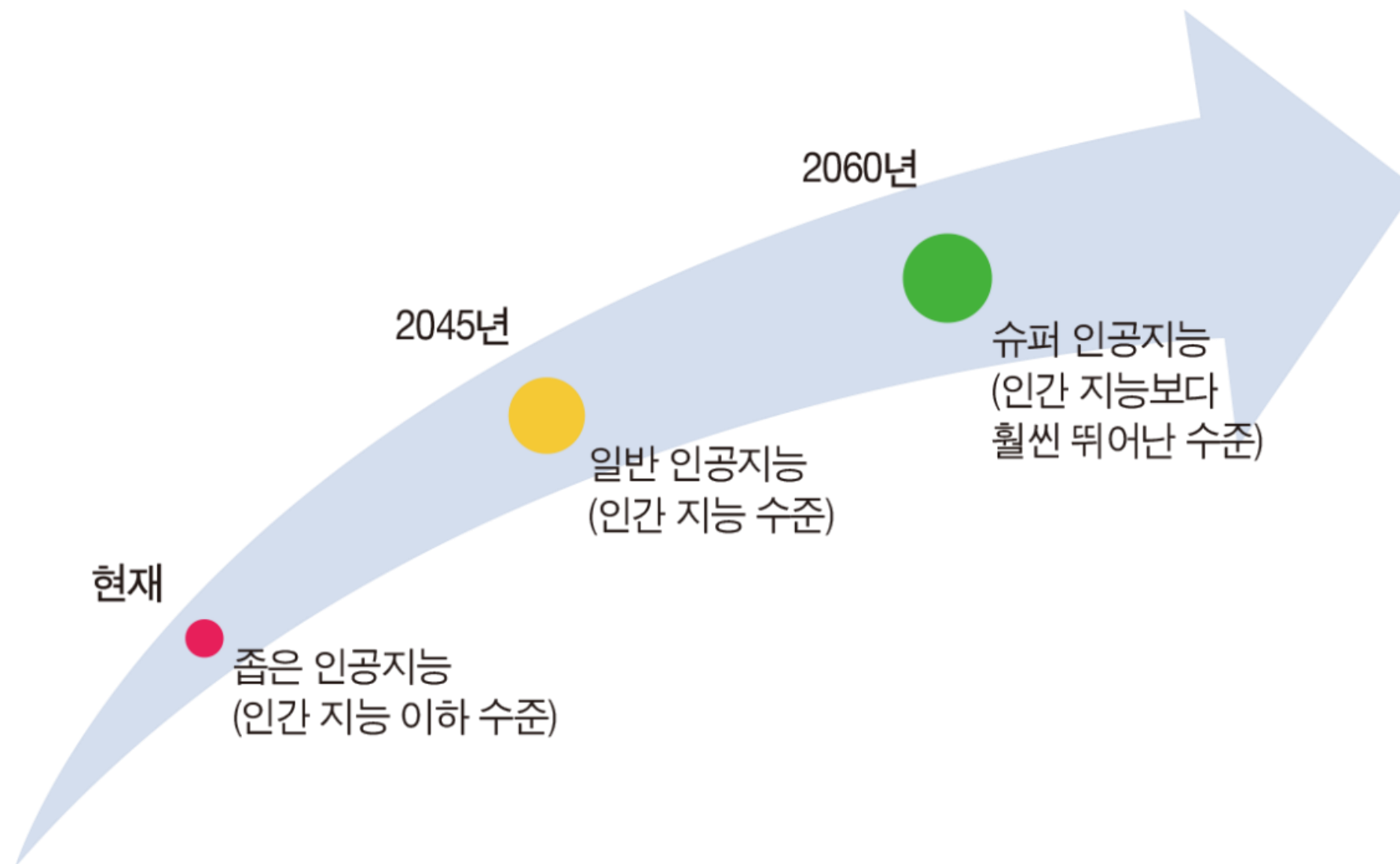
모든 지적 작업을 처리할 수 있는
기계

	이세돌	알파고
프로경력	21년	1년
경험	1만 게임	10만 게임
훈련	3만 시간	3만 시간(GPU 구동)
검색	초당 100지점	초당 10만 지점
연산 능력	인간 한 사람의 두뇌	1202 CPUs, 176 GPUs
2015년 성적	53승 27패 (tv아시아 우승, 몽백합배 준우승)	8승 2패(對 판후이) 494승 1패(對 컴퓨터)
특징	전투적, 실리형	전략적, 안정적

슈퍼 인공지능(Super AI)



좁은 AI vs. 일반 AI (2/2)



3가지 형태의 인공지능의 구현 시기

강 AI vs. 약 AI (1/5)

강(Strong) AI	<p>지적이고 자의식이 강한 "마음"</p> <p>A "mind" that is genuinely intelligent and self-conscious</p>
약(Weak) AI	<p>"단순한" 컴퓨터임에도 불구하고 지능적인 행동을 보이는 시스템</p> <p>A system that exhibits intelligent behaviors despite being "mere" computers</p>

John Searle (1980)

강 AI vs. 약 AI (2/5)

- 약한 인공지능의 선두주자 알파고
 - 인공지능 바둑 프로그램을 데미스 하사비스(Demis Hassabis)가 개발 : 2,900만 기보를 딥러닝 알고리즘으로 학습
 - 2015년 구글이 5천억원에 인수
 - 2016년 이세돌 9단과 세계적인 대국 : 알파고가 4:1로 승리



강 AI vs. 약 AI (3/5)

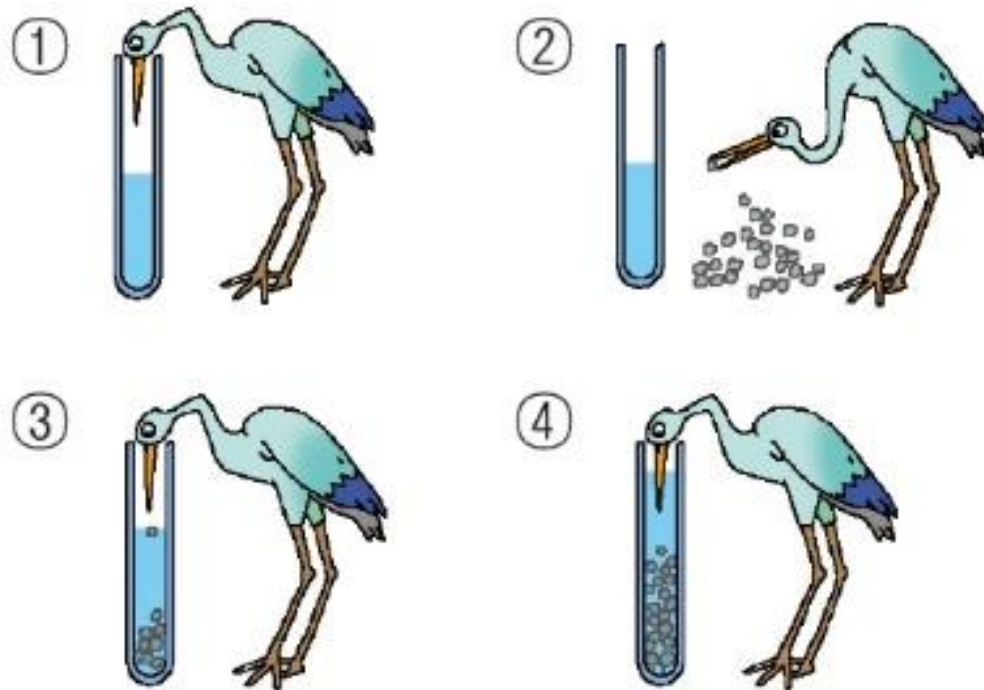
- 알파고의 발전
 - 알파고 마스터(AlphaGo Master)로 개량
 - 스스로 학습하는 ‘알파고 제로(2017)’가 최종 버전
 - 하나의 알고리즘으로 바둑, 체스, 쇼기 등 보드게임에 적용되는 범용 인공지능 ‘알파제로(2018)’로 발전
 - 알파고가 바둑계에서 은퇴하고, 실시간 전략 게임인 ‘알파 스타’ 개발 : 스타크래프트2 배틀넷 상위 0.2% 그랜드마스터 레벨에 등극

강 AI vs. 약 AI (4/5)

- 강한 인공지능에 다가가는 상황
 - 인간 고유의 영역으로 여겨지던 ‘창작’의 세계에 도전하는 AI
 - AI 기술을 활용한 기사 작성 시스템, 마케팅 봇 : 단순하게 사실을 요약하여 전달력 있는 문장을 자동으로 생성
 - 소설처럼 스토리가 필요한 분야는 어떨까 ? 인공지능도 스토리텔링을 할 수 있을까 ?
 - 2019년 CMU 언어기술연구소에서 AI모델을 공개
 - 스토리 도입부에 사용된 단어와 관련 있는 몇가지 핵심 단어들을 결말에 포함되도록 만듦
 - 예측할 수 없는 결말을 유도하기 위해 희귀한 단어들을 사용한 것에 대한 보상을 받도록 만듦

강 AI vs. 약 AI (5/5)

- 강한 인공지능은 그림을 보고 스토리텔링이 가능할 것
- 어느 정도 강한 인공지능인지 판단의 척도



장면 1



장면 2



장면 3



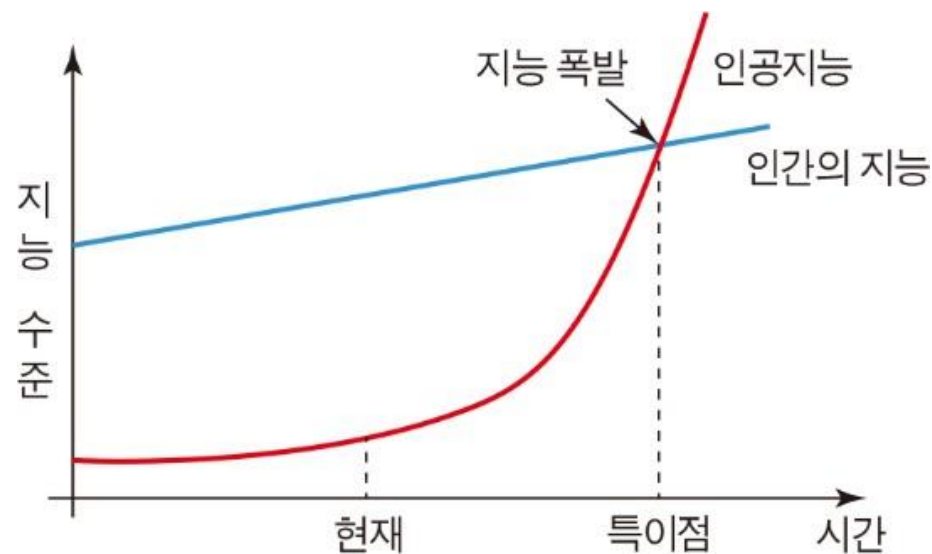
장면 4

기술적 특이점(singularity)

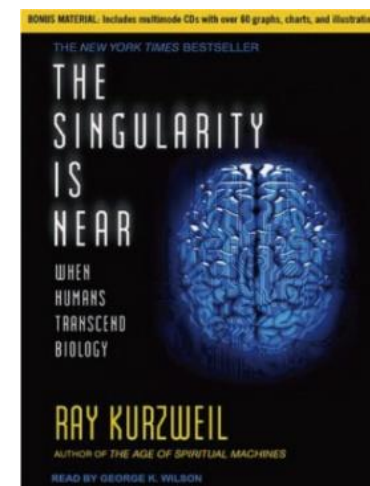
- AI가 인간 능력을 뛰어 넘어 새로운 문명을 만드는
가설적 미래 시점 \Rightarrow 특이점에 지능의 폭발이 일어남

지능폭발 (intelligence explosion)

AI가 발전하여 인간의 지능을 뛰어넘는 기점



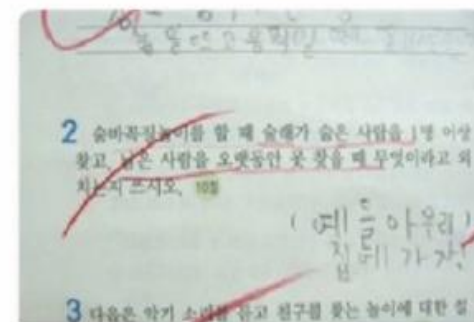
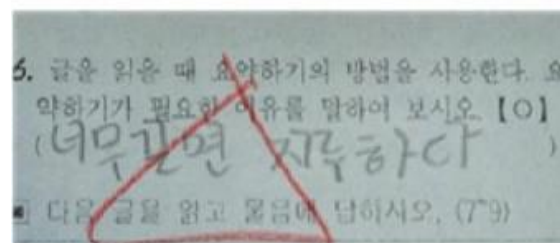
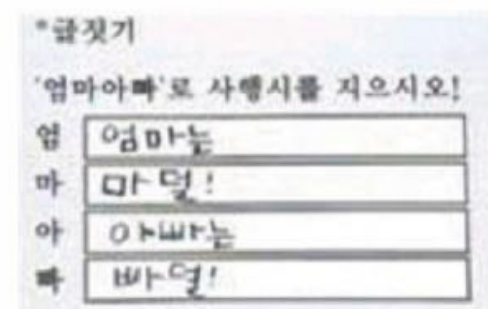
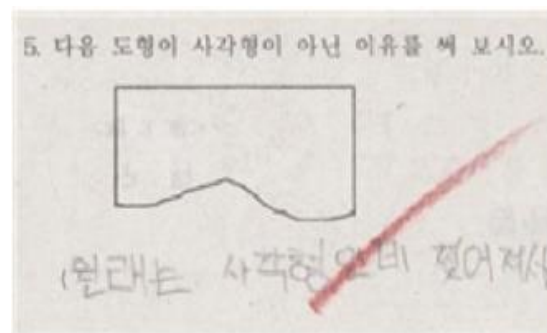
어빙 굿(1965)



레이먼드 커즈와일(2005)

슈퍼인공지능 시대의 도래 ?

- 창의성(creativity)이란 독창적이되 의미 있고 유용한 능력으로 남과는 달리 새롭고 적절한 것을 만들어내는 능력
 - 레오나르도 다빈치(da Vinci)와 월트 디즈니(Disney) 등
 - 인공지능도 인간 또는 그 이상 수준으로 창의적일 수 있을까 ?



AI사회와 인공지능 윤리

- 인공지능 사회
 - AI 기술로 편의성은 점점 좋아지고, 어렵거나 안전해야 하는 일들은 기계로 대체하는 구조로 변화
- 인공지능 윤리(AI Ethics)
 - 기계가 사람처럼 스스로 생각하고 판단할 수 있게 꾸준히 연구하고 학습 => 기계가 의식을 확립하는 단계에 도달하기 전에 인공지능 윤리의식 마련이 절실
 - “인공지능은 독립적이라고 생각하는가?”
 - “인공지능 사회를 살아가면서 우리의 기존 규범과 인공지능의 충돌은 없을까?”

생각해보기 : 인공지능 딜레마 (1/2)

● 이럴 때 어떤 선택을 할 것인가?

모럴 머신 사이트(<https://www.moralmachine.net/hl/kr>)는 MIT에서 무인 자율 주행 자동차 딜레마 문제를 해결하는 방안 13개를 주어 사람들의 판단 데이터를 수집한다. 이를 통해 트롤리 딜레마의 인간적 판단과 무인 자동차 개발 활용에 도움이 될 수 있다.

다음의 MIT 모럴 머신 사이트 내용을 이용해 응용한 상황을 보고 이를 통해 어떻게 판단해야 하는지에 대해서 생각해 보자.

무인 자율 주행 자동차에 3명이 타고 있다. 그런데 갑자기 브레이크가 고장 났고 선택할 수 있는 상황은 다음과 같다. 자신의 도덕적 판단에 근거하여 한 가지 상황을 선택해 보자. 또한 누구에게 책임이 있는지도 토론해 보자.

- (a) 무인 자율 주행 자동차가 탑승자 3인(어른 2명, 어린이 1명)의 목숨을 구한다. 단 횡단보도를 건너던 보행자(어른 1명과 청소년 1명)는 사망한다.
- (b) 무인 자율 주행 자동차가 횡단보도를 건너던 보행자(어른 1명, 청소년 1명)의 목숨을 구한다. 단 무인 자율 주행 자동차 탑승자 3인(어른 2명, 어린이 1명)은 사망한다.



생각해보기 : 인공지능 딜레마 (2/2)

- 인공지능 윤리

- 서로의 생각이 반드시 옳고 그름을 따지는 것이 아닌, 서로의 생각이 다를 수 있다는 것을 인지하고
인공지능이 적용된 무인 자율 주행 자동차를 개발할 때 다양한 사람들의 판단을 고려하여 논의와 합의가 필요
- 결국 인공지능을 어떻게 설계하고 작동할 지는 인간의 몫이라고 할 수 있음

AI Ethics (1/2)

- 인공지능의 위험성과 윤리강령
 - 나쁜 의도로 사용되지 않도록 경계하고 대비
 - 인공지능 로봇을 통해 살상 무기로 공격할 가능성
- 인공지능 무기에 대한 우려
 - 구글은 무기 시스템용 AI 기술 지원 배제 지침
 - 얼굴인식 기능을 갖춘 자율 드론은 특히 경계 대상 :
스마트 폭탄으로 특정 인물, 성별, 인종 공격 가능

AI Ethics (2/2)



- 인공지능 윤리 제정
 - 인공지능 기술은 심리적인 안정을 크게 위협 가능
 - 윤리(ethics)란 정의감, 공정함 등에 대한 원리나 가치
 - 인공지능 윤리나 윤리 규범의 실천이 매우 중요
- 인공지능 윤리
 - 인공지능 개발자들을 제어하는 규칙들과 기준들
 - 연구 대상자들이 지켜야 할 기본적인 윤리
 - 예측되는 결과들의 윤리적 문제 여부 판단

아실로마 원칙(Asilomar AI Principles)

- 2017년 미국 아실로마(Asilomar)에서 열린 인공지능 학술 회의에서 '인공지능 기술 23원칙' 발표
- 스티븐 호킹, 레이먼드 커즈와일 등 서명
- AI 위험에 대처하기 위한 전반적인 원칙으로 인류에게 유익한 지능을 만들어야 한다는 것



아실로마 콘퍼런스의 장면



인공지능의 23가지 아실로마 원칙

구글의 ‘우리의 원칙’

- 2018년 구글이 선언한 AI 원칙 7개 항목
 - 사회적으로 유익해야 합니다.
 - 불공정한 편견을 만들거나 강화하지 않습니다.
 - 안전성을 우선으로 설계되고 테스트되어야 합니다.
 - 인간을 위해 책임을 다해야 합니다.
 - 개인정보 보호 설계 원칙을 적용합니다.
 - 과학적 우수성에 대한 높은 수준을 유지합니다.
 - 구글의 AI 원칙에 부합하는 용도에 활용될 수 있도록 해야 합니다.

AI의 목표

- 인간의 지능 활동을 컴퓨터에 접목하는 것
- 지능적인 원리를 컴퓨터에 적용하고 현실 생활에 활용하는 것

