



Samsung Innovation Campus

| Artificial Intelligence Course

Together for Tomorrow!
Enabling People

Education for Future Generations

Chapter 1.

Introduction to Artificial Intelligence

Artificial Intelligence Course

Chapter Description

◆ Chapter objectives

- ✓ 인공지능의 개념과 관련 학문이 어떤 것이 있는지에 대하여 자세하게 알 수 있도록 한다.
- ✓ 인공지능 트렌드는 어떻게 바뀌고 있는지 관련 산업별 여러 응용 사례를 통해 향후 인공지능 관련 서비스 기획을 할 수 있도록 한다.

◆ Chapter contents

- ✓ Unit 1. The Concept of Artificial Intelligence
- ✓ Unit 2. Applications of Artificial Intelligence
- ✓ Unit 3. Trends in Artificial Intelligence
- ✓ Unit 4. AI 시장
- ✓ Unit 5. Course Roadmap

Unit 1.

The Concept of Artificial Intelligence

- | 1.1 인공지능의 정의
- | 1.2 인공지능의 종류 및 하위 집합
- | 1.3 머신러닝의 정의
- | 1.4 기계학습 관련 학문
- | 1.5 머신러닝 기반 데이터 분석 기법의 유형 및 선택
- | 1.6 머신러닝 기반 데이터 분석 절차
- | 1.7 머신러닝을 사용하는 이유
- | 1.8 머신러닝의 한계

What is Artificial Intelligence

| Definition



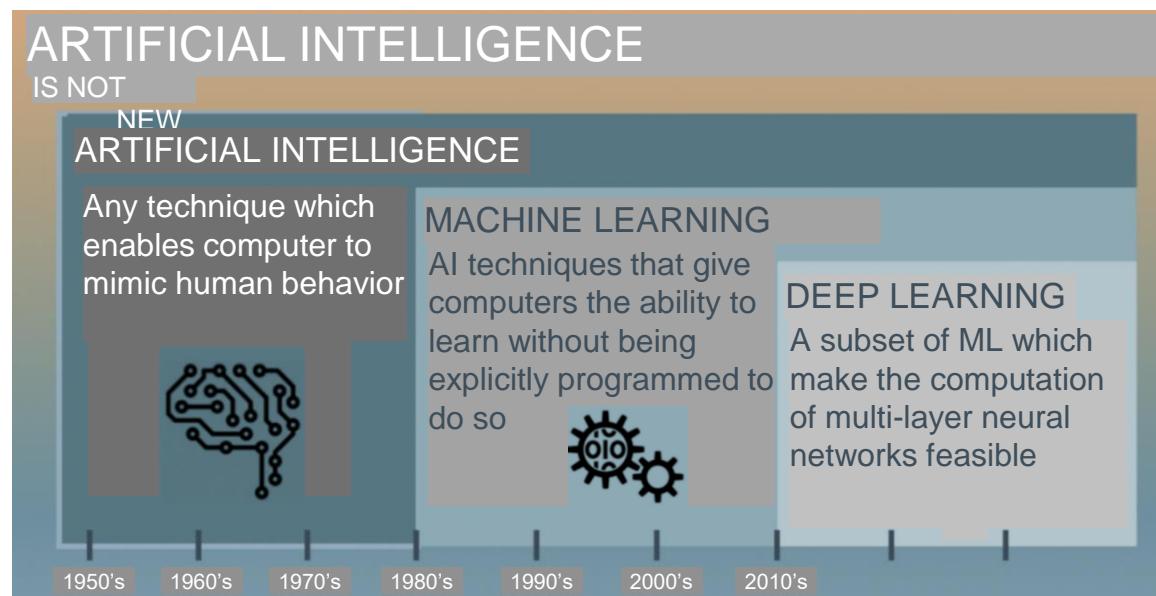
- ▶ **Artificial intelligence (AI)** is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". They work on their own without being encoded with commands. John McCarthy came up with the name "artificial intelligence" in 1955.
- ▶ These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction.

<https://www.wired.com/2011/10/john-mccarthy-father-of-ai-and-lisp-dies-at-84/>

John McCarthy(1927–2011)

| 데이터를 지식으로 바꾸는 지능적인 시스템 구축

- ▶ 현대 기술 시대에는 정형 또는 비정형 데이터가 매우 풍부하다.
- ▶ 20세기 후반에 데이터에서 지식을 추출하여 예측하는 자기 학습(self-learning) 알고리즘과 관련된 인공 지능(Artificial Intelligence, AI)의 하위 분야로 머신 러닝이 출현한다.
- ▶ 사람이 수동으로 대량의 데이터를 분석하여 규칙을 유도하고 모델을 만든다.
- ▶ 머신 러닝이 데이터에서 더 효율적으로 지식을 추출하여 예측 모델과 데이터 기반의 의사 결정 성능을 점진적으로 향상시킬 수 있다.



AI vs. machine learning vs. deep learning

	AI	Machine learning	Deep learning
Optimal data volumes	Varying data volumes	Thousands of data points	Big data: millions of data points
Outputs	Anything from predictions to recommendations to decision-making	Numerical value, like a classification or score	Anything from numerical values to free-form elements, like free text and sound
How it works	Machines are programmed to mimic human activity with human-like accuracy	Uses various types of automated algorithms that learn to model functions and predict future actions from data	Uses neural networks that pass data through many processing layers to interpret data features and relationships
How it's managed	Algorithms require human oversight in order to function properly	Algorithms are directed by data analysts to examine specific variables in data sets	Algorithms are largely self-directed on data analysis once they're put into production

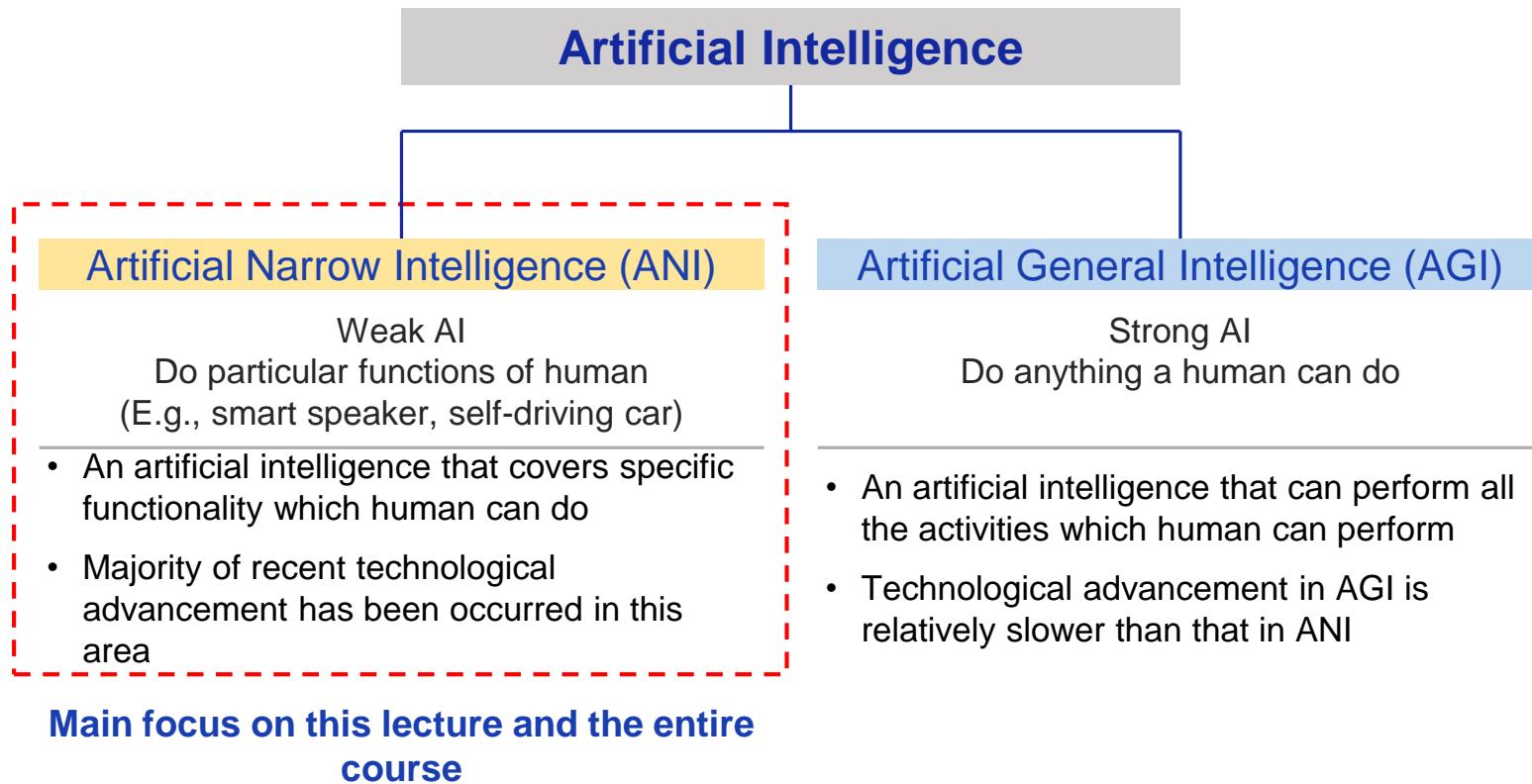
Unit 1.

The Concept of Artificial Intelligence

- | 1.1 인공지능의 정의
- | 1.2 인공지능의 종류 및 하위 집합
- | 1.3 머신러닝의 정의
- | 1.4 기계학습 관련 학문

- | 1.5 머신러닝 기반 데이터 분석 기법
 의
 유형 및 선택
- | 1.6 머신러닝 기반 데이터 분석 절차
- | 1.7 머신러닝을 사용하는 이유
- | 1.8 머신러닝의 한계

Type of Artificial Intelligence



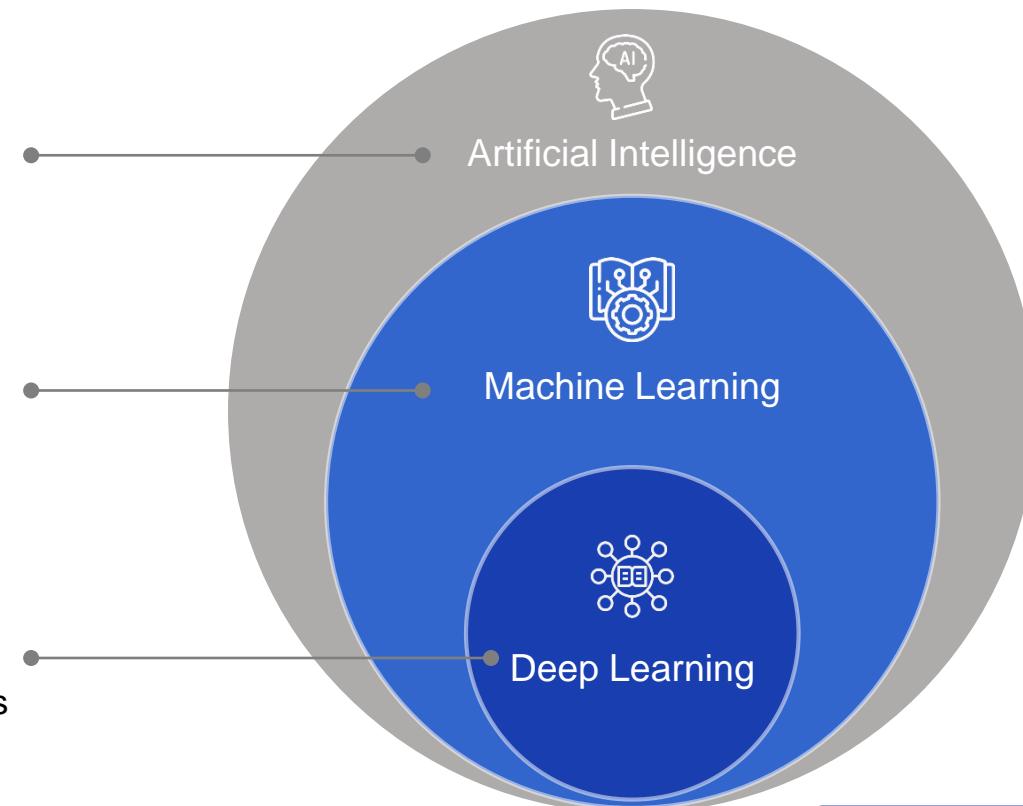
What are subset of AI?

- Subset of Artificial Intelligence

Artificial Intelligence
Any technique which enables computers
to mimic human behavior.

Machine Learning
Subset of AI techniques which use
statistical methods to enable machines to
improve with experiences.

Deep Learning
Subset of ML which make the
computation of multi-layer neural networks
feasible.



Source: KD Nuggets

Unit 1.

The Concept of Artificial Intelligence

| 1.1 인공지능의 정의

| 1.2 인공지능의 종류 및 하위 집합

| 1.3 머신러닝의 정의

| 1.4 기계학습 관련 학문

| 1.5 머신러닝 기반 데이터 분석 기법
의

유형 및 선택

| 1.6 머신러닝 기반 데이터 분석 절차

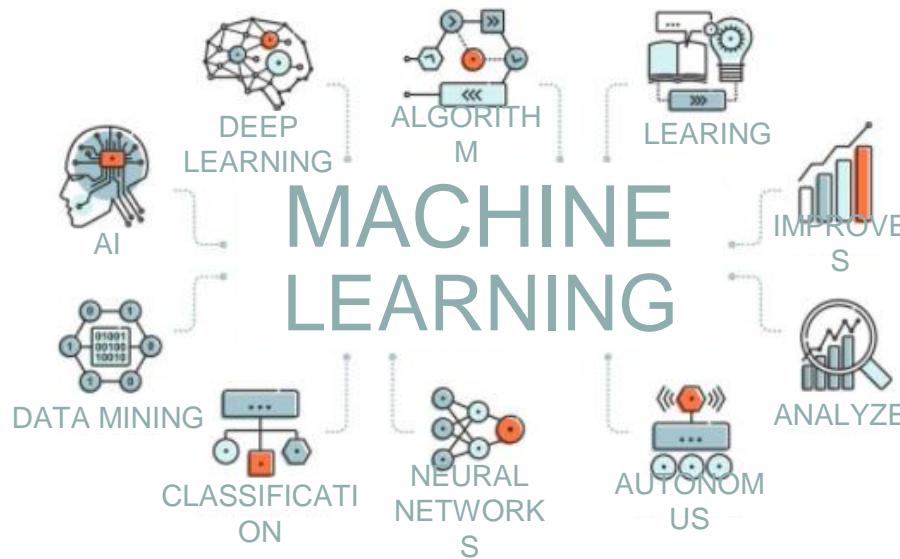
| 1.7 머신러닝을 사용하는 이유

| 1.8 머신러닝의 한계

머신 러닝 (machine learning)이란?

| 경험을 통해 자동으로 개선하는 컴퓨터 알고리즘의 연구로 인공지능의 한 분야로 간주된다.
컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야이다.

- ▶ **아서 새뮤얼 (Arthur Samuel, 1959)**: “기계가 일일이 코드로 명시하지 않은 동작을 데이터로부터 학습하여 실행할 수 있도록 하는 알고리즘을 개발하는 연구 분야”
- ▶ **톰 미첼 (Tom Mitchell, 1977)**: “어떤 작업 (Tasks)에 대하여 컴퓨터 프로그램의 성능 (Performance)을 측정했을 때 경험 (Experience)으로 인해 성능이 향상되었다면, 이 컴퓨터 프로그램은 작업 T와 성능 측정 P에 대해 경험 E로부터 학습한다.”



■ 머신러닝과 딥러닝

- ▶ 예를 들어 만약 만 장의 개와 고양이 사진이 있다고 가정할 경우 그 사진을 머신(컴퓨터)에게 입력한 후 개와 고양이를 분류한다고 하는 작업이 있다고 하자. 이때의 성능은 얼마나 정확하게 개와 고양이를 분류하는 것이다.



- ▶ T라는 작업 (Tasks)은 개와 고양이를 분류하는 것이고, P라는 성능 (Performance)은 개와 고양이를 정확하게 잘 분류하는 성능 지표를 나타낸다.
- ▶ E라는 경험 (Experience, 즉 데이터(만 장의 사진))을 통해 점차 개와 고양이를 분류하는 성능이 좋아진다고 하면 이 프로그램은 학습을 한다고 말할 수 있다.

Unit 1.

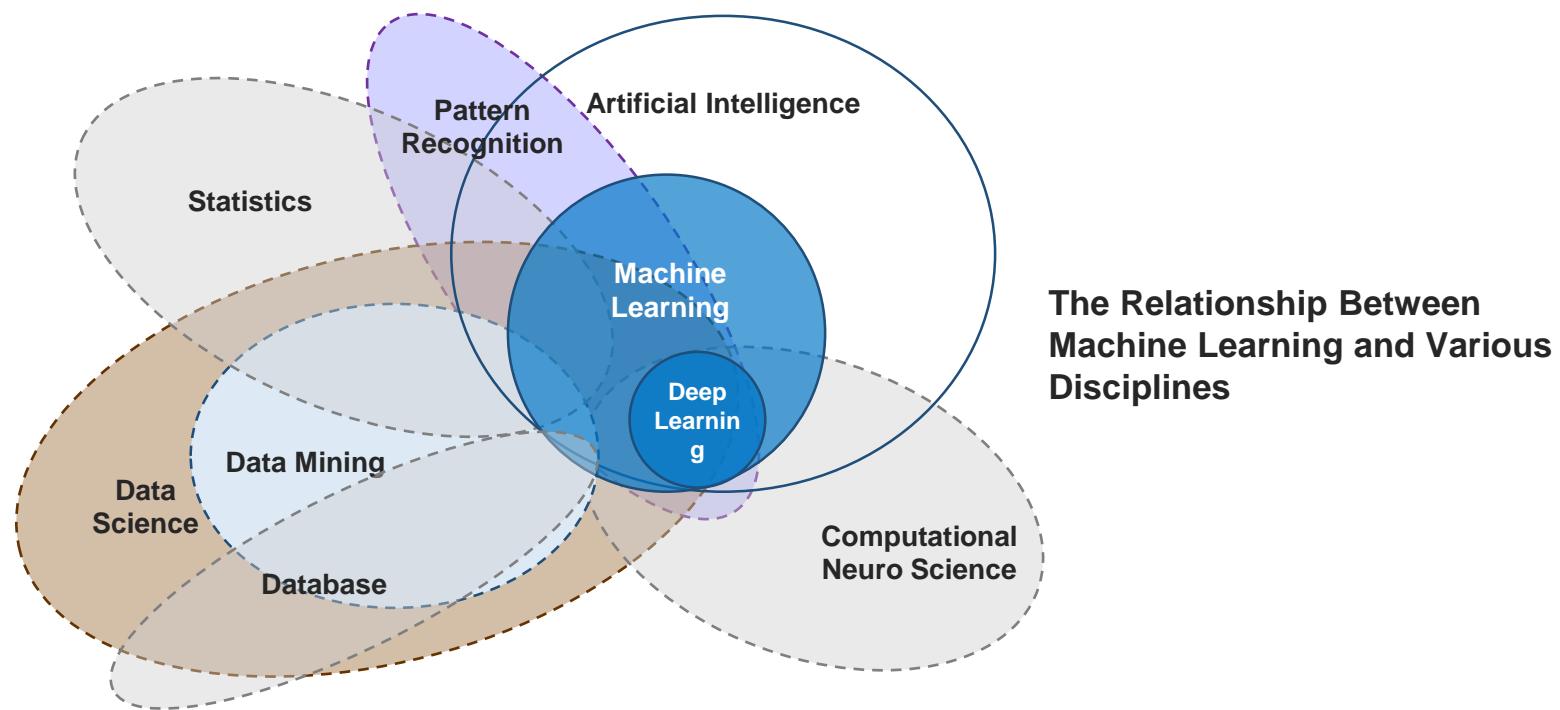
The Concept of Artificial Intelligence

- | 1.1 인공지능의 정의
- | 1.2 인공지능의 종류 및 하위 집합
- | 1.3 머신러닝의 정의
- | 1.4 기계학습 관련 학문

- | 1.5 머신러닝 기반 데이터 분석 기법
 의
 유형 및 선택
- | 1.6 머신러닝 기반 데이터 분석 절차
- | 1.7 머신러닝을 사용하는 이유
- | 1.8 머신러닝의 한계

기계 학습 관련 학문

- 기계 학습은 어느 한 가지 분야의 기술이나 방법론에 국한되기보다는 확률론 및 통계학, 컴퓨터 과학, 데이터베이스론, 인지과학, 신경과학, 패턴인식 등 실로 다양한 분야의 학문적 배경과 성과들이 융합된 학제적 분야라고 할 수 있다.



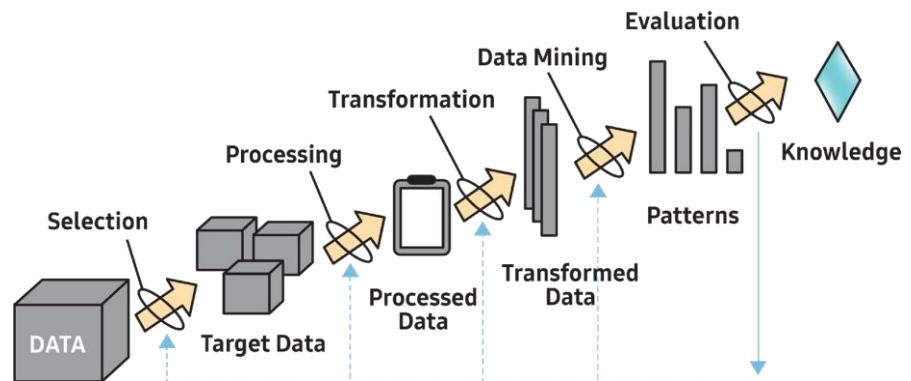
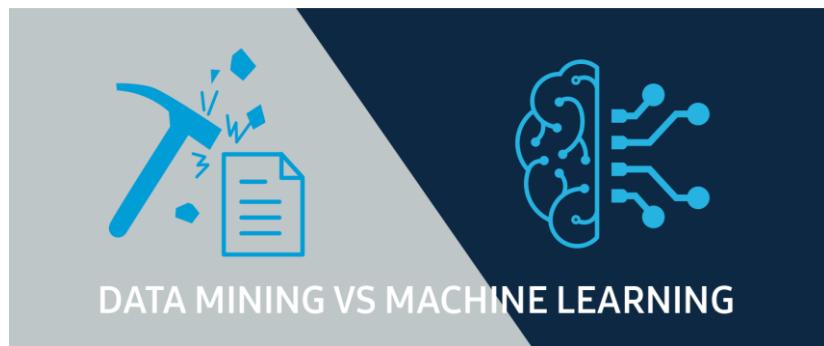
| 기계학습 vs 통계학

- ▶ 통계학은 전통적으로 데이터를 정보로 변환하기 위한 과학적이고 체계적인 방법을 제공하는 이론적 토대라고 할 수 있다. 통계학에서 특히 강조되고 있는 영역은 추론과 검증이며, 주어진 데이터가 연구자의 가설과 이론에 얼마나 부합하는가(혹은 관측치로 부터 도출된 값이 실제 모집단의 모수를 얼마나 정확하게 추정하고 있는가) 등을 설명하기 위해 다양한 방법론과 이론이 구축되어 있다.
- ▶ 한편 기계학습은 명시적인 알고리즘을 설계하기 어렵거나 프로그래밍하기 어려운 작업을 해결하기 위해 주로 사용되며, 대부분의 기계 학습 알고리즘은 먼저 데이터가 생성한 잠재적인 메커니즘의 특징을 파악하여 복잡한 관계를 정량화한 후, 이 식별된 패턴을 사용하여 새로운 데이터에 대한 예측을 실시할 경우 사용된다.
- ▶ 얼핏 생각하면 통계학과 머신러닝의 접근 방법이 서로 정반대라고 생각될 수 있으나, 강조하는 측면이나 관점에 다소 차이가 있을 뿐 근간을 이루는 방법론은 매우 유사하다고 할 수 있다. 실제로 기계학습의 많은 방법론들이 통계학에 기반을 둔 통계적 학습(Statistical Learning)에 근간을 두고 있다고 할 수 있다.



- 기계학습 vs 데이터마이닝

- ▶ 데이터마이닝(Datamining)이란:
 - 대용량의 데이터 창고[광석 더미]로부터 유용한 정보[금 조각]를 캐내는(mining) 작업을 의미함
 - 대용량 데이터에 존재하는 데이터 간의 관계, 패턴, 규칙 등을 찾아내고 모형화해서 기업의 경쟁력 확보를 위한 의사결정을 돋는 일련의 과정이다.
- ▶ 기계학습은 데이터로부터 유용한 규칙, 지식 표현 혹은 판단 기준 등을 추출한다는 점에서 데이터 마이닝과도 관련이 깊다.
- ▶ 데이터마이닝이 주로 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 의미 있는 규칙이나 패턴을 발견하고 이를 지식화 하는 과정이라고 한다면, 기계학습은 주어진 입력 데이터를 컴퓨터 프로그램이 학습하여 예측을 수행하고 스스로의 예측 성능을 향상시키는 과정과 이를 위한 알고리즘을 연구하고 구축하는 기술이라는 점에서 다소 차이가 있다고 할 수 있다.



- 통계분석 vs 데이터마이닝

- ▶ 전통적 통계분석

- 대상집단이 있으며, 모집단의 분포 혹은 모형 등 여러 가지 가정을 전제로 하게 되며 이 전제 조건하에서 분석을 실시

- 표본 (sample)의 관찰을 통해 모집단 (population) 의 모수(parameter) 전체를 추론 (Inference)하는 과정

- ▶ 데이터마이닝

- 표본조사 / 실험에서 필연적으로 수반되는 분포라든가 모형에 대한 전제조건이 필요하지 않음

- 모집단의 전체자료를 이용하여 필요한 정보/지식을 추출하는 과정

- 대용량 자료여야 한다는 전제조건 있음

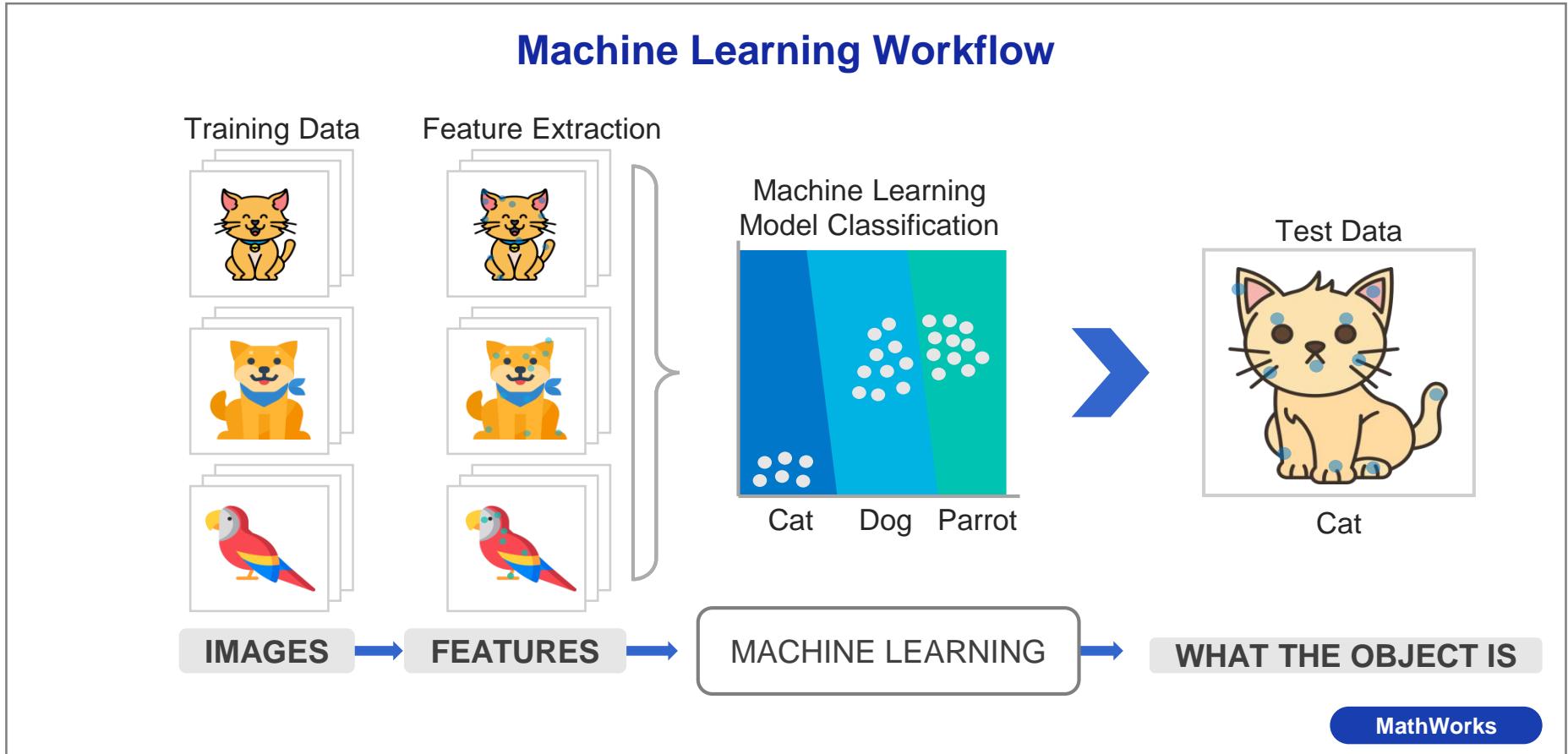


Unit 1.

The Concept of Artificial Intelligence

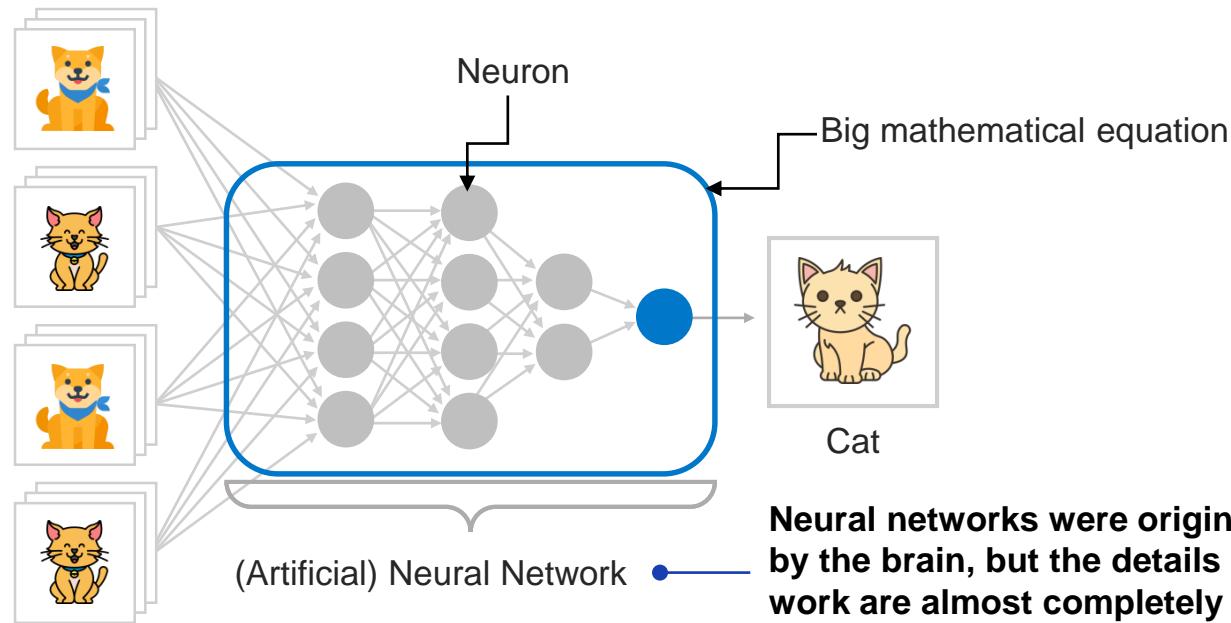
- | 1.1 인공지능의 정의
- | 1.2 인공지능의 종류 및 하위 집합
- | 1.3 머신러닝의 정의
- | 1.4 기계학습 관련 학문
- | 1.5 머신러닝 기반 데이터 분석 기법의 유형 및 선택
- | 1.6 머신러닝 기반 데이터 분석 절차
- | 1.7 머신러닝을 사용하는 이유
- | 1.8 머신러닝의 한계

| 머신러닝 vs. 딥러닝



| 머신러닝 vs. 딥러닝

Deep Learning Workflow



Neural networks were originally inspired by the brain, but the details of how they work are almost completely unrelated to how biological brains work.

MathWorks, Coursera

■ 머신러닝 기술을 적용해서 대용량의 데이터를 분석하면 겉으로는 보이지 않던 패턴을 발견할 수 있는데 이를 데이터마이닝이라고 한다. 데이터마이닝은 아래 4개의 문제를 다룬다.

예측 (Prediction)

주어진 Data에 근거하여 Model을 만들고 이 Model을 이용하여 새로운 Case 들에 대한 예측을 하는 작업

Ex 조업에 투여되는 성분, 제조 환경 (압력, 온도, 습도)에 의해 품질 spec을 예측

분류 (Classification)

일련의 범주들이 사전에 분류되어 있을 때, 특정한 case가 어디에 속하는가를 결정

Ex 품질등급이 양호/보통/불량일 때 새로운 제품의 품질등급 결정

군집 (Clustering)

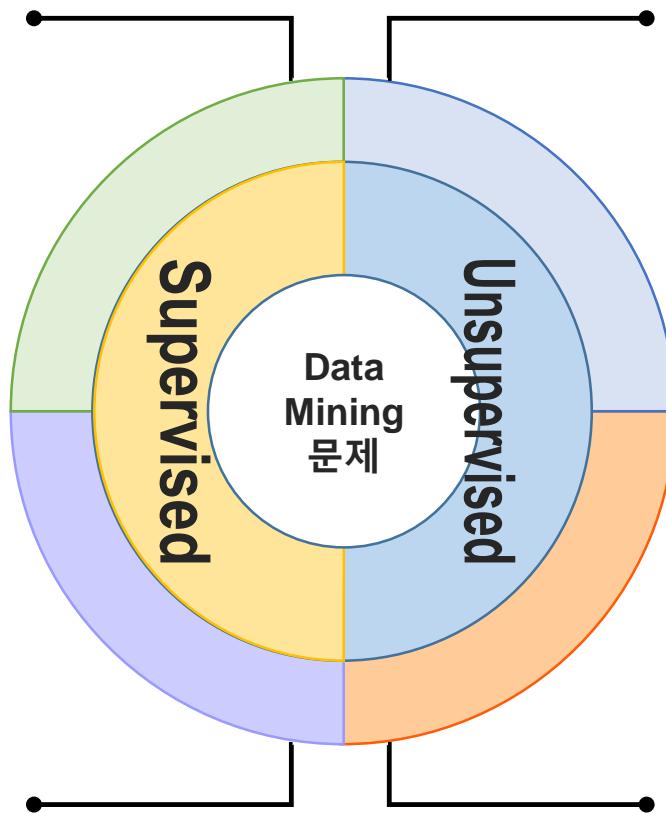
데이터의 여러 속성을 비교하여 유사한 특성을 갖는 항목들을 함께 군집 시키는 것

Ex 여러 공정 중 유사 특성을 갖는 공정간의 군집화

연관 규칙 (Association Rule)

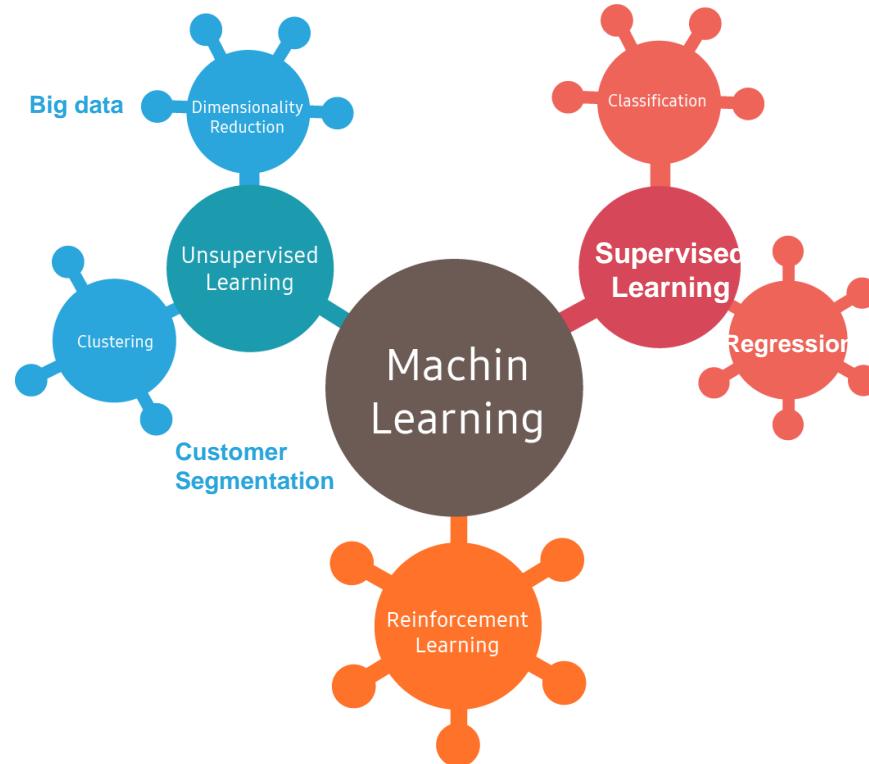
한 패턴의 출현이 다른 패턴 출현을 암시하는 속성이나 항목 간의 관계를 파악

Ex 한 공정에서 이상 패턴이 전체 공정에서 나타날 현상을 예측



I 기계학습 기반 데이터 분석 기법의 유형

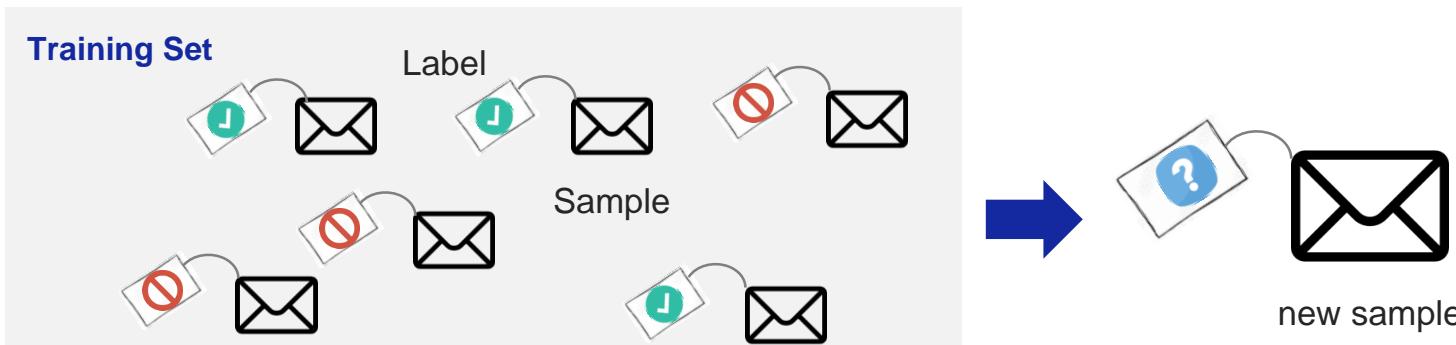
- ▶ 기계학습 기반의 데이터 분석 기법을 분류하는 기준은 다양한 관점에서 여러 가지 기준이 있을 수 있으나, 일반적으로 목적변수(혹은 반응변수, 목표변수, 출력 목표값 등으로 표현) 존재 여부 등에 따라, 지도 학습(Supervised Learning)과 자율학습 혹은 비지도 학습 (Unsupervised Learning)으로 분류하며, 추가적으로 강화학습 (Reinforcement Learning), 준 지도 학습(Semi-Supervised Learning) 등으로 구분하기도 한다.



| 지도 학습 (Supervised Learning)

- ▶ 지도 학습은 설명변수(혹은 독립변수, 특성(Feature) 등으로 표현)와 목적 변수(혹은 반응변수, 종속변수, 목표 변수, 출력값 등으로 표현) 간의 관계성을 표현해 내거나 미래 관측을 예측해 내는 것에 초점이 있으며 주로 인식, 분류, 진단, 예측 등의 문제 해결에 적합하다.
- ▶ 지도학습에 속하는 대표적인 기법은 설명 혹은 예측하고자 하는 목적 변수(혹은 반응변수, 종속변수)의 형태가 수치형(양적변수)인가 범주형(질적변수)인가에 따라 분류와 수치예측 방법으로 다시 나눌 수 있다.

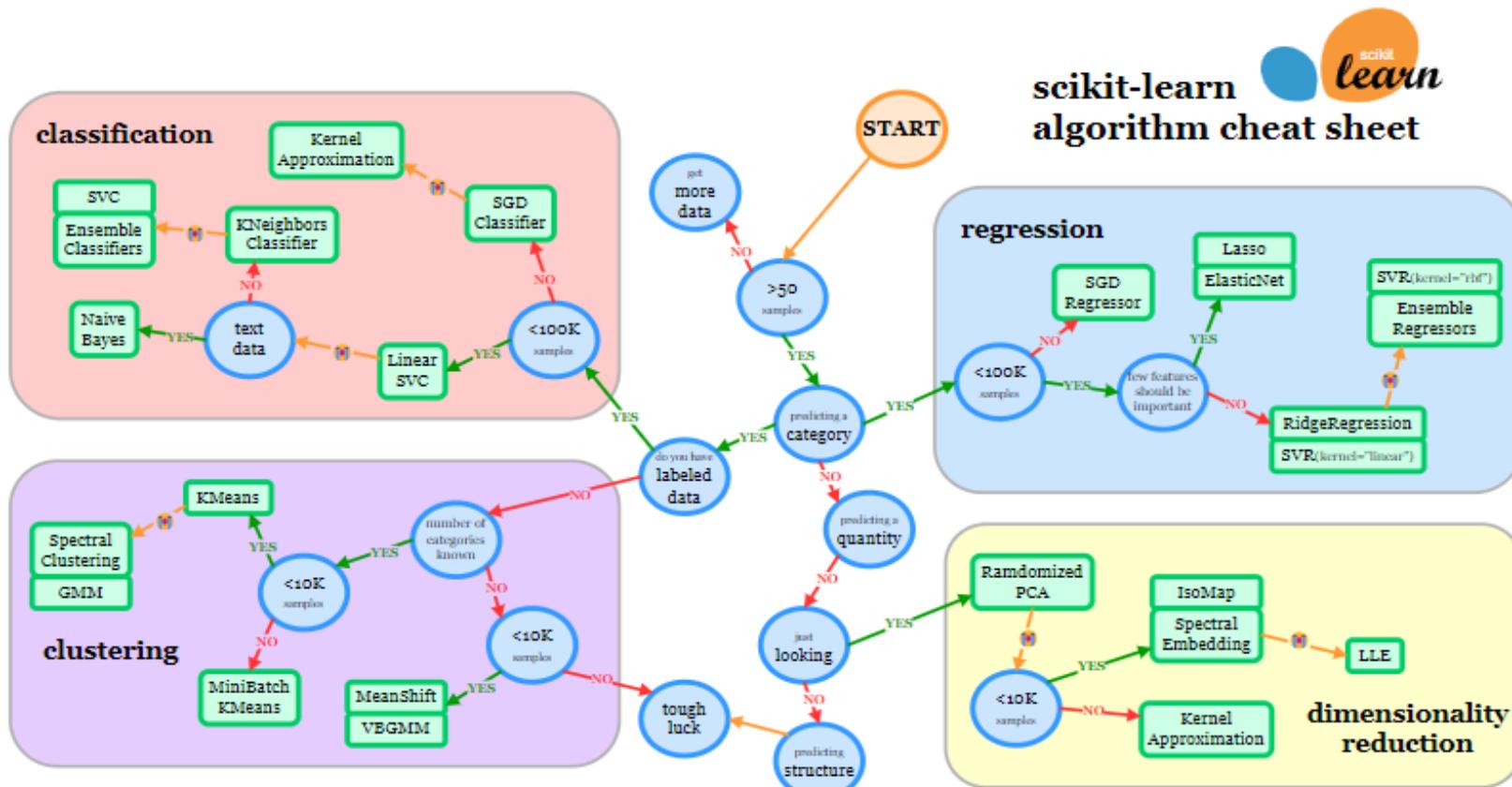
스팸 분류를 위한 레이블된 훈련 세트(지도 학습의 예)



- ▶ 알고리즘에 주입하는 훈련 데이터에 레이블 (label) 이라는 원하는 답이 포함된다.

지도 학습 (Supervised Learning)

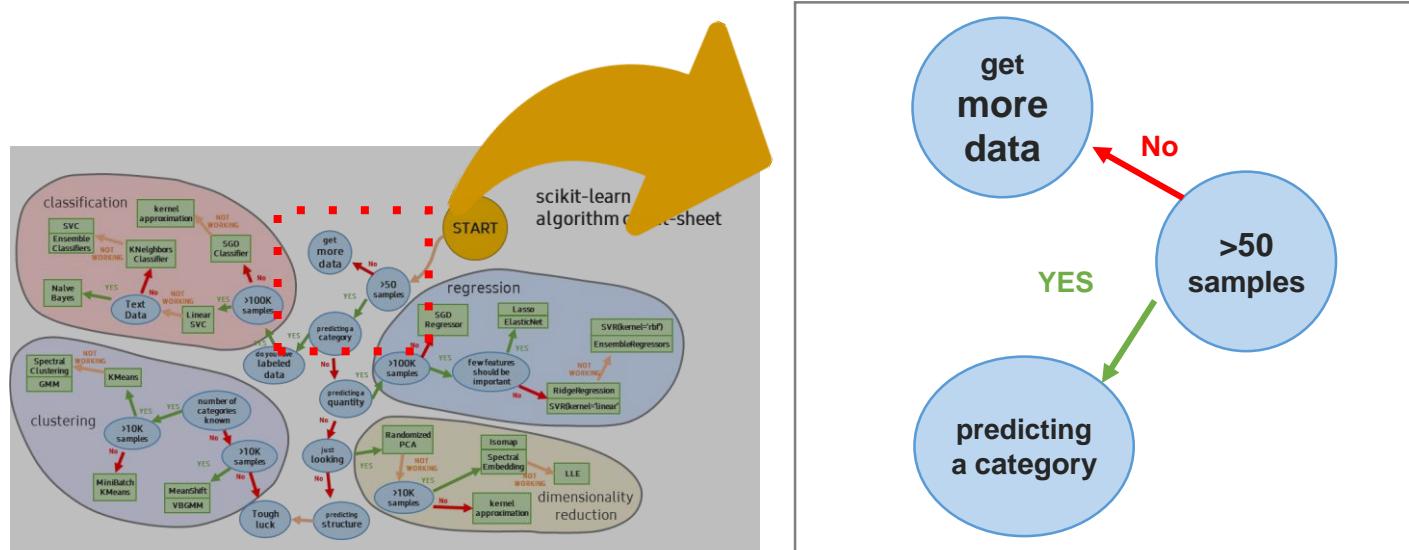
- ▶ 데이터마이닝은 통계학과 머신러닝 (또는 인공지능)의 결합이다. 머신러닝으로 해결해야 할 문제는 데이터마이닝의 문제와 큰 틀을 같이 한다.
 - ▶ 아래 그림은 파이썬 머신러닝 라이브러리인 사이킷런으로 문제를 해결해 나가는 가이드를 보여주고 있다.



| 지도 학습 (Supervised Learning)

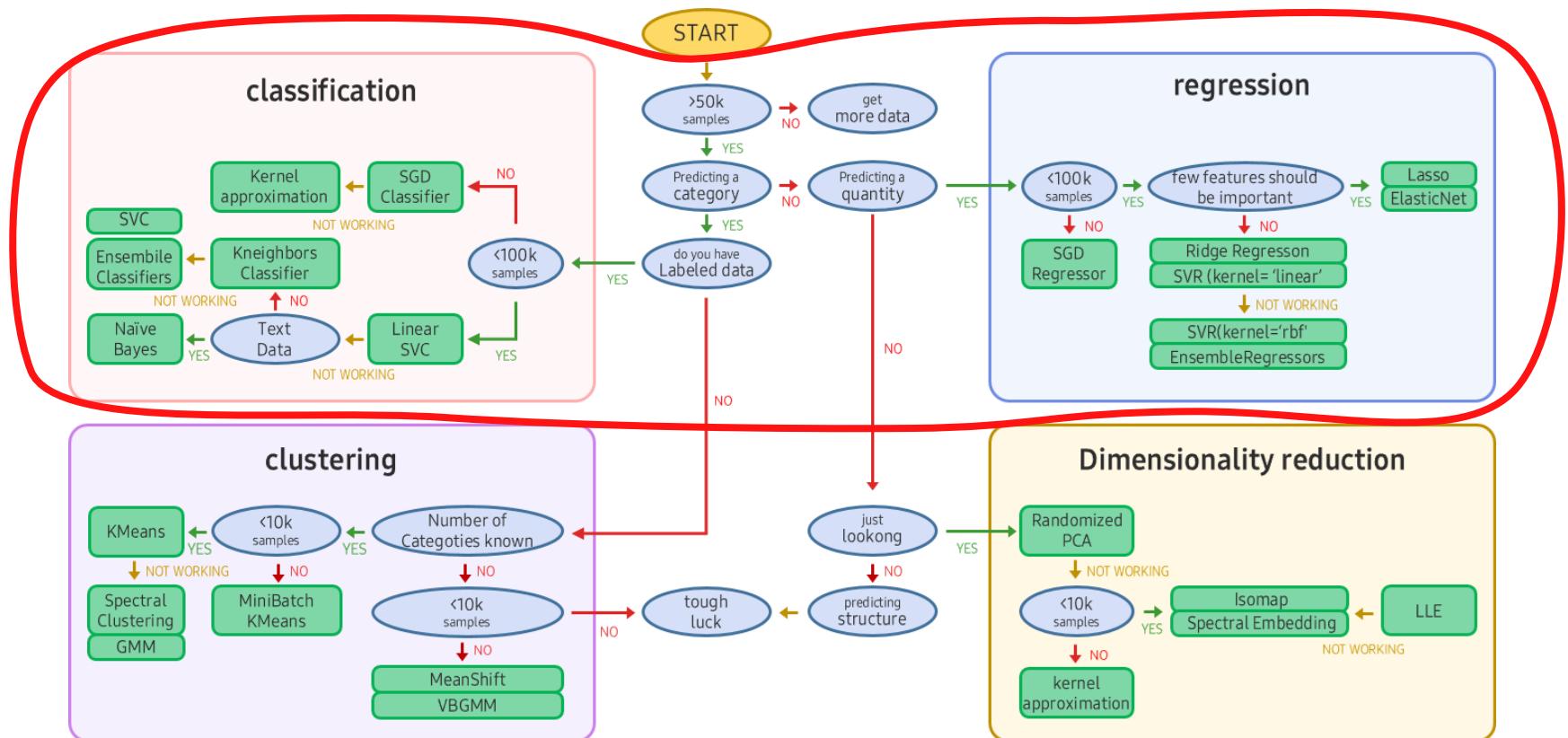
▶ 데이터가 적은 경우

- 아래의 그림을 보면 데이터가 적을 때는 통계학 기반의 분석을 할 수밖에 없다.
- 통계학에서는 모집단의 특징 (모수)을 추정하기 위해서 표본의 개수를 약 30개를 최소 수준으로 이해한다.
- 표본의 개수가 커질 때 모수와 가까워진다.
- 최소 수준인 약 30개의 데이터를 기반으로 어떤 가정을 할 수 있다. 여기서의 가정은 대수의 법칙을 통한 샘플이 30개 이상일 경우 정규분포를 따른다는 가정을 한다는 것이다.
- 데이터마이닝 또는 머신러닝에서는 대용량의 데이터가 반드시 필요한데, 기준은 대략 10만 건 이상으로 본다.



| 지도 학습 (Supervised Learning)

- 아래의 그림에서 보면 예측하는 문제 중 정답이 있는 경우 그 정답인 수치형인 경우에는 회귀, 그 정답이 범주형인 경우에는 분류인 것을 확인할 수 있다.



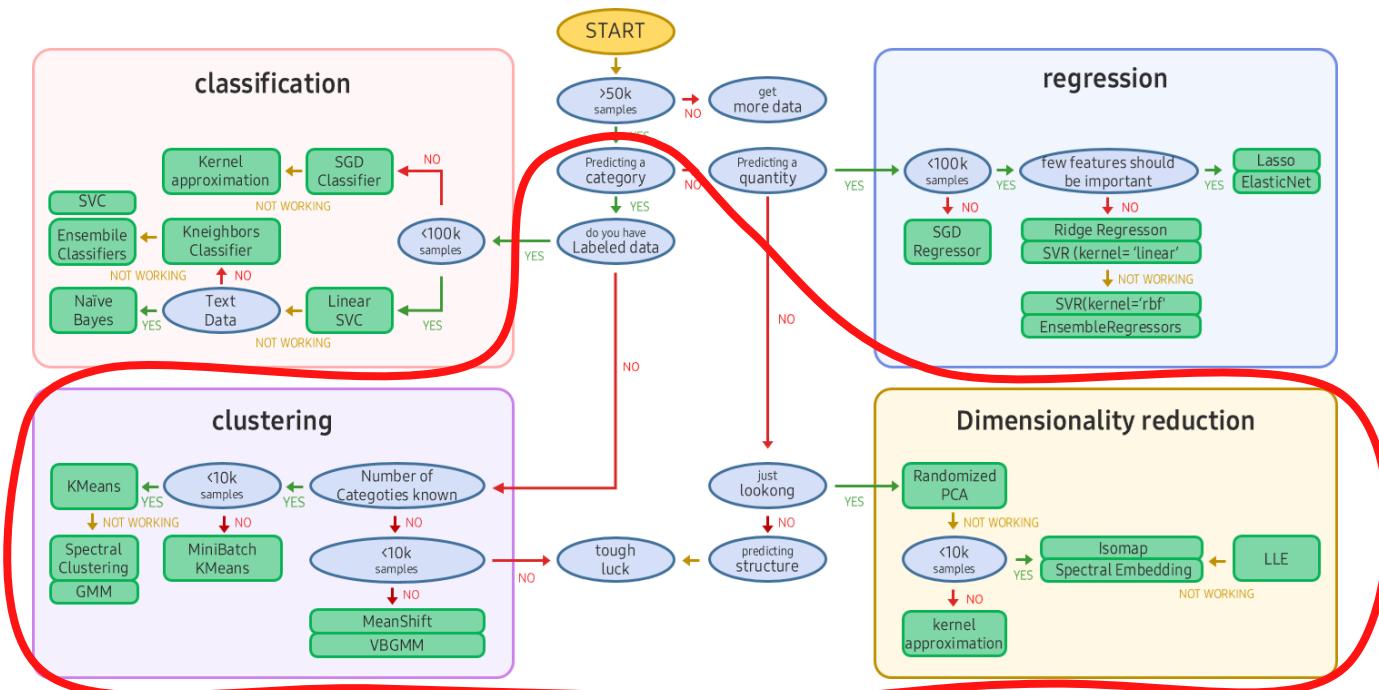
| 지도 학습 (Supervised Learning)

- ▶ 주요 기법 및 알고리즘

분류	수치예측 (혹은 회귀)
K- 최근접 이웃(K-Nearest Neighbors)	선형 회귀 (Linear Regression)
로지스틱 회귀 (Logistic Regression)	확장된 회귀분석 (ex: 다항회귀, 비선형 회귀, 벌점화 회귀 등)
인공 신경망 분석 (Artificial Neural Network)	인공 신경망 분석 (Artificial Neural Network)
의사결정트리 (Decision Tree)	의사결정트리 (Decision Tree)
서포트 벡터 머신 (Support Vector Machine)	서포트 벡터 머신 (회귀) (Support Vector Machine (Regression))
나이브 베이즈 (Naïve Bayes)	PLS (Partial Least Squares)
양상블 기법 (랜덤 포레스트 등)	양상블 기법 (랜덤 포레스트 등)

자율학습 혹은 비지도 학습 (Unsupervised Learning)

- ▶ 자율학습 혹은 비지도 학습은 목적변수(혹은 반응변수, 종속변수, 목표변수, 출력값)에 대한 정보 없이 학습이 이루어지는 형태를 말하며, 예측의 문제보다는 주로 현상의 기술 (Description)이나 특징 도출, 패턴 도출 등의 문제에 많이 활용된다.
 - ▶ 일반적으로 명확하고 뚜렷한 예측 목적이 있는 지도학습 기법에 비해 자율학습 기법은 사전정보가 없는 상태에서 유용한 정보나 패턴을 탐색적으로 발견하고자 하는 데이터 마이닝의 성격이 더 강하다고 볼 수 있다



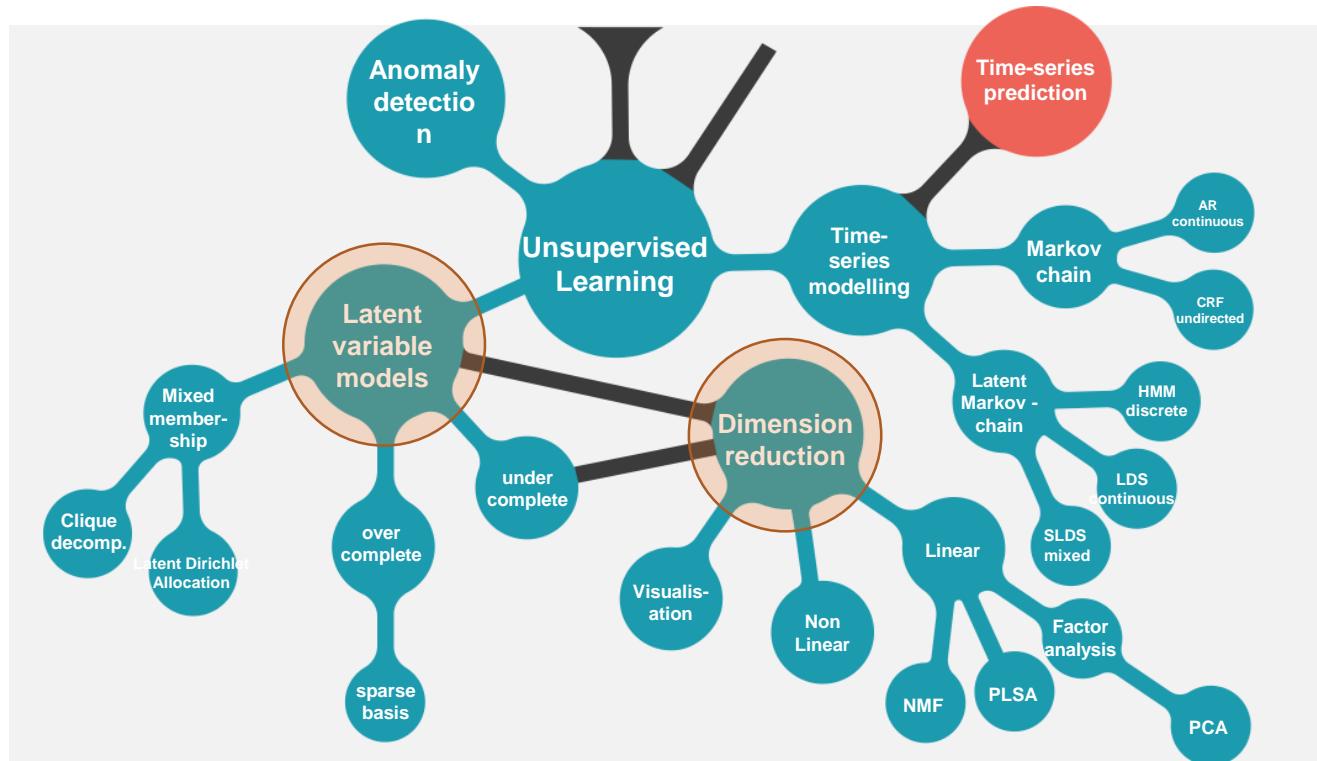
- 자율학습 혹은 비지도 학습 (Unsupervised Learning)

- ▶ 비지도 학습: 훈련 데이터에 레이블이 없어서, 시스템이 아무런 도움 없이 학습해야 한다.
- ▶ 중요한 비지도학습 알고리즘들

군집	시각화와 차원 축소	연관 규칙 학습
<ul style="list-style-type: none">• K-평균• DBSCAN• 계층 군집 분석• 이상치 탐지와 특이치 탐지• 원-클래스 SVM• 아이솔레이션 포레스트	<ul style="list-style-type: none">• 주성분 분석 (PCA)• 커널 PCA• 지역적 선형 임베딩• t-SNE	<ul style="list-style-type: none">• 어프라이어리 (Apriori)• 이클렛 (Eclat)

| 자율학습 혹은 비지도 학습 (Unsupervised Learning)

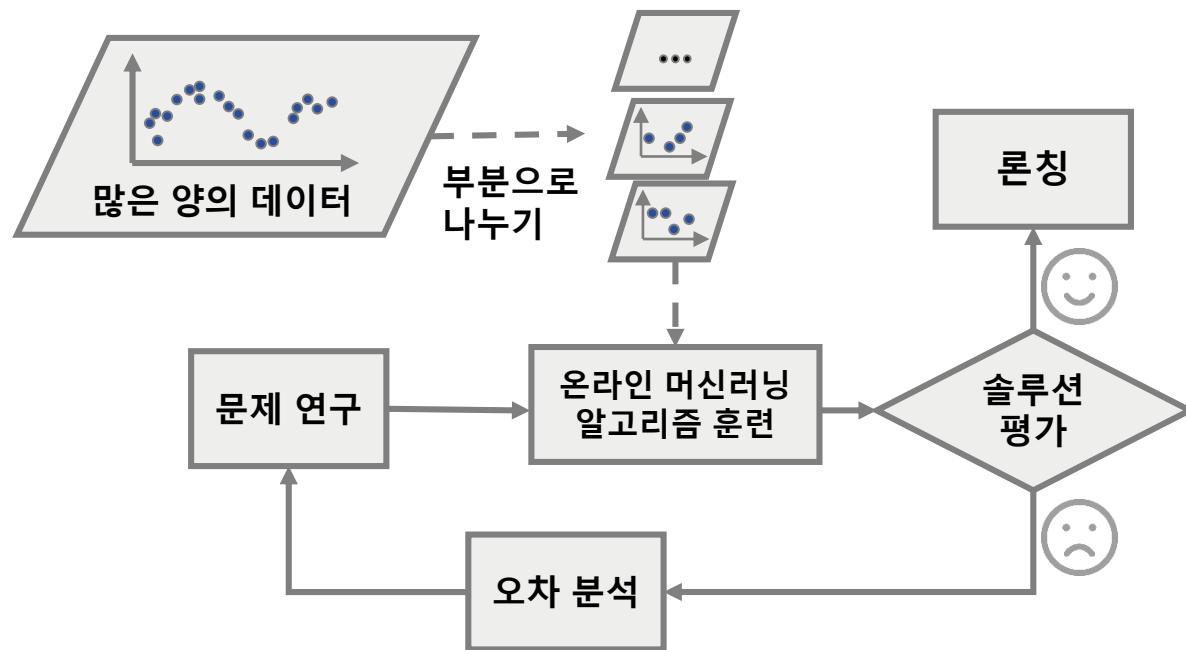
- ▶ 자율학습 혹은 비지도 학습에 속하는 대표적인 기법은 군집화(Clustering), 차원축소 기법, 연관관계분석(장바구니 분석), 자율학습 인공 신경망(SOM 등)의 기법이 있으며, 최근 관심이 높아지고 있는 딥러닝(Deep Learning) 기법에서도 입력 특성들의 차원을 축소하는 단계에서 자율학습 기법이 적용된다고 볼 수 있는 등 그 활용성은 매우 다양하다고 하겠다.



| 배치 학습과 온라인 학습

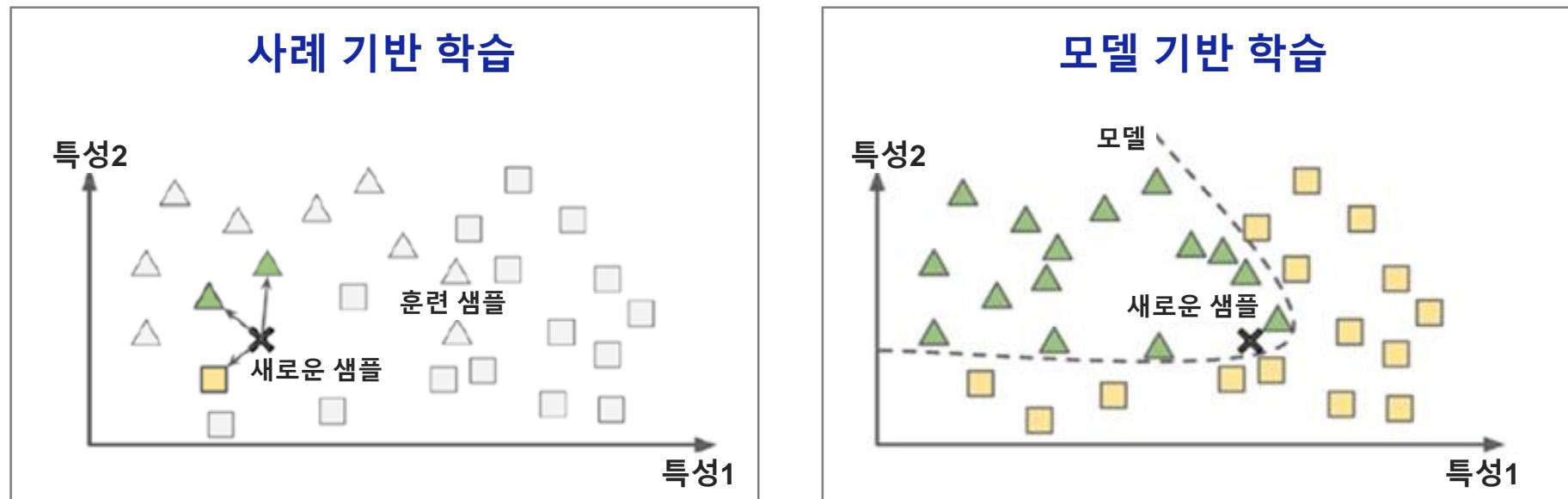
- ▶ **배치 학습**: 시스템이 점진적으로 학습할 수 없다.
- ▶ **온라인 학습**: 데이터를 순차적으로 한 개씩 또는 미니 배치 (mini-batch)라 부르는 작은 묶음 단위로 주입하여 시스템을 훈련한다.

온라인 학습을 사용한 대량의 데이터 처리



| 사례 기반 학습과 모델 기반 학습

- ▶ **사례 기반 학습**: 시스템이 훈련 샘플을 기억함으로써 학습한다. 유사도 측정을 사용해 새로운 데이터와 학습한 샘플을 비교하는 식으로 일반화한다.
- ▶ **모델 기반 학습**: 샘플로부터 모델을 만들어 예측에 사용한다.



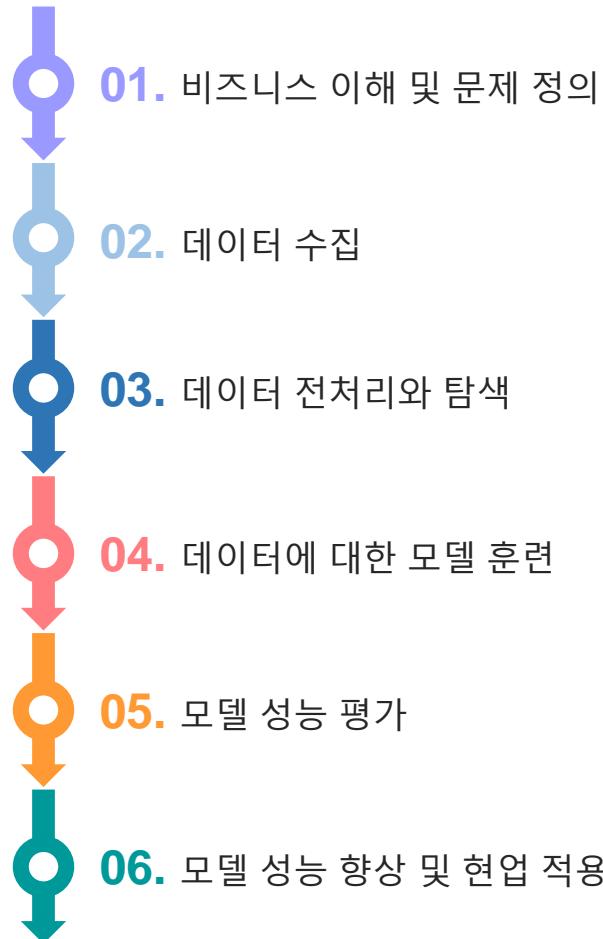
Unit 1.

The Concept of Artificial Intelligence

- | 1.1 인공지능의 정의
- | 1.2 인공지능의 종류 및 하위 집합
- | 1.3 머신러닝의 정의
- | 1.4 기계학습 관련 학문

- | 1.5 머신러닝 기반 데이터 분석 기법
 의
 유형 및 선택
- | 1.6 머신러닝 기반 데이터 분석 절차
- | 1.7 머신러닝을 사용하는 이유
- | 1.8 머신러닝의 한계

| 머신러닝 기반 데이터 분석 수행 시 일반적으로 다음의 절차를 따르게 된다





01. 비즈니스 이해 및 문제 정의

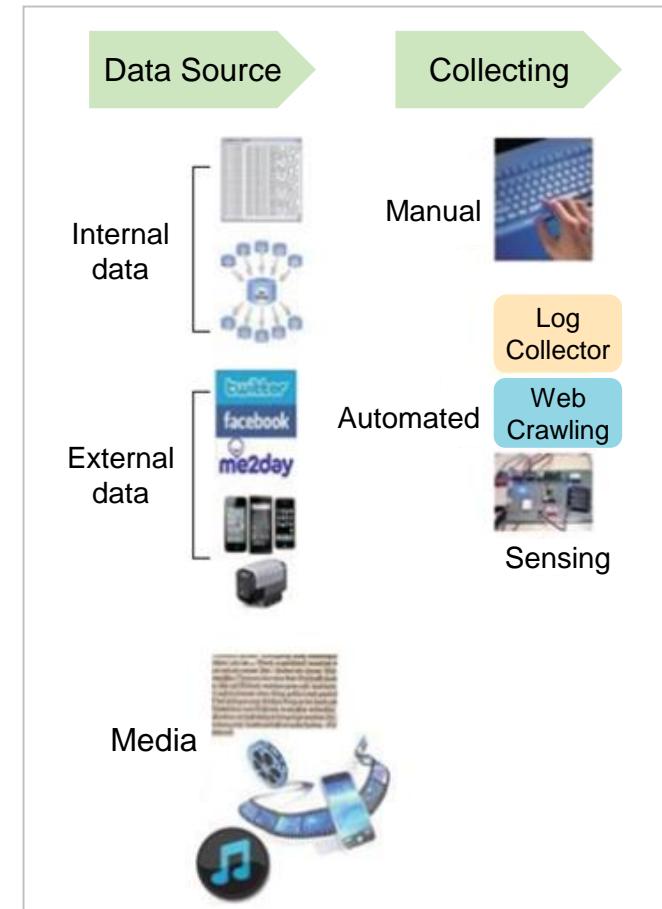
- ▶ 모든 분석 주제가 그렇듯, 머신러닝 기반 데이터 분석에서도 마찬가지로 해결하려는 문제를 정의하는 것이 가장 우선이다.
- ▶ 자신이 해결하려는 문제를 이해하고, 그 문제를 해결하기 위한 비즈니스 도메인 이해 및 문제를 파악해가는 과정을 반복해 가면서 문제를 재정의하고 해결책을 모색하는 단계를 반복적으로 거치면서 구체적인 목적과 필요한 데이터에 대한 이해가 명확해지게 된다.
- ▶ 또한, 문제 정의 과정과 필요한 데이터 형태에 대해 구상을 하는 과정에서 자연스럽게 어떤 머신러닝 기법을 적용하게 될지 일차적인 잠정적 의사결정 과정을 거치게 된다.





02. 데이터 수집

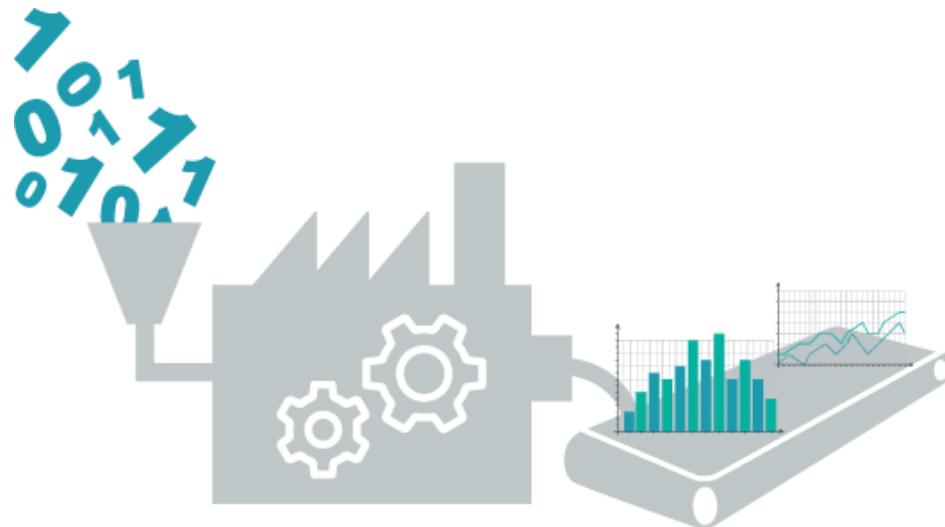
- ▶ 해결하고자 하는 분석 이슈가 명확해졌다면, 분석을 위해 필요한 데이터를 수집하게 된다.
물론 분석하려는 데이터가 이미 주어져 있는 상태에서 해당 데이터에 적합한 분석 이슈를 고민하게 되는 경우도 있지만, 그 경우에도 분석을 위해 추가로 필요한 데이터가 발생할 수 있으므로 필요한 데이터 수집은 올바른 분석을 위해 항상 고려해야 하는 요소이다.
- ▶ 내부 데이터 저장소(데이터웨어하우스 혹은 데이터 마트 등)에서 SQL을 통하여 데이터를 추출하거나
하둡 기반의 빅데이터 플랫폼에서 데이터를 추출하는 경우가 일반적이라고 볼 수 있으나, 어떤 경우에는
외부 데이터가 필요할 수도 있으므로 웹 사이트에서 필요한 데이터를 스크래핑 형태로 수집하거나 API 등을 통해 데이터를 수집해야 될 경우가 발생할 수도 있다.

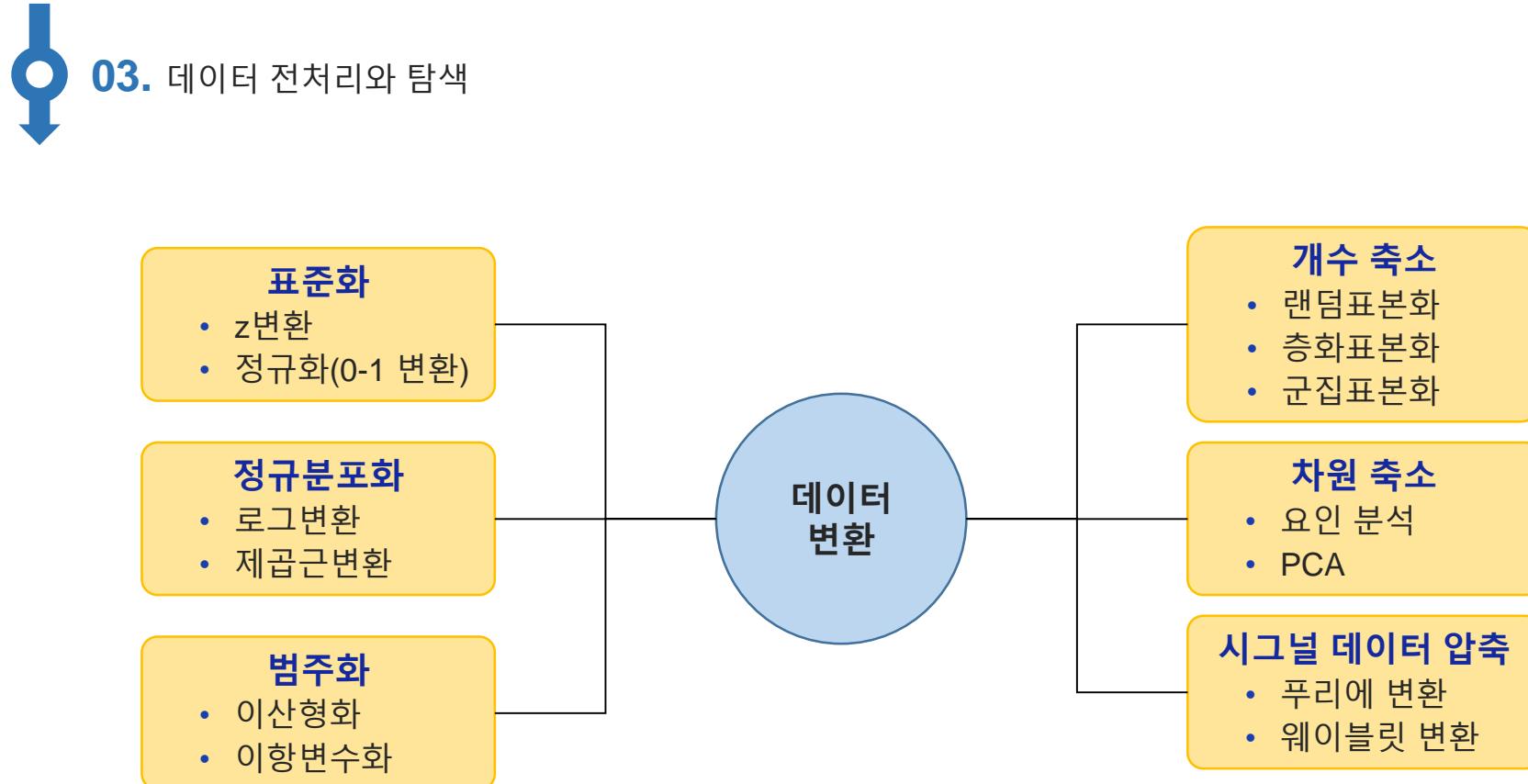




03. 데이터 전처리와 탐색

- ▶ 필요한 데이터를 수집되었다면, 머신러닝을 적용하기에 적당한 형태로 데이터를 전처리하고 변환하게 된다. 머신러닝 기반 데이터 분석 결과의 질은 필요한 기법이나 알고리즘에 도 좌우되지만, 데이터의 질에 따라 크게 좌우되므로 데이터 전처리와 변환 및 탐색 단계가 매우 중요하다.
- ▶ 실제로 대부분의 분석과 마찬가지로 머신러닝 기반 데이터 분석 프로세스 중에서도 이 단계가 가장 많은 시간과 노력을 들이게 되는 단계라고 할 수 있다.







03. 데이터 전처리와 탐색

▶ 정규분포화

- 로그 변환: 입력되는 데이터가 역함수 분포를 나타내는 경우 log를 사용해서 정규분포로 변환
- 제곱근 변환: 정규분포가 아닌 데이터를 제곱근을 사용해서 정규분포 변환

▶ 범주화

- 이산형 (discretization): 다수의 구간으로 연속형 변수를 범주화
- 이항변수화 (binarization): 0과 1의 두 개의 값으로 더미 변수화

▶ 개수 축소

- 단순 임의 추출
- 체계적 추출
- 층화 임의 추출
- 군집 추출
- 다단계 추출

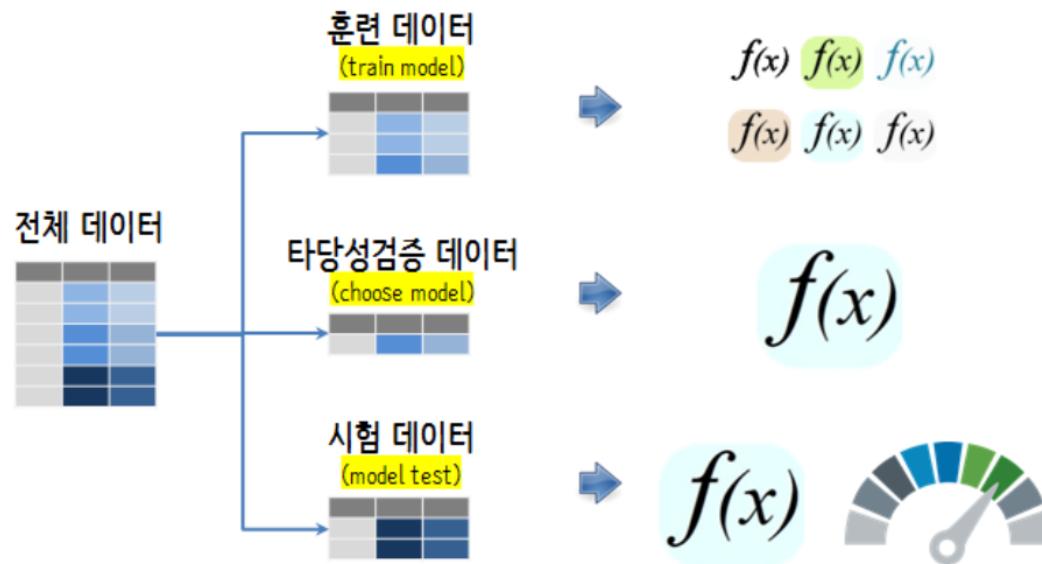
▶ 차원축소

- 요인분석: 잠재적인 변수를 찾는 방법으로 공통적인 요인을 찾아낸다.
- 주성분 분석



04. 데이터에 대한 모델 훈련

- ▶ 데이터 전처리와 탐색을 거친 후 사용하고자 하는 머신러닝 기법을 적용하여 데이터를 학습하는 단계이다.
- ▶ 물론 지도 학습의 경우는 모델 훈련을 위해 데이터를 학습용 데이터와 검증용 및 평가용 데이터로 분할하거나, 교차 검증 등에 대한 설계를 거친 후 모델 훈련을 하게 된다.
- ▶ 자율학습의 경우는 목적 값을 가지지 않기 때문에 모델 훈련을 한다기보다는 바로 분석을 통한 패턴 도출의 과정을 수행한다고 볼 수 있다.





05. 모델 성능 평가

- 일반적으로 머신러닝 기반 학습 모델은 모델 훈련에 사용한 훈련 데이터에 편향된 결과값을 내는 경향이 많으므로, 머신러닝 알고리즘이 데이터로부터 얼마나 잘 학습했는지를 평가하기 위해 평가 데이터 세트를 이용하여 모델의 정확도를 평가하게 된다.
- 자율학습의 경우에는 평가 데이터 세트를 두지 않는 경우가 일반적이므로, 교차 검증보다는 분석 과정에서 도출되는 통계치나 규칙들의 해석 가능성 등에 초점을 두고 성능을 평가하게 된다.

		True/Actual		
		Cat (🐱)	Fish (🐠)	Hen (🐓)
Predicted	Cat (🐱)	4	6	3
	Fish (🐠)	1	2	0
	Hen (🐓)	1	2	6

출처 : <https://towardsdatascience.com/multi-class-metrics-made-simple-part-i-precision-and-recall-9250280bddc2>



06. 모델 성능 향상 및 협업 적용

- ▶ 일반적으로 단일 머신러닝 분석 프로세스로 해결하고자 하는 이슈가 단번에 해결되는 경우는 거의 없으며, 지속적으로 모델 파라미터나 추정 방법 등을 변화시켜서 모델의 성능 향상을 꾀하게 된다.
- ▶ 때로는 다른 알고리즘을 적용하여 원래 적용한 알고리즘과 성능을 비교하게 된다. 모델 성능이 어느 정도여야 만족스러운 성능인가에 관한 판단은 분석하고자 하는 이슈나 비즈니스 도메인에 따라 달라질 수 있으며, 어떤 경우든 모델 성능이 충분히 향상되었다는 판단에 대한 절대적인 기준이 존재하는 것은 아니다.
- ▶ 그렇지만, 분석가는 여러 가지 알고리즘을 비교 적용해 보고, 동일 알고리즘 내에서도 파라미터나 추정 방법 등을 바꿔가며 만족스러운 수준이라고 판단될 때까지 모델 성능 향상을 모색해 볼 필요가 있다.
- ▶ 이런 과정이 끝난 후 모델이 만족할 만큼 결과값이 도출된다면, 당초 의도했던 비즈니스 이슈에 적용할 수 있다. 경우에 따라서는 자동화나 시스템 연계를 위한 추가적인 개발 작업이 필요하게 될 수도 있다.



Unit 1.

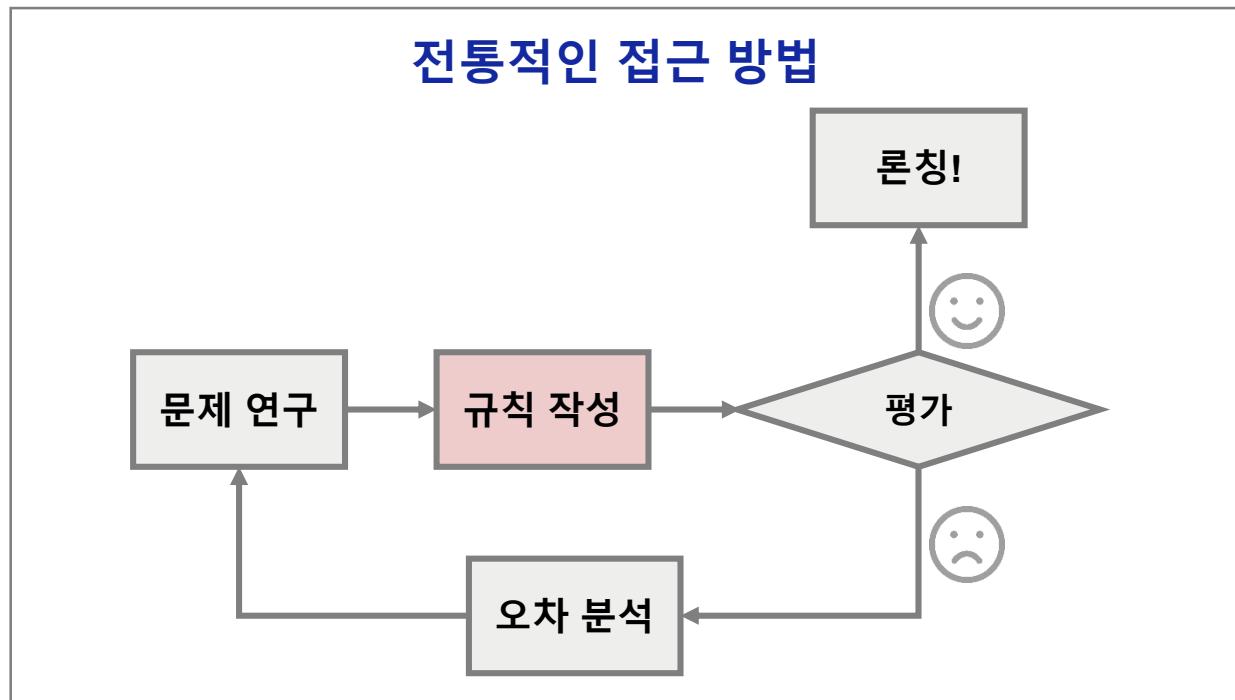
The Concept of Artificial Intelligence

- | 1.1 인공지능의 정의
- | 1.2 인공지능의 종류 및 하위 집합
- | 1.3 머신러닝의 정의
- | 1.4 기계학습 관련 학문

- | 1.5 머신러닝 기반 데이터 분석 기법
 의
 유형 및 선택
- | 1.6 머신러닝 기반 데이터 분석 절차
- | 1.7 머신러닝을 사용하는 이유
- | 1.8 머신러닝의 한계

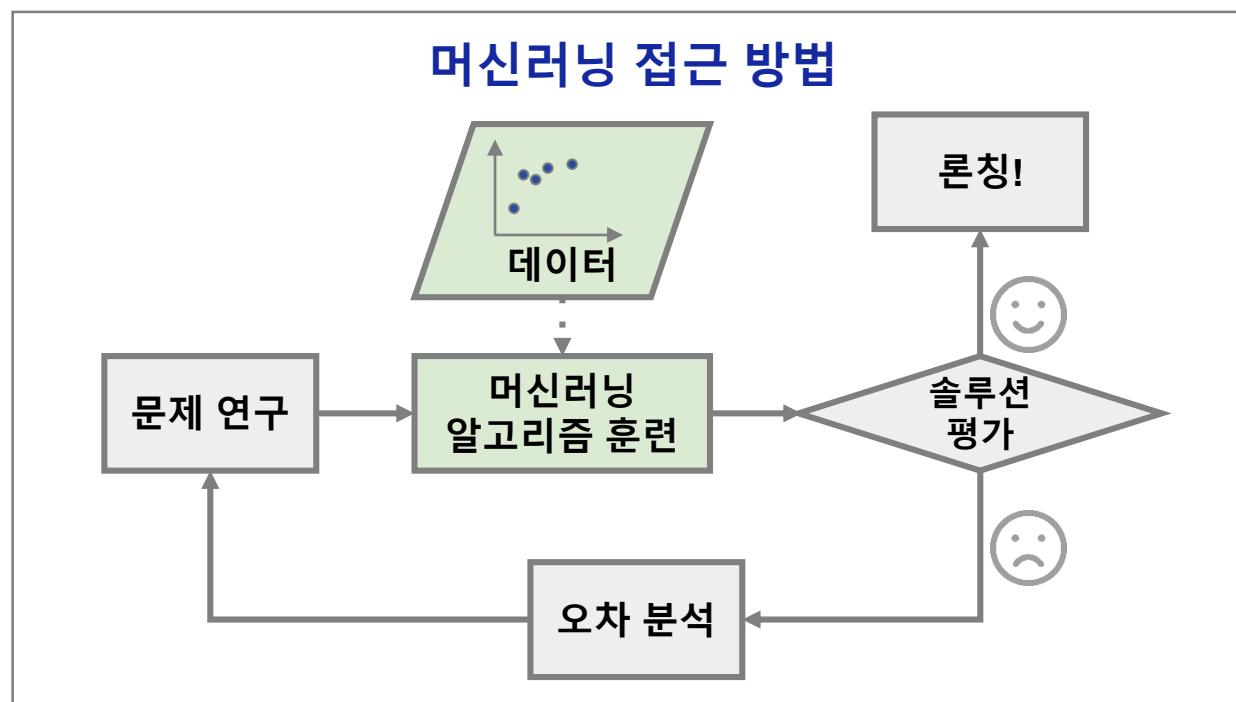
왜 머신러닝을 사용하는가?

- ▶ 전통적 프로그래밍 기법으로는 규칙이 점점 길고 복잡해지므로 유지 보수하기 매우 힘들다.
하지만 머신러닝 기법에 기반을 둔 스팸 필터는 일반 메일에 비해 스팸에 자주 나타나는 패턴을 감지하여 어떤 단어와 구절이 스팸 메일을 판단하는데 좋은 기준인지 자동으로 학습한다.
- ▶ 머신러닝은 전통적인 방식으로는 너무 복잡하거나 알려진 알고리즘이 없는 분야 (예: 음성인식)에 적합하다.



머신러닝의 장점

- ▶ 기존 솔루션으로는 많은 수동 조정과 규칙이 필요한 문제: 하나의 머신러닝 모델이 코드를 간단하게 만들고 전통적인 방법보다 더 잘 수행되도록 할 수 있다.
- ▶ 전통적인 방식으로는 해결 방법이 없는 복잡한 문제: 가장 뛰어난 머신러닝 기법으로 해결 방법을 찾을 수 있다.
- ▶ 유동적인 환경: 머신러닝 시스템은 새로운 데이터에서도 적응이 가능하다.
- ▶ 복잡한 문제와 대량의 데이터에서 통찰 얻고자 하는 경우



Unit 1.

The Concept of Artificial Intelligence

| 1.1 인공지능의 정의

| 1.2 인공지능의 종류 및 하위 집합

| 1.3 머신러닝의 정의

| 1.4 기계학습 관련 학문

| 1.5 머신러닝 기반 데이터 분석 기법
의

유형 및 선택

| 1.6 머신러닝 기반 데이터 분석 절차

| 1.7 머신러닝을 사용하는 이유

| 1.8 머신러닝의 한계

Limitations of Machine Learning

- 충분하지 않은 양의 훈련 데이터
 - ▶ 대부분의 머신러닝 알고리즘이 잘 작동하려면 데이터 많아야 합니다.
 - ▶ 아주 간단한 문제에서조차도 수천 개의 데이터가 필요하고 이미지나 음성인식 같은 복잡한 문제라면 수백만 개가 필요할지도 모릅니다. (이미 만들어진 모델을 재사용 할 수 없을 경우).
 - ▶ 훈련 데이터를 추가로 모으는 것이 항상 쉽거나 저렴한 일은 아니므로, 아직은 알고리즘을 무시할 수 없다.

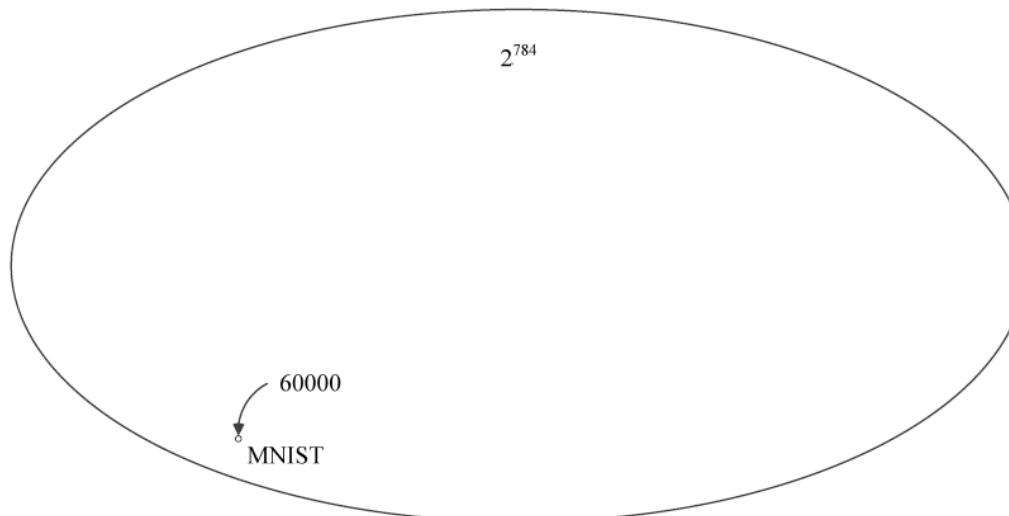


그림 1-9 방대한 특징 공간과 희소한 데이터베이스

- 대표성 없는 훈련 데이터

- ▶ 샘플링 잡음 (우연에 의한 대표성 없는 데이터)
- ▶ 샘플링 편향 (매우 큰 샘플도 표본 추출 방법이 잘못되면 대표성을 띠지 못함)

샘플링 잡음
(Sampling Noise)



샘플링 편향
(Sampling Bias)



- 낮은 품질의 데이터

- ▶ 훈련 데이터가 에러, 이상치, 잡음으로 가득하다면 머신러닝 시스템이 내재된 패턴을 찾기 어려워 잘 작동하지 않을 것입니다.
- ▶ 정제가 필요한 경우 : 일부 샘플이 이상치라는게 명확하면 무시하거나 고치는 것이 좋다.

- 관련 없는 특성

- ▶ 훈련데이터에 관련 있는 특성, 특성이 충분해야 시스템이 학습할 수 있다.
- ▶ 성공적인 머신러닝 프로젝트의 핵심 요소는 훈련에 사용할 좋은 특성을 찾는 것 (feature engineering)
- ▶ Feature selection: 가지고 있는 특성 중에서 훈련에 가장 유용한 특성 선택
- ▶ Feature extraction: 특성을 결합하여 더 유용한 특성을 만듭니다 (차원 축소).

- 훈련 데이터 과대적합

- ▶ 규제 (regularization) : 모델을 단순하게 하고 과대적합의 위험을 감소시키기 위해 모델에 제약을 가하는 것

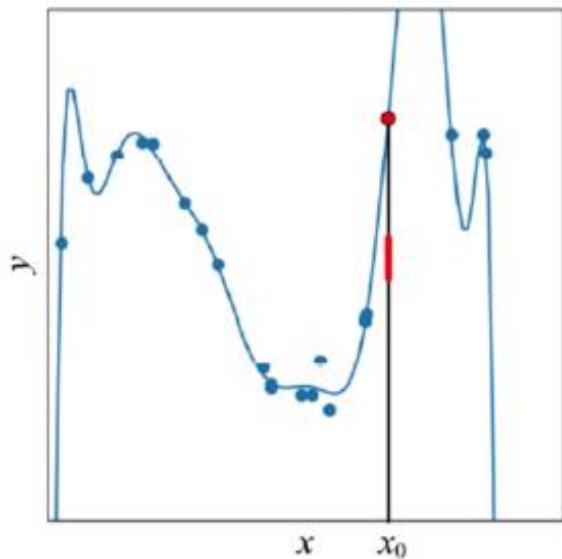
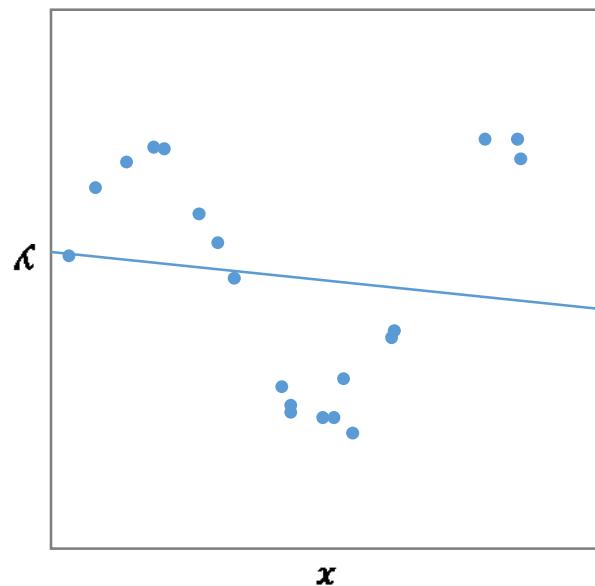


그림 1-14 과잉적합되었을 때 부정확한 예측 현상

- **훈련 데이터 과소적합**

- ▶ 모델 파라미터가 더 많은 강력한 모델을 선택
- ▶ 학습 알고리즘에서 더 좋은 특성 제공 (특성공학)
- ▶ 모델의 제약을 줄임 (규제 하이퍼파라미터 감소)



Unit 2.

Applications of Artificial Intelligence

- | 2.1 인공지능의 응용분야
- | 2.2 Image Recognition
- | 2.3 Computer Vision & Machine Vision 분야
- | 2.4 언어지능 분야

머신러닝 덕택에 이메일 스팸 필터, 편리한 텍스트와 음성 인식 소프트웨어, 믿을 수 있는 웹 검색 엔진, 안전하고 효율적인 자율 주행 자동차도 사용할 수 있을 것이다.

의료 애플리케이션에서도 큰 진전이 있었다.

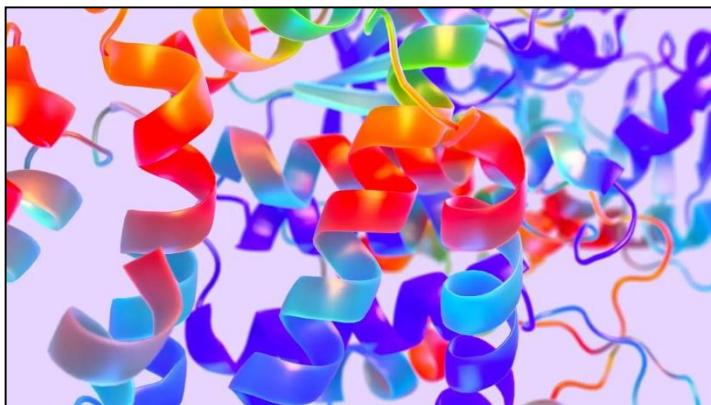
Ex 딥러닝 모델을 사용하여 피부 암을 거의 사람이 할 수 있는 정도 수준의 정확도로 진단할 수 있다는 것을 보여주었다.



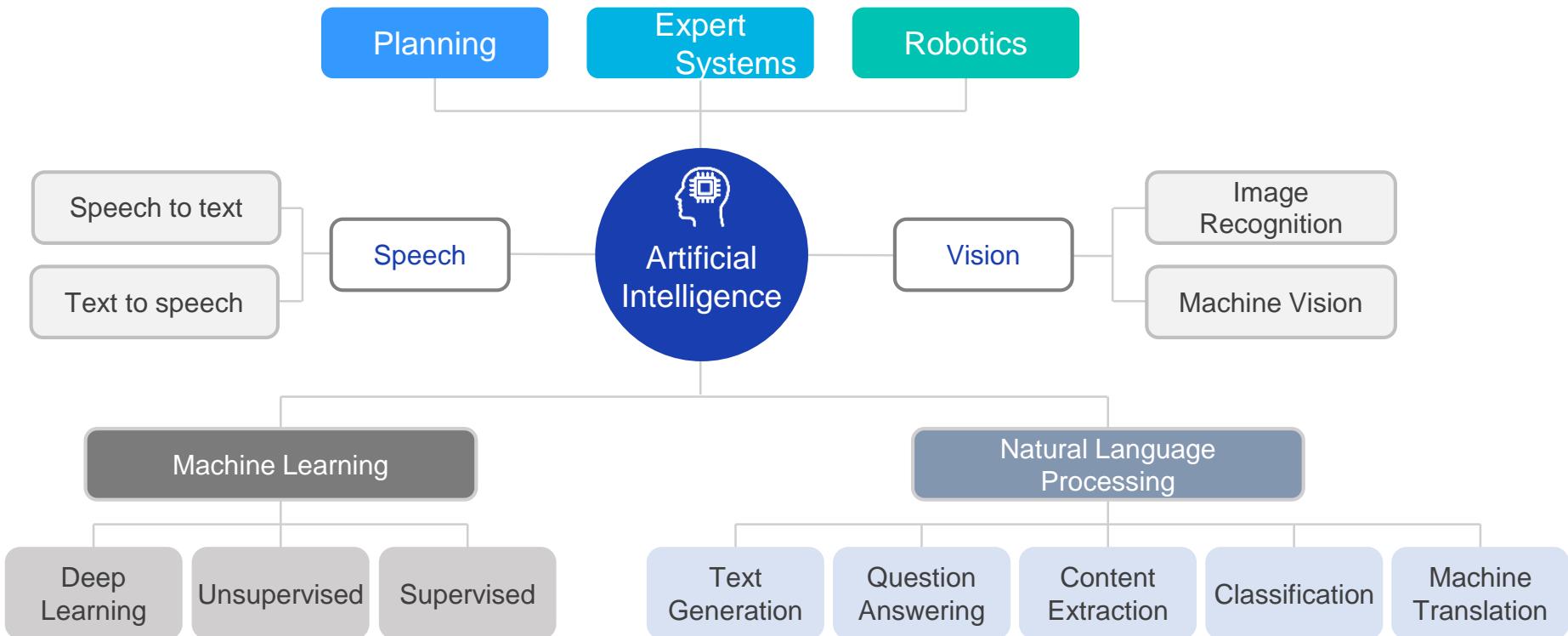
Andre Esteva, et al., “**Dermatologist-level classification of skin cancer with deep neural networks**”, *Nature*, volume 542(2017), pages 115–118

<https://www.nature.com/articles/nature21056>

딥러닝으로 3D 단백질 구조를 예측하여 처음으로 물리학 기반 방식의 성능을 뛰어넘었다.



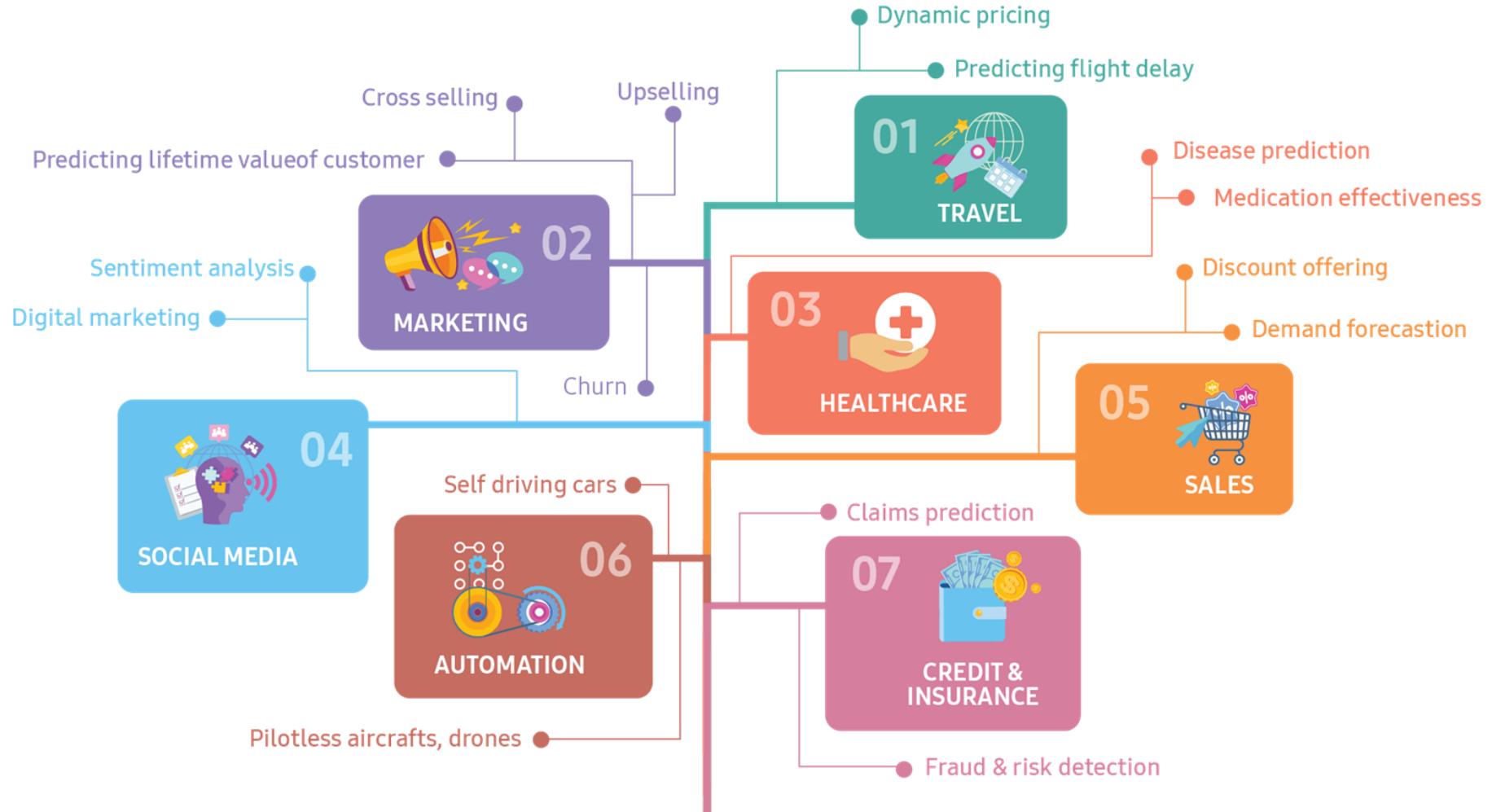
머신러닝은 점점 더 중요해지며, 우리 일상생활에서도 아주 큰 역할을 하고 있고 다음과 같은 응용분야로 구분된다.



Applications of Artificial Intelligence

- ▶ 이미지 분류 작업: 생산 라인에서 제품 이미지를 분석해 자동으로 분류
- ▶ 시맨틱 분할 작업: 뇌를 스캔하여 종양 진단
- ▶ 텍스트 분류 (자연어 처리): 자동으로 뉴스 기사 분류
- ▶ 텍스트 분류: 토론 포럼에서 부정적인 코멘트를 자동으로 구분
- ▶ 텍스트 요약: 긴 문서를 자동으로 요약
- ▶ 자연어 이해: 챗봇 (chatbot) 또는 개인 비서 만들기
- ▶ 회귀 분석: 회사의 내년도 수익을 예측하기
- ▶ 음성 인식: 음성 명령에 반응하는 앱
- ▶ 이상치 탐지: 신용 카드 부정 거래 감지
- ▶ 군집 작업: 구매 이력을 기반으로 고객을 나누고 각 집합마다 다른 마케팅 전략을 계획
- ▶ 데이터 시각화: 고차원의 복잡한 데이터셋을 명확하고 의미 있는 그래프로 표현하기
- ▶ 추천 시스템: 과거 구매 이력을 기반으로 고객이 관심을 가질 수 있는 상품 추천하기
- ▶ 강화 학습: 지능형 게임 봇 (bot) 만들기

Applications of Artificial Intelligence



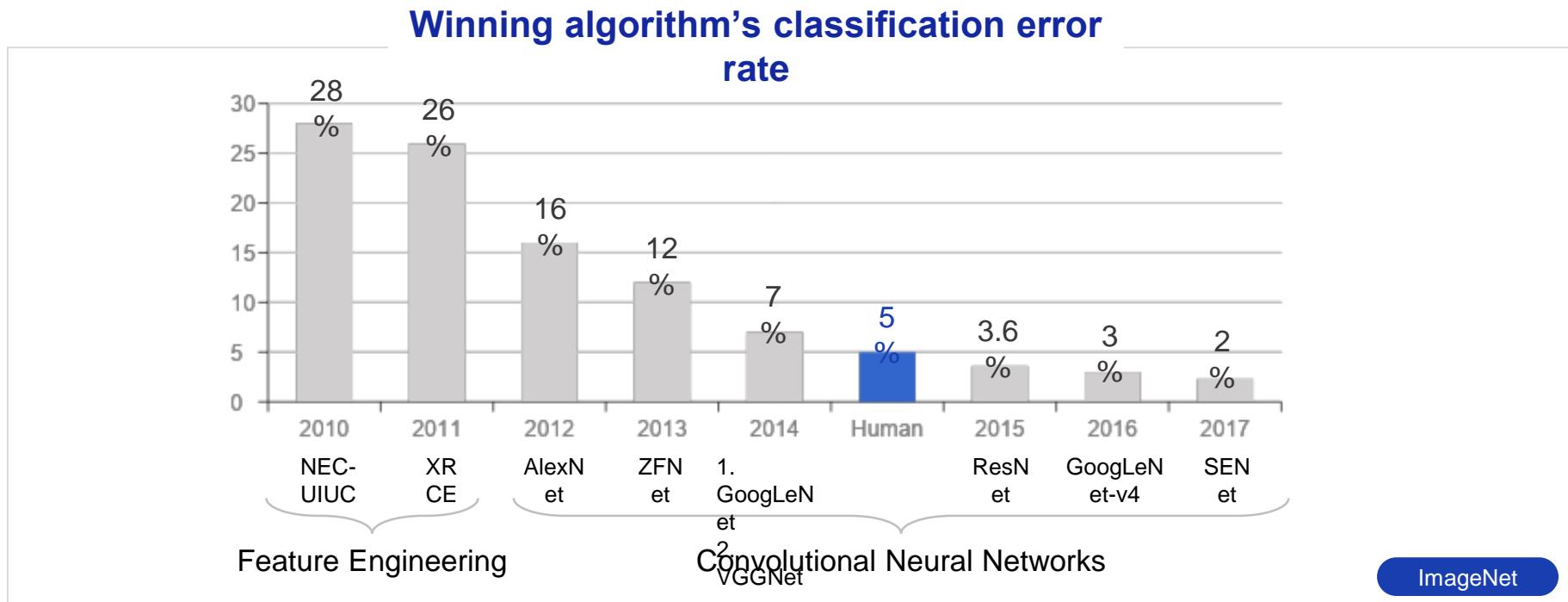
Unit 2.

Applications of Artificial Intelligence

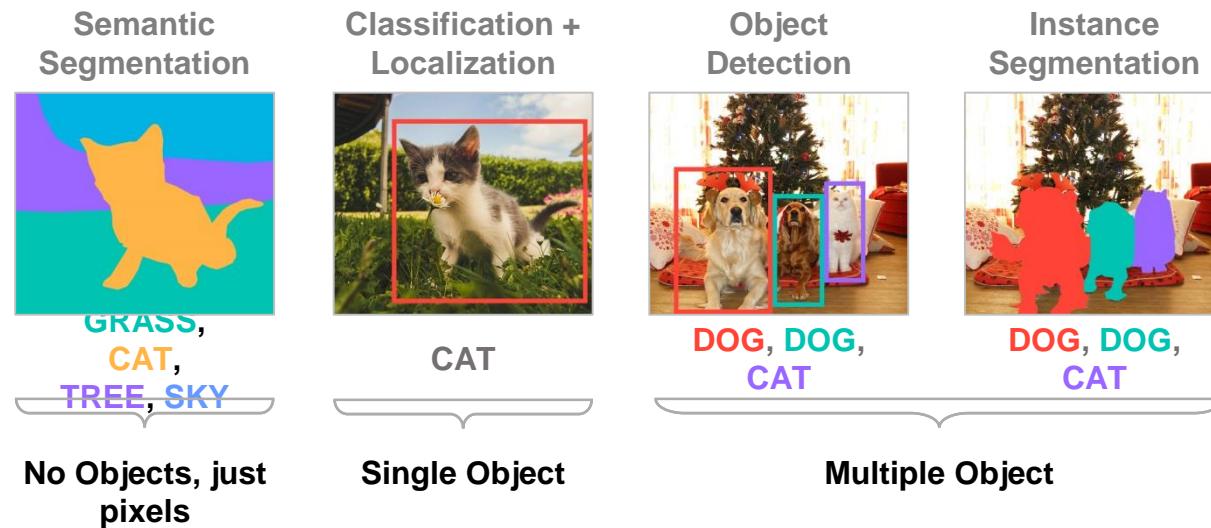
- | 2.1 인공지능의 응용분야
- | 2.2 Image Recognition
- | 2.3 Computer Vision & Machine Vision 분야
- | 2.4 언어지능 분야

Image Recognition

- The benchmark ImageNet classification error rate fell from 26 percent in 2011 to 3.1 percent in 2016, surpassing far beyond the human capability of 5% of error rate.
- It was not until the adoption of Convolutional Neural Networks that accurate of image recognition was dramatically improved.



- Image recognition refers to technologies that identify places, logos, people, objects, buildings, and several other variables in images.
- Apart from image recognition, computer vision also includes event detection, object recognition, learning, image reconstruction and video tracking.



https://mlwhiz.com/blog/2018/09/22/object_detection/

Unit 2.

Applications of Artificial Intelligence

- | 2.1 인공지능의 응용분야
- | 2.2 Image Recognition
- | 2.3 Computer Vision & Machine Vision 분야
- | 2.4 언어지능 분야

Computer Vision & Machine Vision

- Computer vision is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos.
- Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information.
- The applications of computer vision are various. They include agriculture, geoscience, biometrics, augmented reality, medical image analysis, robotics, industrial quality inspection, security and surveillance.



Scientific
Research



Natural
Resources
Management



Protection and
Security

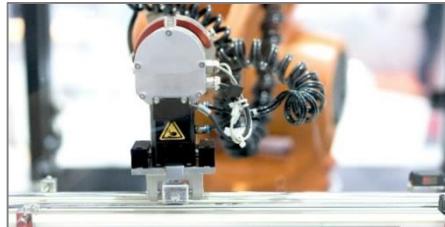


Marketing

[Wikipedia, Vidolab](#)

Computer Vision & Machine Vision

- Machine vision as a systems engineering discipline can be considered distinct from computer vision, a form of computer science. It attempts to integrate existing technologies in new ways and apply them to solve real world problems
- Machine vision refers to many technologies, software and hardware products, integrated systems, actions, methods and expertise
- Machine vision (MV) is the technology and methods used to provide imaging-based automatic inspection and analysis for such applications as automatic inspection, process control, and robot guidance, usually in industry



Guidanc
e



Gaugin
g



Defect Detection



Packaging
Inspection

Wikipedia, Vidolab, DevisionX

Acquire Automation – Machine Vision

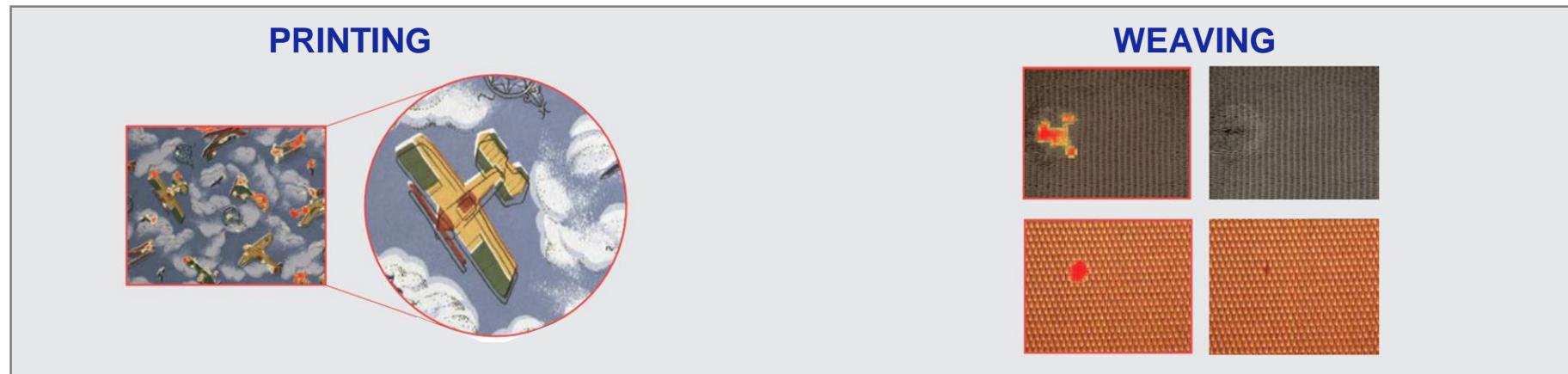
- ▶ Acquire Automation은 정부 또는 규제 기관에서 요구하는 수준에도 이상적으로 기업의 제품 포장 및 품질 검증할 수 있는 자동화된 광범위한 비전 솔루션을 갖고 있다.
- ▶ 그중 한 가지를 예로 들면 병에 대해 360도로 검사하여 제품을 올바르게 포장하였는지 확인하고, 모든 구성 요소가 제대로 포함되어 있는지 검사한다.
- ▶ 세부적으로 캡의 폐쇄 유무, 밀봉, 위치 및 기울기, 라벨 식별, 색상 확인, 바코드 확인 등의 기능을 수행한다.
- ▶ 또한 머신 비전 기술을 통해 데이터를 수집 및 제공하여 실시간 생산 통계를 제공하고 프로세스 제어를 촉진한다.



<https://acquireautomation.com/>

| 코그넥스 (Cognex) – VisionPro ViDi

- ▶ 코그넥스는 머신비전 시스템, 소프트웨어, 센서 제조업체이며 공장 자동화에 최적화된 딥러닝 기반 이미지 분석 소프트웨어 VisionPro ViDi를 출시하였다.
- ▶ VisionPro ViDi는 룰 기반 알고리즘으로 프로그래밍하기 어려운 결함 감지, 질감 및 재질 분류, 조립 검사, 문자 판독을 딥러닝으로 해결할 수 있다.
- ▶ 산업 이미지 분석이라는 특수 목적으로 활용하므로 수 백장 정도의 극소량 이미지트만 필요하고 트레이닝 및 검증 시간이 짧다. 컴퓨팅 비용 또한 저렴하다.
- ▶ 비전 분야의 비 전문가가 사용할 수 있도록 타기팅하고 실제 공장 조건에 맞춰 개발하였다.



Cognex's brochure

I Focal

- ▶ Focal은 AI와 컴퓨터 비전 기반 품절 감지 시스템을 서비스한다.
- ▶ 작고 저렴한 카메라를 매장 통로마다 설치하고 30분마다 한 번씩 사진을 찍는다.
- ▶ 이미지에서 품절된 상품을 인식하고 종업원에게 차트를 제공한다.
- ▶ 사람 종업원이 품절 상품을 인식하는 데는 하루에 약 4시간이 걸린다고 주장한다.
- ▶ 인건비가 증가함에 따라 무인 매장을 운영하는 소매점이 늘고 있으나, 인력이 줄어듦에 따라 재고를 확인하고 보충하는 작업은 더욱 더뎌진 것이 현실이다.
- ▶ 또한 계산할 때 바코드 스캔 대신 컨베이어 벨트 위에서 Focal 카메라로 스캔하고 계산할 수 있다. 종업원이 포장에 집중하면 거래 시간을 60% 단축할 수 있다.
- ▶ 시간별 재고 현황과 품절 상품과 경쟁 상품의 비교 분석을 차트 형식으로 제공하여 매장 운영 인사이트를 얻을 수 있다.



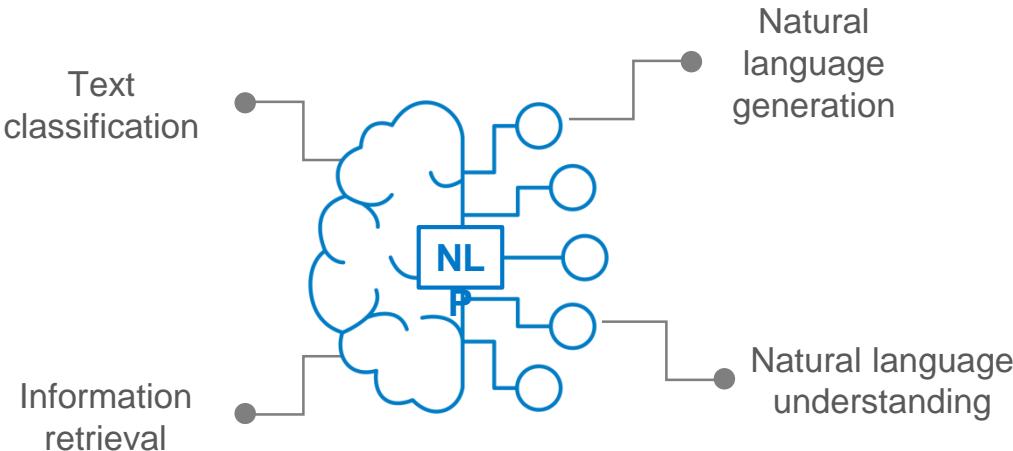
Unit 2.

Applications of Artificial Intelligence

- | 2.1 인공지능의 응용분야
- | 2.2 Image Recognition
- | 2.3 Computer Vision & Machine Vision 분야
- | 2.4 언어지능 분야

Natural Language Processing

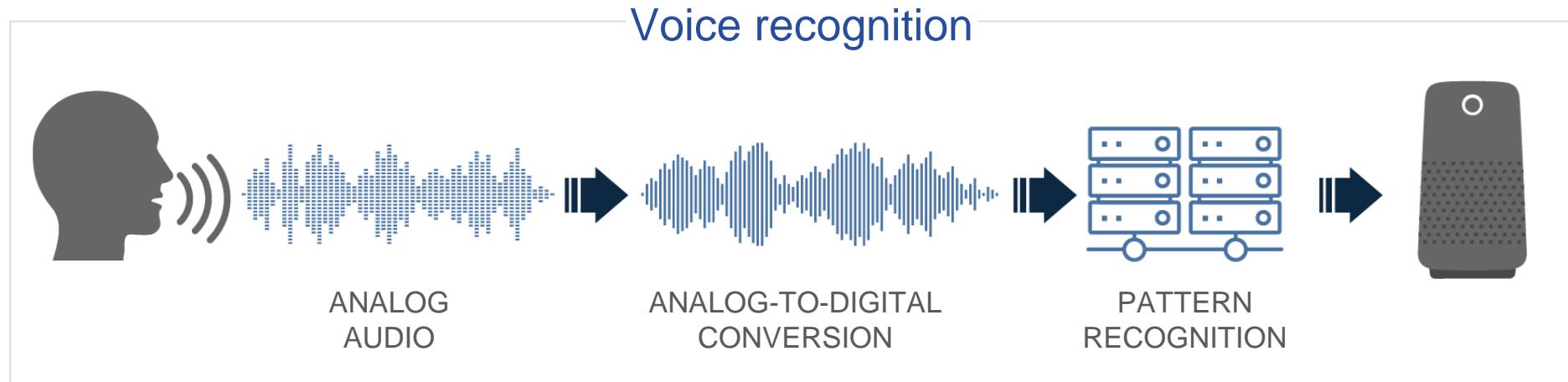
- ▶ Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.
- ▶ Challenges in NLP frequently involve speech recognition, natural language understanding, and natural language generation.
- ▶ Machine Translation, Information Retrieval, Question Answering, Information Extraction, Summarization are main application of natural language processing techniques.



Wikipedia, Expert System

I Speech & Voice Recognition

- ▶ Speech Recognition: recognize words, sentences, and contents spoken by anyone.
 - General dictation, transcribing, using a computer hands-free, medical transcription, automated customer service, etc.
- ▶ Voice Recognition: recognize the accent, pitch, or intonation of a person regardless of language he or she speaks.
 - Speaker verification and speaker identification.



■ 바비 (Barbie) – Hello Barbie

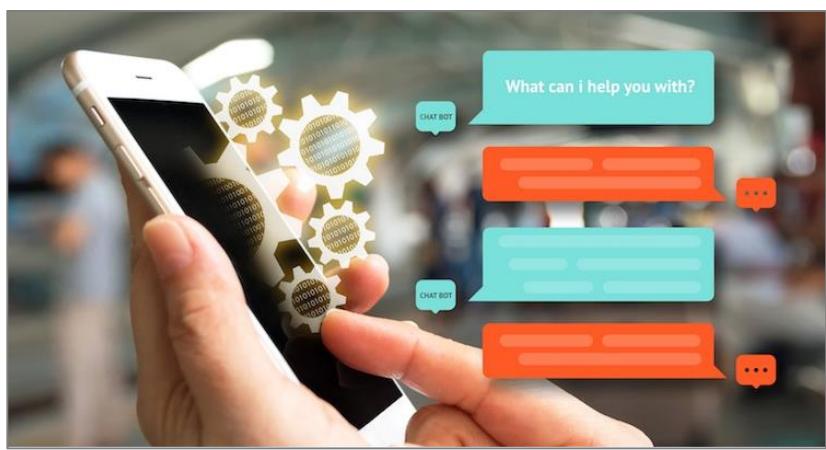
- ▶ 세계적인 인형 제조업체 바비는 바비 인형과 대화할 수 있는 AI 장난감을 개발했다.
 - ▶ Hello Barbie는 자연어 처리, AI를 사용하여 아동의 말을 분석하고 이에 대응한다.
 - ▶ 인형에는 대화 버튼을 누르면 내장된 마이크가 아동이 하는 말을 녹음하고 서버로 전송한 다음 AI가 적절한 응답을 선택하여 인형의 스피커로 다시 전송한다.
 - ▶ Hello Barbie는 아이들이 한 말은 기억했다가 나중에 다시 응답할 수 있다.
- Ex** 좋아하는 가수가 테일러 스위프트라고 말하면 몇 주 후에도 좋아하는 가수가 누구인지에 대해서 대답한다.
- ▶ AI는 3세에서 9세까지의 아동의 대화를 예측할 수 있으며 최대 200회까지 아동과 대화 교환이 가능하다.



<http://hellobarbiefaq.mattel.com/>

I Personetics - Assist

- ▶ Assist는 대화식 AI 기능에 소비자 금융 및 금융 산업에 대한 지식을 결합한 도메인 특화 자문형 챗봇 서비스이다.
- ▶ 인건비 없이 편리하게 24시간 고객 서비스를 제공할 수 있어 글로벌 은행으로부터 인기를 얻고 있다.
- ▶ Assist는 고객의 최신 거래 데이터를 대화에 통합하여 개인의 재무활동 및 은행 관계를 기반으로 응답을 제공한다.
- ▶ 또한 예측 분석을 사용하여 고객의 질문과 문제를 예측하여 고객보다 한발 앞서 적시에 적절한 인사이트와 조언을 제공한다.
- ▶ 자연어 처리를 기반으로 구축되어 송금, 예약, 비밀번호 변경 등의 업무를 대신할 수 있다.
- ▶ Facebook Messenger 같은 메시지 플랫폼과 아마존 Alexa 등의 AI 비서 플랫폼까지 고객 상호작용 채널에 배포할 수 있다.



<https://tearsheet.co/modern-banking-experience/toes-in-the-water-banks-play-around-with-chatbots/>

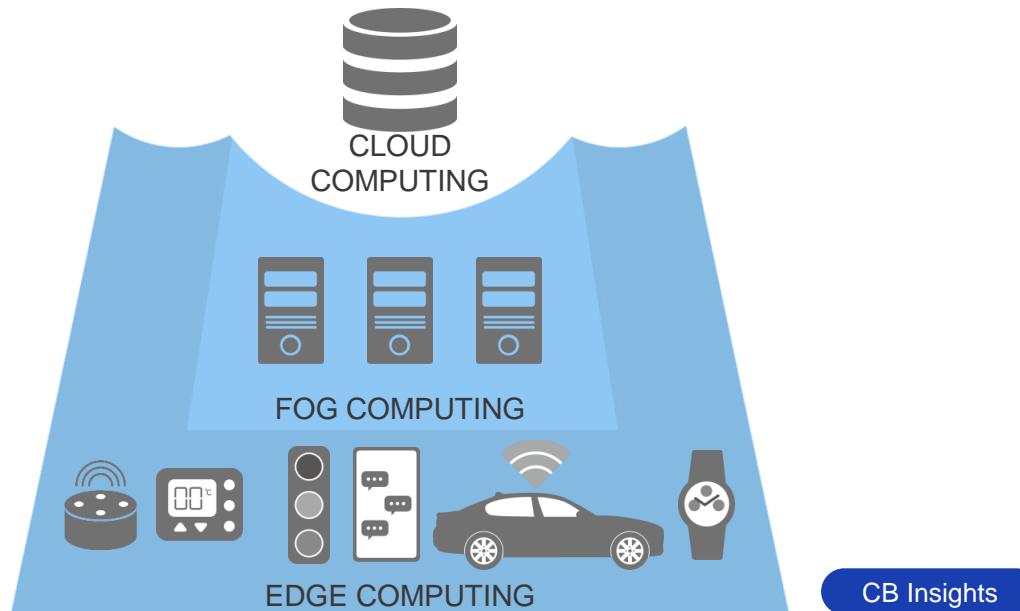
Unit 3.

Trends in Artificial Intelligence

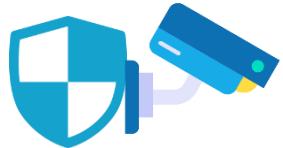
- 3.1 Edge AI
- 3.2 Medical Imaging & Diagnostics
- 3.3 Autonomous Vehicle
- 3.4 Reinforcement Learning
- 3.5 Conversational AI
- 3.6 GAN, XAI, Synthetic Training Data

Definition and Concept

- Running AI algorithms on edge devices, such as a smartphone or a car or even a wearable device, instead of communicating with a central cloud or server provides devices the ability to process information locally and respond more quickly to situations.
- For example, an autonomous vehicle has to respond in real-time to what's happening on the road, and function in areas with no internet connectivity. Decisions are time-sensitive and latency could prove fatal.



Initiatives



Edge AI
use case



In-home smart cameras can recognize that a person(s) has entered an area



Instantaneous driving decisions



Vision for baby monitors, drones, robots, and other devices that can respond to situations without internet connection



IQ cameras,
Deep Lens



AI
processor



custom
AI chip



Myriad X

CB Insights

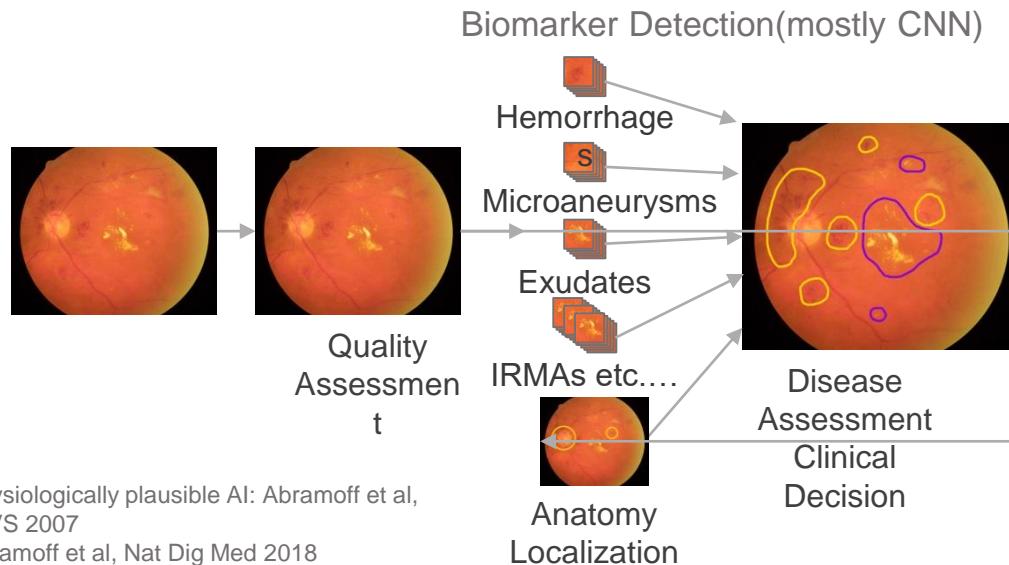
Unit 3.

Trends in Artificial Intelligence

- 3.1 Edge AI
- 3.2 Medical Imaging & Diagnostics
- 3.3 Autonomous Vehicle
- 3.4 Reinforcement Learning
- 3.5 Conversational AI
- 3.6 GAN, XAI, Synthetic Training Data

AI-as-a-medical-device

Autonomous AI algorithm based on biomarkers



- In April 2018, the FDA approved AI based software that screens patients for diabetic retinopathy without the need for a second opinion from a medical specialist.
- The software, IDx-DR, correctly identified patients with “more than mild diabetic retinopathy” 87.4% of the time, and identified those who did not have it 89.5% of the time.
- IDx is one of the many AI based software products approved by the FDA for clinical commercial applications in recent months.

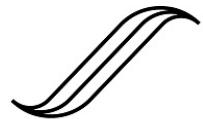
CB Insights, IDx

AI-as-a-medical-device



Diabetic Retinopathy

- ▶ FDA clearance
- ▶ AI based software that screens patients for diabetic retinopathy without confirmation of medical specialist



A R T E R Y S

Liver and Lung AI
Lesion

- ▶ FDA clearance
- ▶ GE Ventures-backed startup
- ▶ Oncology AI suite initially focused on spotting lung and liver lesions.



Viz.ai

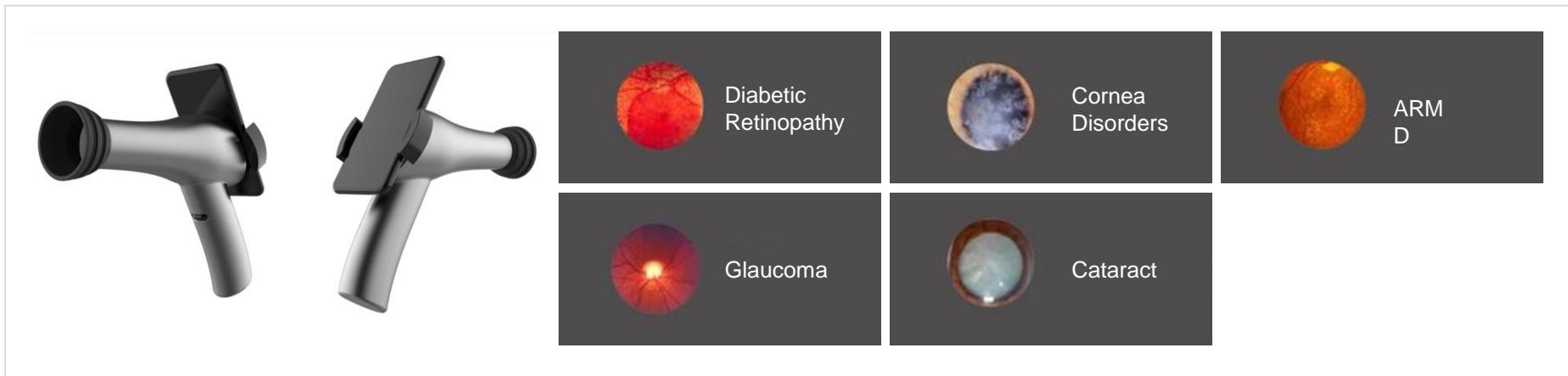
CT-Scan Analysis for
Strokes

- ▶ FDA clearance
- ▶ Analyze CT scans and notify healthcare providers of potential strokes in patients
- ▶ \$21M Series A round from Google Ventures and Kleiner Perkins Caufield & Byers

CB Insights

Portable Ophthalmoscope Based on AI (Project Powered by Samsung)

- Developed a portable diagnostic device and diagnosis support algorithm based on image processing for underserved people in Vietnam.
- 40,000 images are being accumulated per month.
- Detect anomaly from eye images based on Machine learning and classify the image for triage.



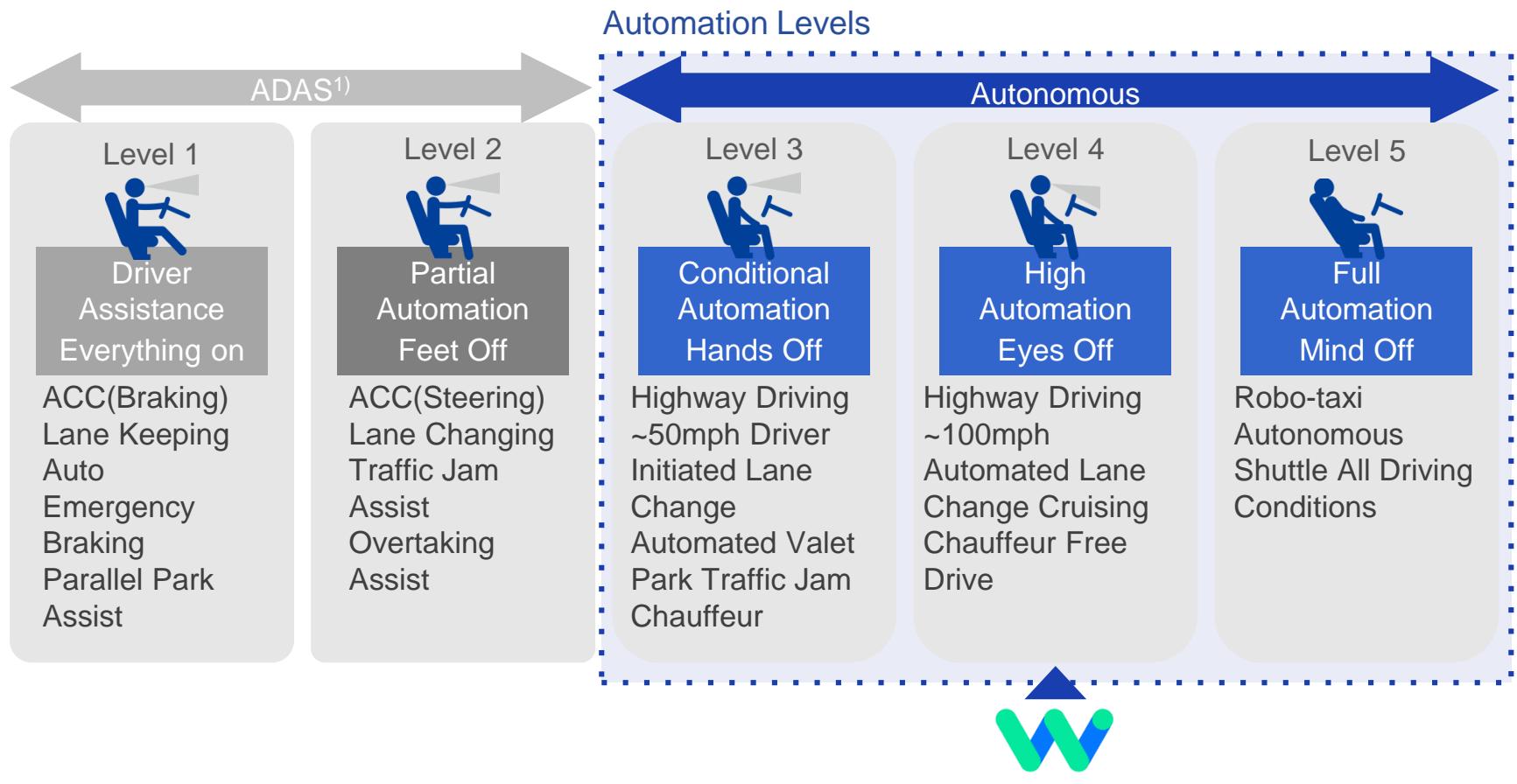
Samsung, Project BOM

Unit 3.

Trends in Artificial Intelligence

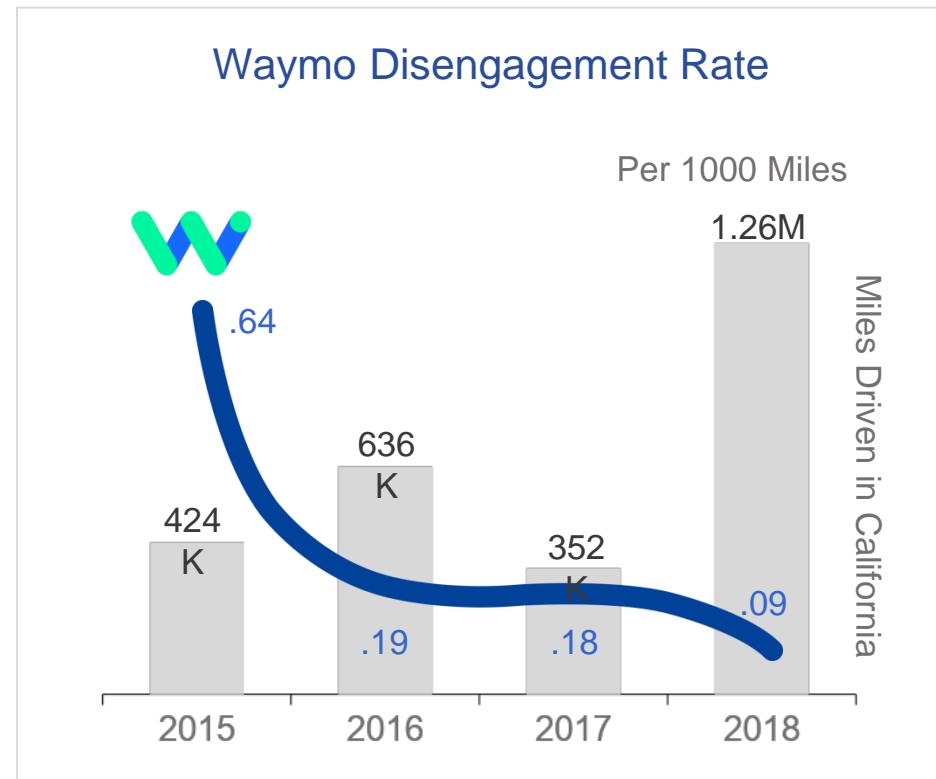
- 3.1 Edge AI
- 3.2 Medical Imaging & Diagnostics
- 3.3 Autonomous Vehicle
- 3.4 Reinforcement Learning
- 3.5 Conversational AI
- 3.6 GAN, XAI, Synthetic Training Data

I Level of Automation



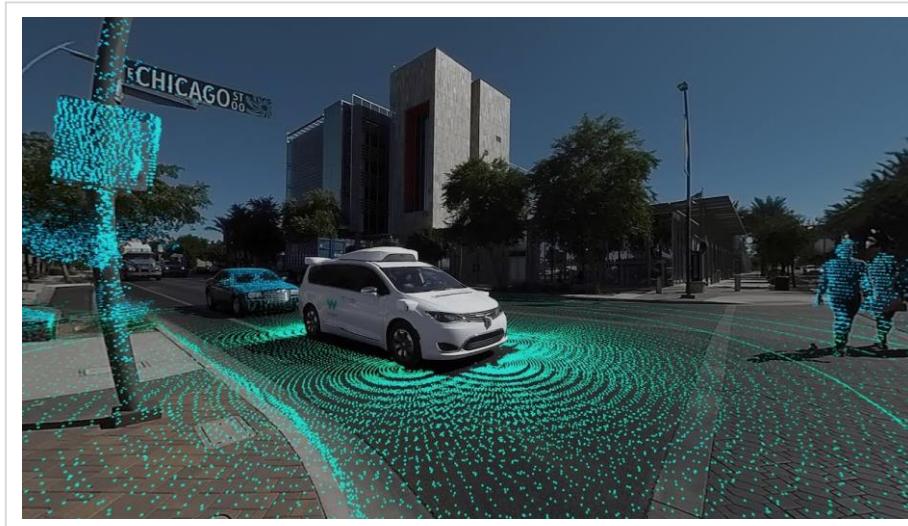
Miles per Disengagement

- ▶ All companies actively testing self-driving cars on public roads in California are required to disclose the number of miles driven and the frequency in which human drivers were forced to take control of their driverless vehicles - also known as a "disengagement".
- ▶ Waymo said its autonomous system only disengaged at a rate of once every 11,017 miles. GM Cruise reported the second-lowest disengagement rate, with a safety driver taking over every 5,205 miles
- ▶ Other startups making progress include Zoox (0.50 disengagements per 1,000 miles), Nuro (0.97 disengagements per 1,000 miles), and Pony.ai (0.98 disengagements per 1,000 miles).



Global Market

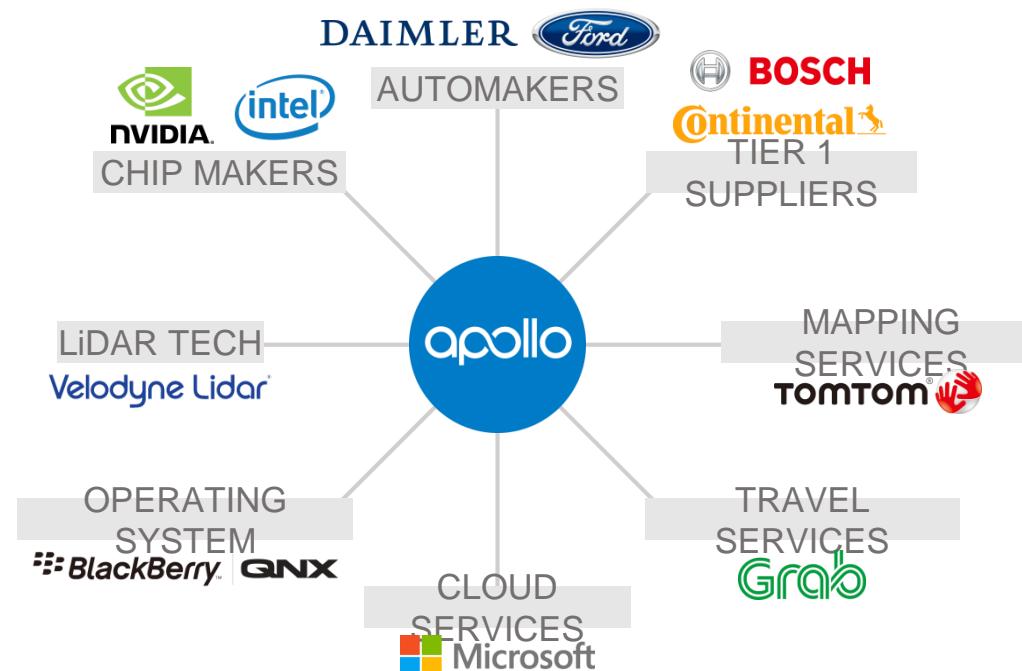
- ▶ Google has made a name for itself in the auto space. Its self-driving project Waymo is the first autonomous vehicle developer to deploy a commercial fleet of AVs.
- ▶ Investors remain confident in companies developing the full autonomous driving stack, pouring hundreds of millions of dollars into GM's Cruise Automation (\$750M from Honda in October 2018 and \$900M from SoftBank in May prior) and Zoox (\$500M in July 2018). Other startups here include Drive.ai, Pony.ai, and Nuro.



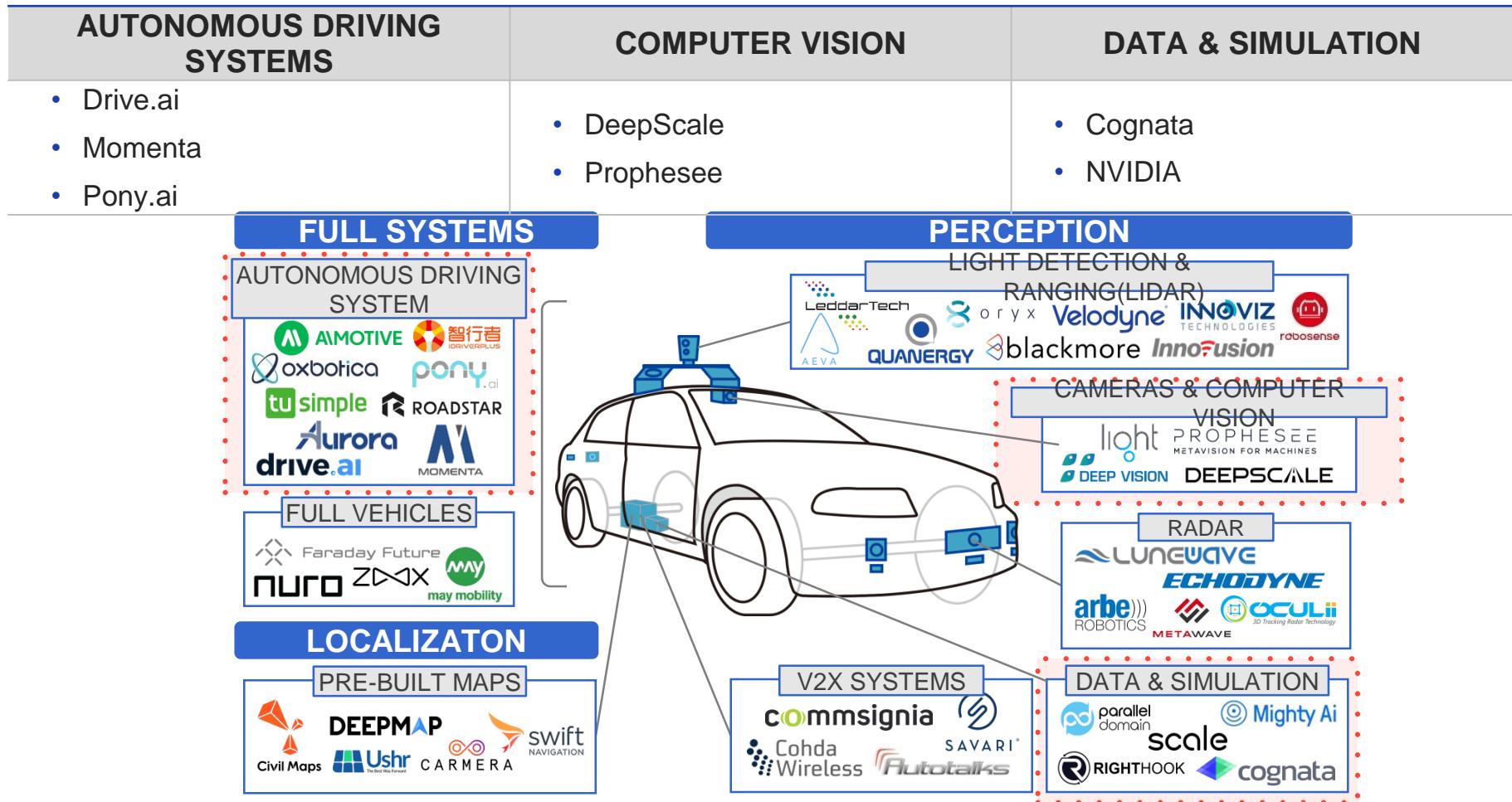
<https://www.forbes.com/sites/alanohnsman/2018/03/02/waymo-is-millions-of-miles-ahead-in-robot-car-tests-does-it-need-a-billion-more/#c854b381ef4c>

China Market

- ▶ The Chinese science ministry announced last year that the nation's first wave of open AI platforms will rely heavily on Baidu for autonomous driving
- ▶ In April 2017, Baidu announced a one-of-a-kind open platform, called Apollo, for autonomous driving solutions, roping in partners from across the globe



Unbundling the Autonomous Vehicle



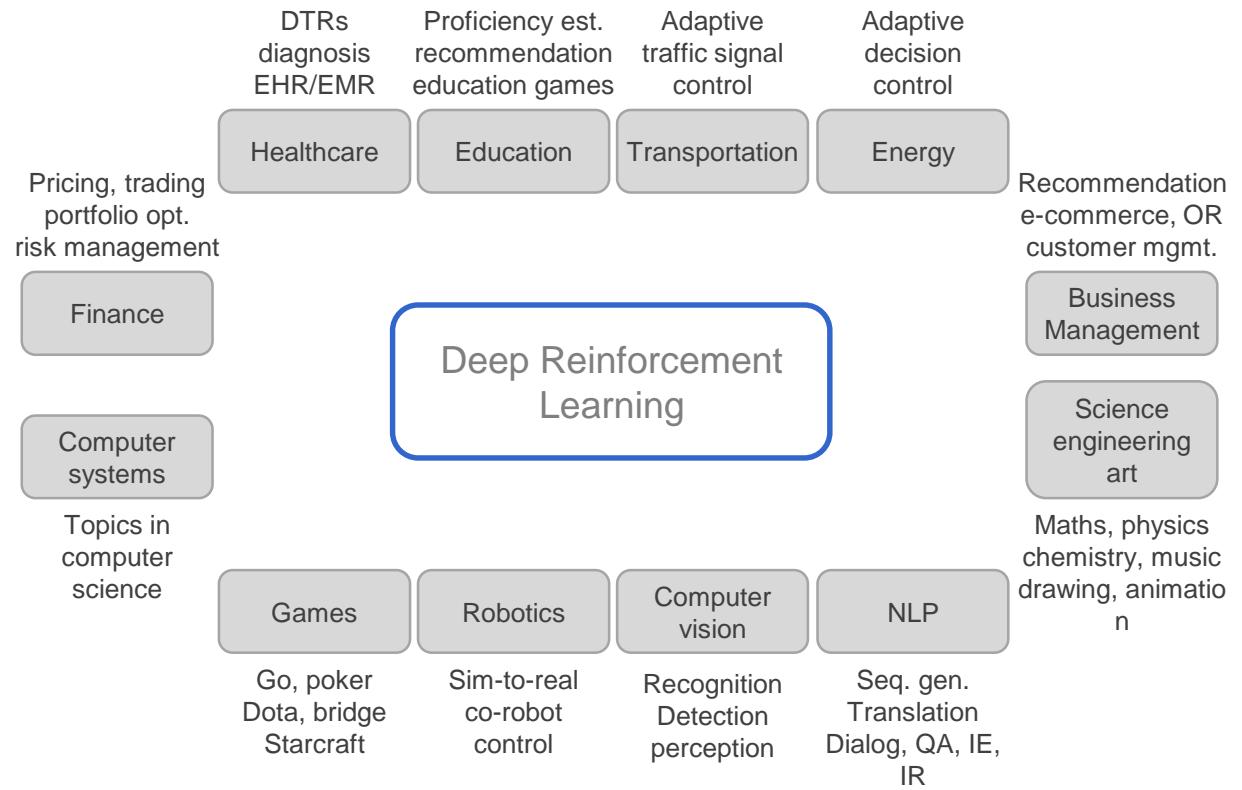
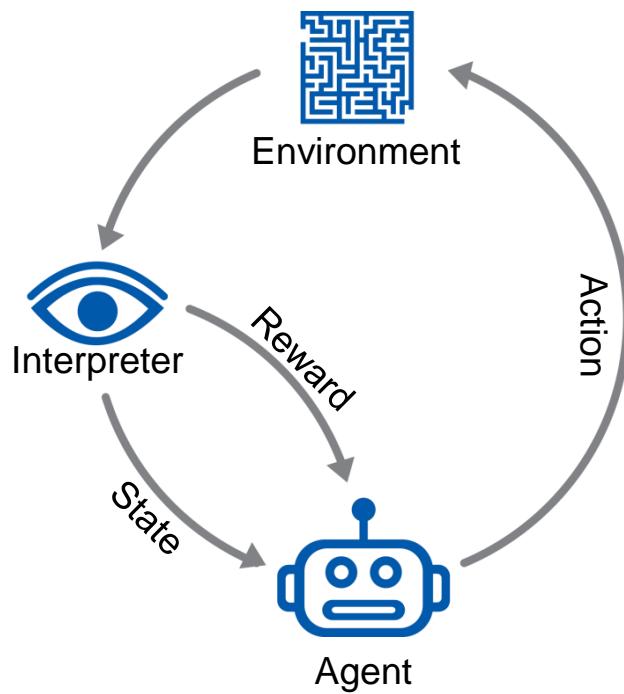
Unit 3.

Trends in Artificial Intelligence

- 3.1 Edge AI
- 3.2 Medical Imaging & Diagnostics
- 3.3 Autonomous Vehicle
- 3.4 Reinforcement Learning
- 3.5 Conversational AI
- 3.6 GAN, XAI, Synthetic Training Data

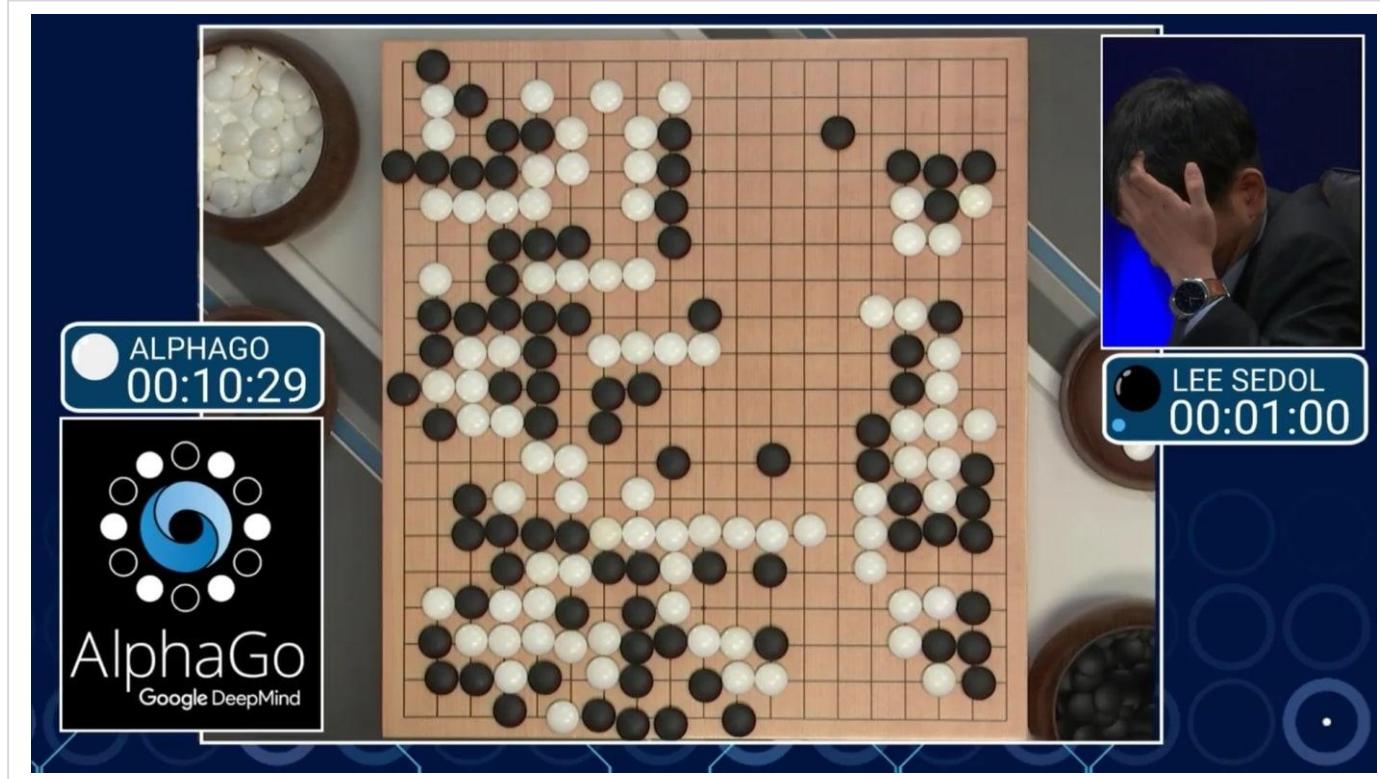
Reinforcement Learning

- Definition and Application



Yuxi Li, Deep Reinforcement Learning, arXiv, 2018

I Definition and Application



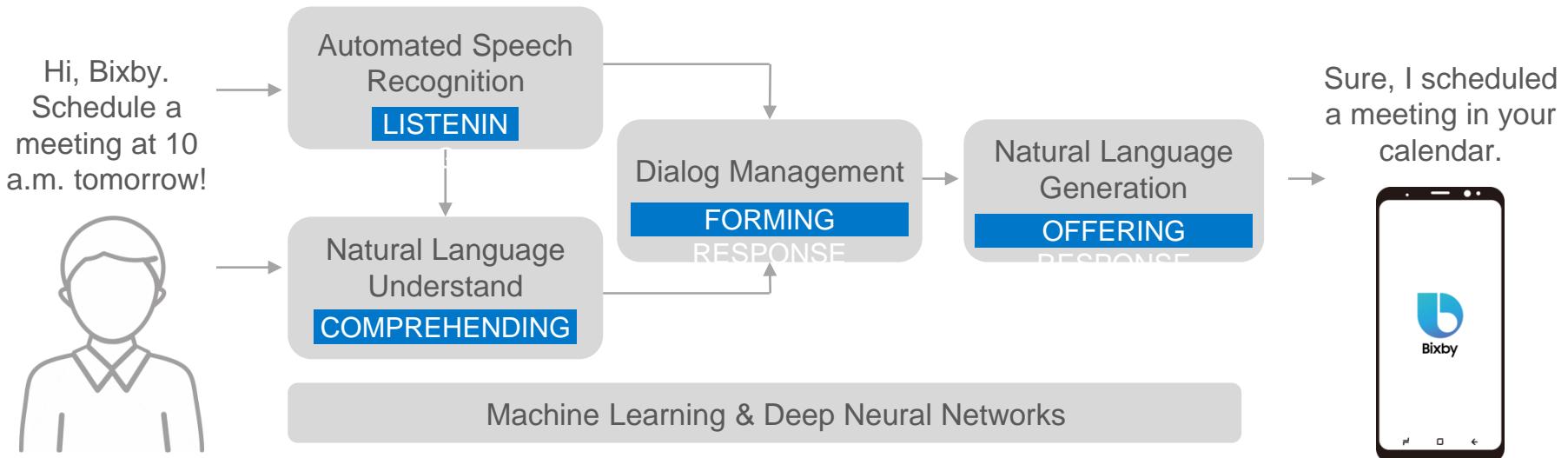
<https://dimensionless.in/reinforcement-learning-super-mario-alpha-go/>

Unit 3.

Trends in Artificial Intelligence

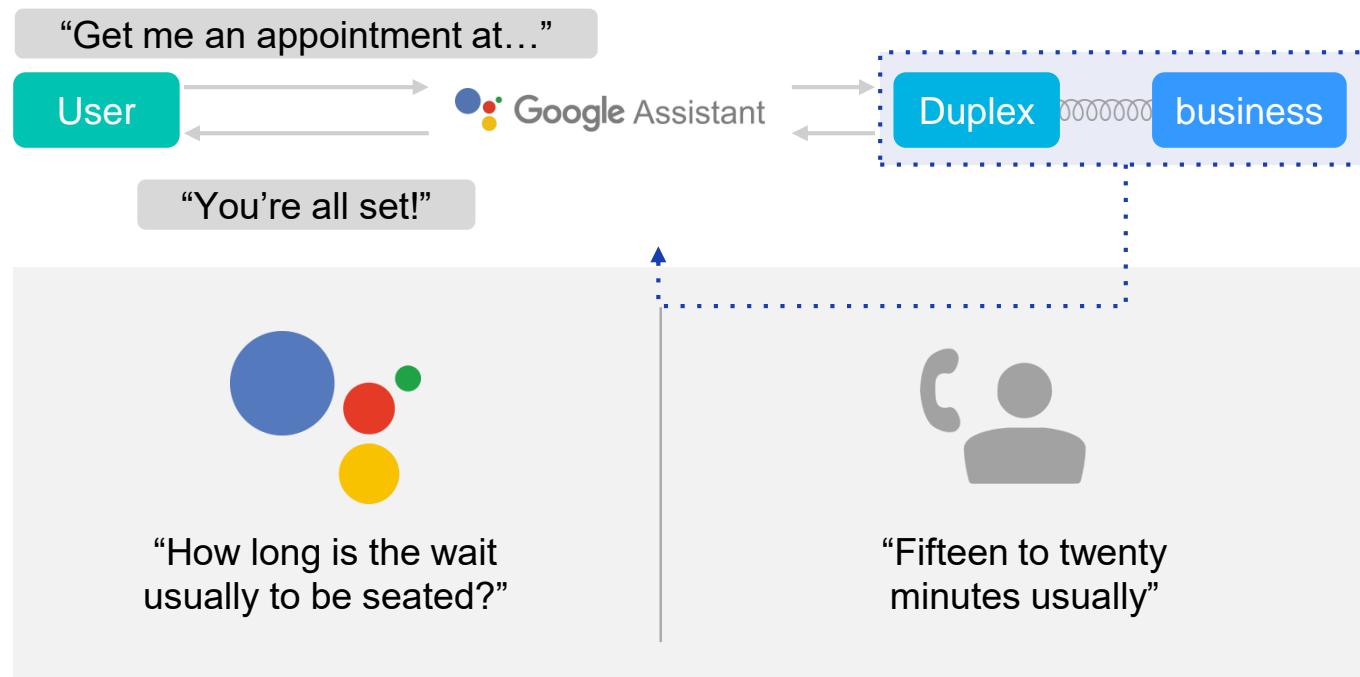
- 3.1 Edge AI
- 3.2 Medical Imaging & Diagnostics
- 3.3 Autonomous Vehicle
- 3.4 Reinforcement Learning
- 3.5 Conversational AI
- 3.6 GAN, XAI, Synthetic Training Data

Voice Assistant



CB Insights, Interactions

Duplex



CB Insights, Google

Unit 3.

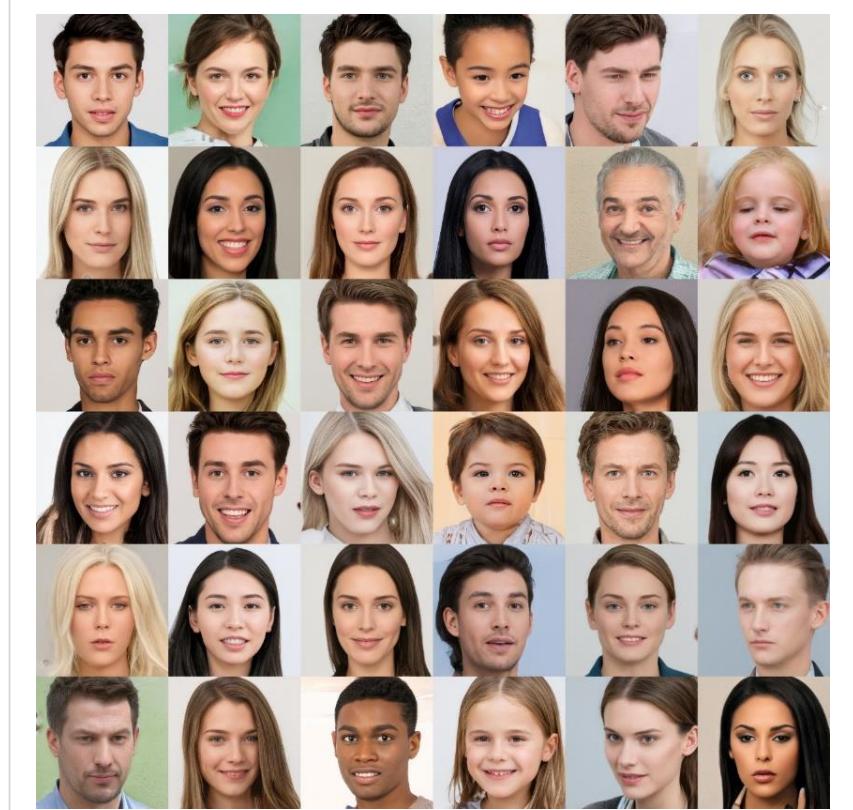
Trends in Artificial Intelligence

- 3.1 Edge AI
- 3.2 Medical Imaging & Diagnostics
- 3.3 Autonomous Vehicle
- 3.4 Reinforcement Learning
- 3.5 Conversational AI
- 3.6 GAN, XAI, Synthetic Training Data

Generative Adversarial Network (GAN)

| Deep fake

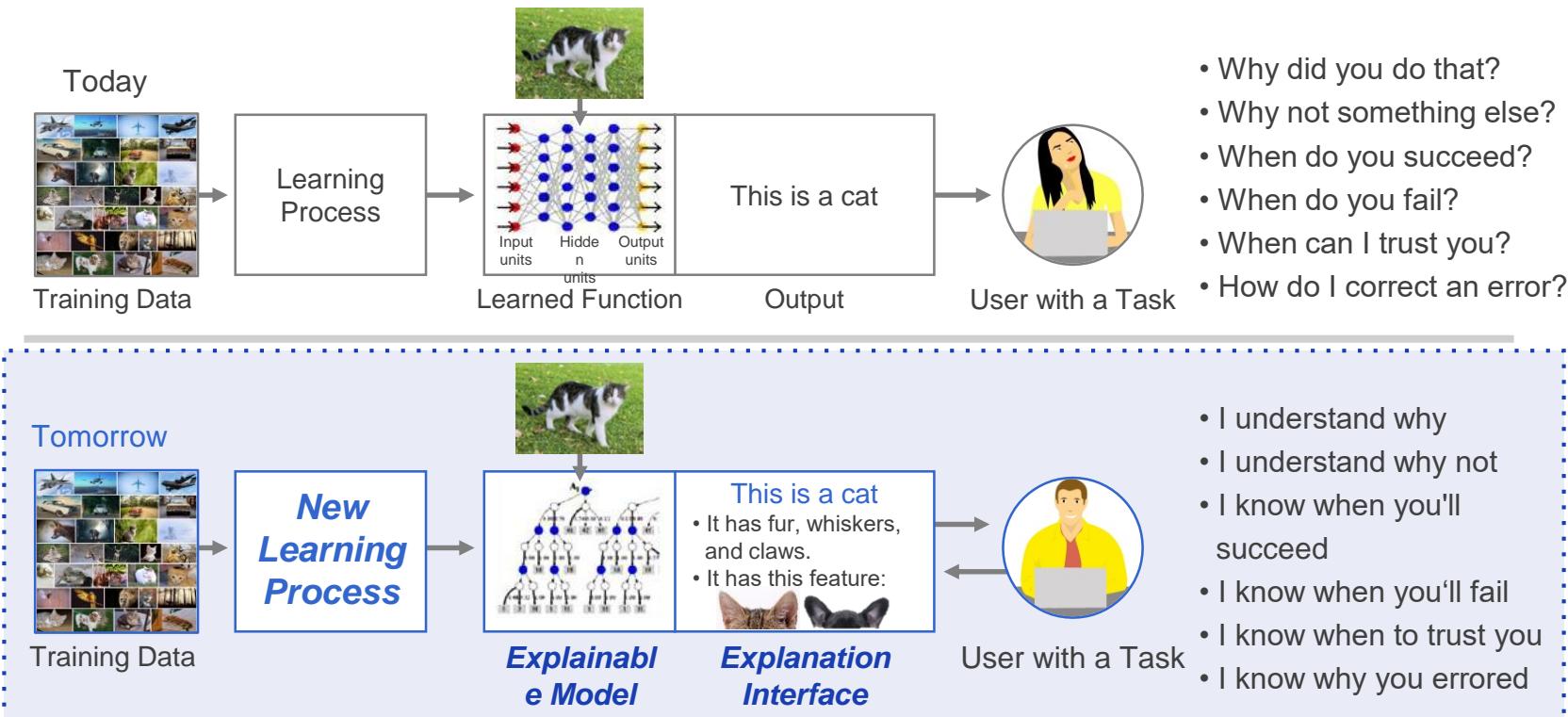
- ▶ Pictures on this page are all generated by artificial intelligence.
- ▶ Generative Adversarial Network, also known as GAN, brought a groundbreaking innovations in artificial intelligence field, especially in generating fake images and videos.
- ▶ In 2018, researchers at Carnegie Mellon University created a video in which two people showed exactly the same facial expression simultaneously – it is called deep fake.
- ▶ With recycled-GAN algorithm, the quality of fake video was significantly enhanced.



Carnegie Mellon University

XAI (eXplainable AI)

| What are we trying to do?

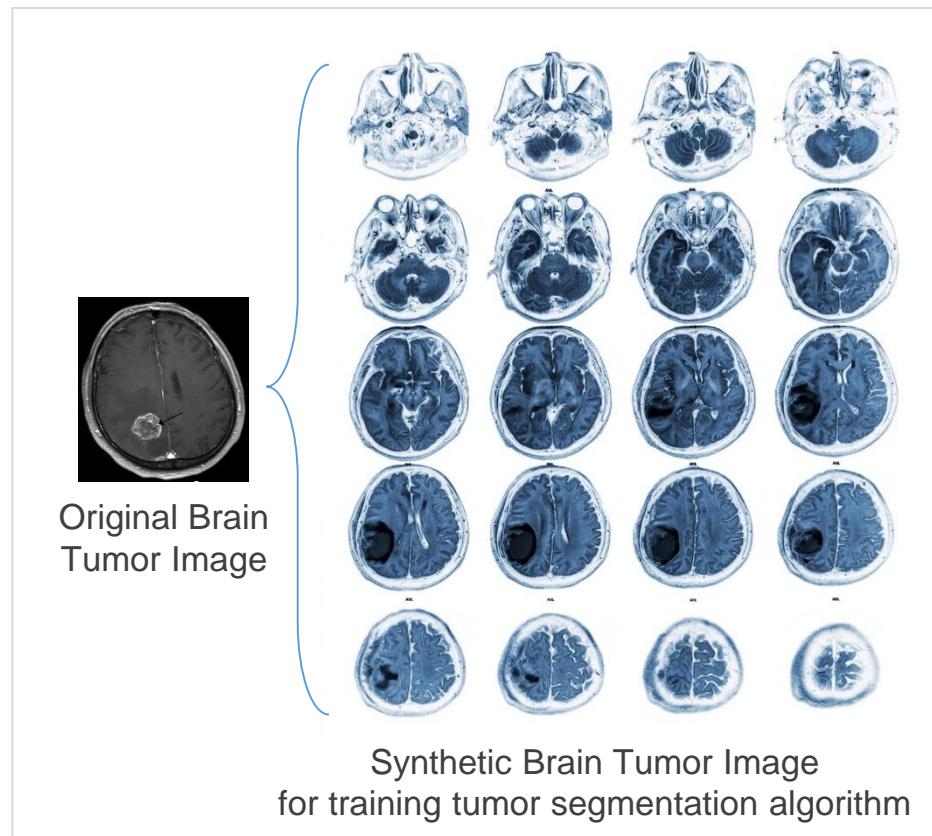


DARPA

Synthetic Training Data

I Synthetic Brain Tumor MRI images

- ▶ Researchers generated synthetic medical image of brain tumor by training a generative adversarial network algorithm with publicly available MRI data
- ▶ With synthetically generated data,
 - they were able to improve the accuracy of tumor segmentation.
 - and they used the generative model as an anonymization tool which mitigates the risk of using patients' medial data without permission.



DARPA

Unit 4.

AI 트렌드 및 시장

| 4.1 AI 트렌드

| 4.2 AI 시장

| 4.3 에너지 분야

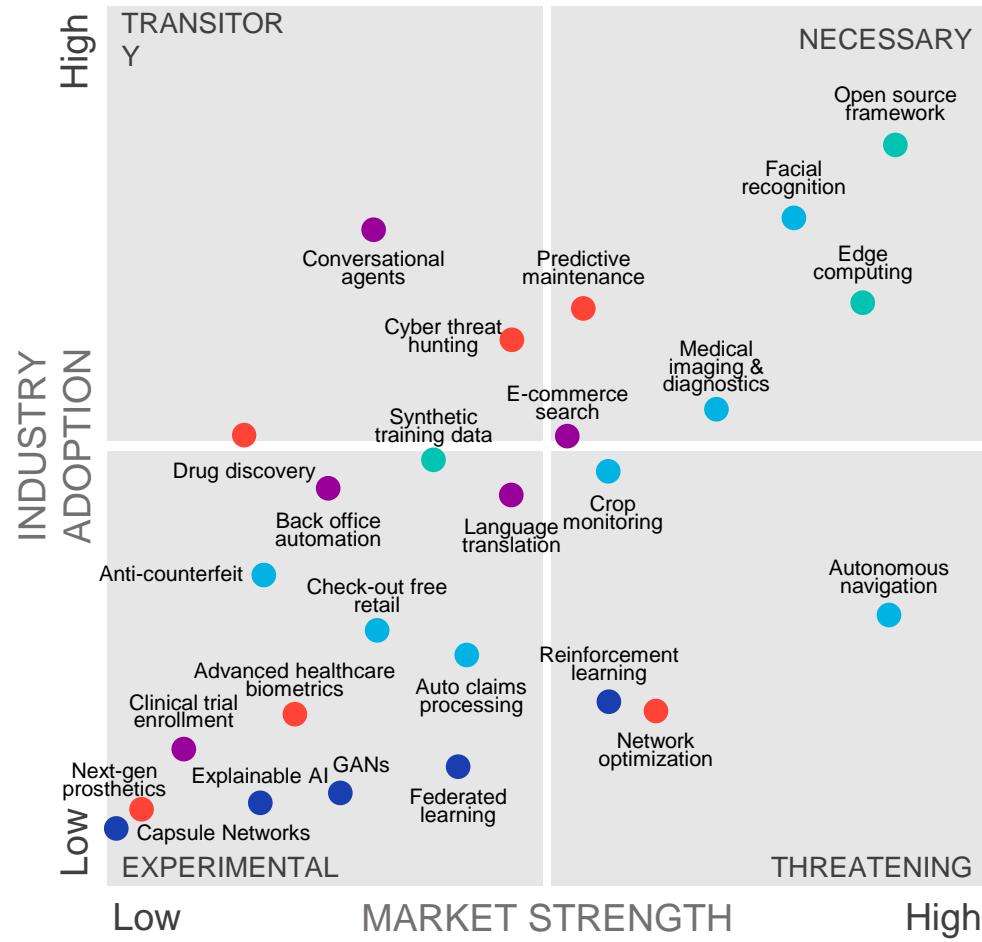
| 4.4 금융서비스 분야

| 4.5 AI를 활용한 정부

| 4.6 보건 의료분야

| 4.7 사물인터넷 및 AI 활용 농업 분야

Trend Matrix



- Application: Computer vision
- Application: Natural language processing / synthesis
- Application: Predictive intelligence
- Architecture
- Infrastructure

Artificial Intelligence Trends 2019, CB Insights

Unit 4.

AI 트렌드 및 시장

| 4.1 AI 트렌드

| 4.2 AI 시장

| 4.3 에너지 분야

| 4.4 금융서비스 분야

| 4.5 AI를 활용한 정부

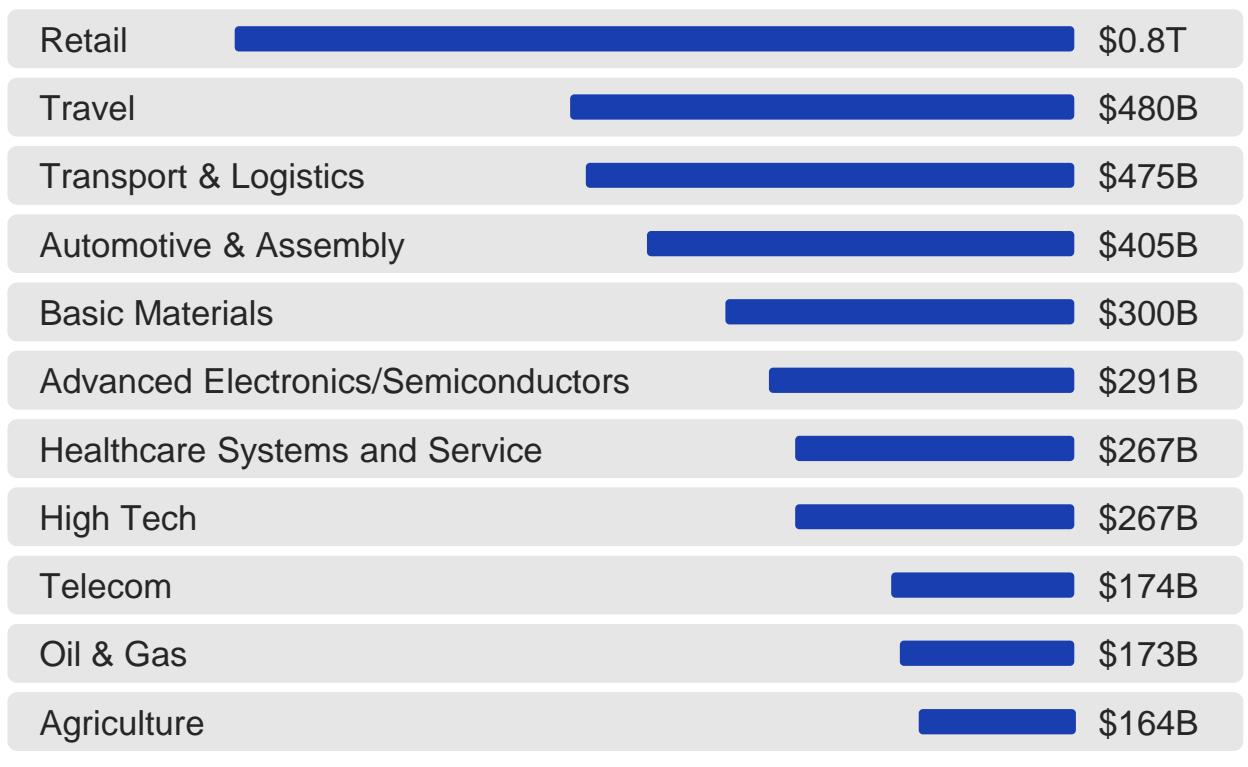
| 4.6 보건 의료분야

| 4.7 사물인터넷 및 AI 활용 농업 분야

AI Value Creation by 2030

- Innovation created by artificial intelligence (AI) started from software industry and will soon revolutionize all the other industries. Value created by artificial intelligence will reach 13 trillion dollars by 2030.

\$13
Trillion



McKinsey Global Institute

Unit 4.

AI 트렌드 및 시장

| 4.1 AI 트렌드

| 4.2 AI 시장

| 4.3 에너지 분야

| 4.4 금융서비스 분야

| 4.5 AI를 활용한 정부

| 4.6 보건 의료분야

| 4.7 사물인터넷 및 AI 활용 농업 분야

AI in Sustainable Energy

- Green AI establishes a connection between computing power and carbon emissions, allowing one to manage the carbon cost of large AI computing demands to some extent.
 - ▶ Today, greenhouse emissions from the information and communication technology (ICT) industry account for 2% of total emissions (as much as all air traffic), warning that failure to confirm this will increase up to 14% of global emissions. AI itself can be used for eco-friendly projects.
 - ▶ Examples
 - Space Intelligence, a British startup, applies machine learning and AI to satellite data to solve environmental problems such as re-seasoning, allowing the industry to take corrective action.
 - Google has promised not to use carbon by 2030. At the recent Google I/O 2021 event, the company explained these efforts in detail by announcing plans to use geothermal energy to power data centers in Nevada.
 - Kate Brandt, Google's head of sustainability, said, "We are the first company to sign a contract for next-generation geothermal power."



Green AI는 컴퓨팅 파워와 탄소 배출량 간의 연결을 설정하여 대규모 AI 컴퓨팅 수요의 탄소 비용을 어느 정도 관리할 수 있도록 한다.

The collage consists of four images arranged in a 2x2 grid. The top-left image shows a green globe with a network of lines and a 'ECO' logo, representing the connection between computing power and carbon emissions. The top-right image shows a field of wind turbines at sunset with the text 'AI is Helping Forecast the Wind, Manage Wind Farms'. The bottom-left image shows a modern data center building with the text 'Green Data Centers Seen as Helping Manage AI Power Demands'. The bottom-right image shows industrial robots in a factory setting with the text 'Marine Battery Production Making Gains Through Automation'.

Green AI Seeks to Connect Compute Power to Carbon Emissions

May 27, 2021

AI is Helping Forecast the Wind, Manage Wind Farms

Green Data Centers Seen as Helping Manage AI Power Demands

Marine Battery Production Making Gains Through Automation

<https://www.aitrends.com/category/energy/>

Unit 4.

AI 트렌드 및 시장

| 4.1 AI 트렌드

| 4.2 AI 시장

| 4.3 에너지 분야

| 4.4 금융서비스 분야

| 4.5 AI를 활용한 정부

| 4.6 보건 의료분야

| 4.7 사물인터넷 및 AI 활용 농업 분야

AI in Financial Services

| 2020년에는 사이버 공격과 신원 사기 손실이 급격히 증가했으며, 이는 대유행이 원격 근무를 표준으로 만들고 더 높은 수준의 보호를 달성하기 위해 AI와 생체 인식이 결합되는 단계를 설정했기 때문입니다.

- ▶ 한 연구에 따르면 전 세계 은행들은 2020년 2월과 4월 사이에 사이버 공격이 238% 급증했습니다.
- ▶ Javelin Strategy & Research의 연구에 따르면 Pymnts.com 의 최근 계정에 따르면 사기꾼이 훔친 개인 정보를 사용하여 합성 ID를 생성함에 따라 ID 사기 손실이 작년에 560억 달러로 증가했습니다.
- ▶ 또한 자동화된 봇 공격은 7월과 12월 사이에 다양한 산업 분야의 기업을 대상으로 1억 건이 증가했습니다.
- ▶ 더 나은 보호를 위해 노력하는 회사는 고객의 삶을 더 어렵게 만들 위험이 있습니다. 또 다른 연구에서는 금융 기관의 40%가 적법한 고객의 온라인 행동을 사기꾼의 행동으로 자주 착각하는 것으로 나타났습니다.

| 보안 이점을 제공하는 AI 및 생체 인식

- ▶ AI는 개인의 신원을 확인할 때 더 높은 정확도와 속도를 제공하거나 사이버 범죄자가 자격 증명만 제공하여 정보에 액세스할 수 없도록 생체 데이터를 통합하는 등 디지털 신원 사기를 방지하는 데 사용할 수 있습니다.

| 팬데믹 변화된 소비자 금융 행동, 신원 사기 유발

- ▶ 세계적 대유행은 소비자 금융 행동에 극적인 영향을 미쳤습니다. 소비자는 2020년에 집에서 더 많은 시간을 보내고 전년도보다 거래량이 줄었으며 스트리밍 서비스, 디지털 상거래 및 결제에 크게 의존했습니다.
- ▶ 그들은 또한 일과 개인 생활 모두에 대해 이메일과 문자를 통해 더 많은 통신을 했습니다.



<https://www.aitrends.com/category/financial-services/>

Unit 4.

AI 트렌드 및 시장

| 4.1 AI 트렌드

| 4.2 AI 시장

| 4.3 에너지 분야

| 4.4 금융서비스 분야

| 4.5 AI를 활용한 정부

| 4.6 보건 의료분야

| 4.7 사물인터넷 및 AI 활용 농업 분야

AI in Government

- Wisconsin Department of Workforce Development(DWD)는 팬데믹 기간 동안 급증한 백로그인 AI 클라우드 컴퓨팅 도구의 도움으로 수천 건의 실업 보험 청구에 대한 2020년 백로그를 정리했습니다.
 - ▶ DWD는 매주 평균 157,000건의 청구를 처리하고 있으며 전염병이 시작된 이후 총 20억 달러의 실업 수당을 지급했습니다. GCN 의 최근 계정에 따르면 이 기관은 이제 영업일 기준 3일 이내에 자금을 해제할 수 있습니다.
 - ▶ Google 공공 부문의 주 및 지방 정부 책임자인 Brent Mitchell에 따르면 이 과정은 몇 주 또는 몇 달이 걸렸습니다. 복구는 DWD가 백로그를 돋기 위해 1,300명을 고용, 계약 또는 재배정함으로써 도움이 되었습니다.
 - ▶ DWD는 AI 및 머신 러닝 제품과 함께 Google Cloud의 DocAI 서비스를 사용했습니다. DocAI는 문서에서 중요한 데 이터 추출을 자동화하고 워크플로를 간소화하는 방법을 제안하여 가치를 추가하는 데 사용됩니다. Human-in-the-Loop AI 기능은 사람의 검토를 추가하여 정확도를 높입니다.



“With design thinking, close partnerships with state officials, and combination of modern technologies, DWD’s solutions are customized to maximize benefits to components,” said Mitchell.

AI in Government



AI Helping State and Local Governments Meet Increasing Demands

Pentagon Cancels JEDI Cloud Contract After Legal Battle Caused Delay

Executive Interview: Paul Nemitz, Principal Adviser on Justice Policy for the European Commission, Brussels

<https://www.aitrends.com/category/ai-in-government/>

Unit 4.

AI 트렌드 및 시장

| 4.1 AI 트렌드

| 4.2 AI 시장

| 4.3 에너지 분야

| 4.4 금융서비스 분야

| 4.5 AI를 활용한 정부

| 4.6 보건 의료분야

| 4.7 사물인터넷 및 AI 활용 농업 분야

AI in Healthcare

미국 식품의약국(FDA)은 지난 2월 AI를 통합한 혁신적인 의료기기에 대한 FDA의 승인을 구하는 제조업체가 앞으로 나아가야 할 방향을 설명 하는 의료기기 실행 계획으로 AI 및 기계 학습 소프트웨어를 발표했습니다.

- ▶ 보고서 작성자는 "인공 지능(AI) 및 머신 러닝(ML) 기술은 매일 의료 서비스를 제공하는 동안 생성되는 방대한 양의 데이터에서 새롭고 중요한 통찰력을 도출하여 의료 서비스를 혁신할 잠재력이 있습니다."라고 말했습니다. . .
- ▶ 이 보고서는 FDA가 승인 프로세스를 업데이트하라는 요청에 따라 AI를 사용하여 의료 기기를 진행하는 방법에 대한 논의를 시작하려는 2019년 노력의 후속 조치입니다. FDA 계획 작성자는 "FDA의 전통적인 의료 기기 규제 패러다임은 적응형 인공 지능 및 기계 학습 기술을 위해 설계되지 않았습니다."라고 말합니다.
- ▶ FDA 실행 계획은 장치가 개발됨에 따라 주기적인 업데이트를 위한 프로세스와 함께 장치가 어떻게 작동할 것인지에 대한 제조업체의 약속을 찾습니다. 이러한 방식으로 FDA와 제조업체는 시판 전 개발에서 시판 후 성능까지 소프트웨어 제품을 평가할 수 있습니다.
- ▶ 이를 통해 FDA는 의료 기기에서 AI 및 ML 기반 소프트웨어의 힘을 수용하는 동시에 환자의 안전을 보장할 수 있습니다.

AI in Healthcare

| 의료 기기 분야에서 그는 다음을 포함하여 AI SaMD의 유망한 대상 용도를 식별했습니다.

- ▶ 심장병 진단. 고감도 심장 트로포닌 I 농도와 짹을 이루는 연령 및 성별을 통합하는 기계 학습 알고리즘(심근 허혈 상해 지수)은 3,013명의 환자의 데이터를 활용하여 AI 플랫폼을 훈련하는데 사용되었습니다.
- ▶ 그런 다음 이 플랫폼은 심근경색이 의심되는 7,998명의 환자를 대상으로 테스트되었습니다. 82.5%의 민감도와 92.2%의 특이도로 의사를 능가하는 것으로 나타났습니다.
- ▶ 망막병증 감지. 당뇨병성 망막증(DR)은 전 세계적으로 예방 가능한 실명의 주요 원인 중 하나입니다. A의 연구 안과의 미국 아카데미에서 출판, 당뇨병 환자에서 75,137 개의 안저 이미지의 총 기차에 사용 DR 가진 사람들에서 건강한 fundi를 구별하는 인공 지능 엔진을 테스트했다. 그 결과 민감도와 특이도가 각각 94%와 98%로 인상적이었습니다.
- ▶ 바이탈 사인 모니터링용 바이오센서. 바이오센서 기반 장치는 방대한 데이터 세트를 생성합니다. AI는 추세와 질병 발생 확률을 예측하는데 사용될 수 있습니다.
- ▶ 현장 진료(POC) 진단을 위한 심장 모니터링 기반 바이오센서에 AI를 통합한 것이 좋은 예입니다. 머신 러닝 알고리즘은 실시간 건강 모니터링을 위해 마이크로칩 기반 심장 바이오센서와 함께 사용되며 적시에 정확한 임상 결정을 제공합니다.

AI in Healthcare



<https://www.aitrends.com/category/healthcare/>

Unit 4.

AI 트렌드 및 시장

| 4.1 AI 트렌드

| 4.2 AI 시장

| 4.3 에너지 분야

| 4.4 금융서비스 분야

| 4.5 AI를 활용한 정부

| 4.6 보건 의료분야

| 4.7 사물인터넷 및 AI 활용 농업 분야

AI in Healthcare

- 샌프란시스코의 Plenty에 있는 농업 기술(ag-tech) 스타트업은 AI와 로봇을 사용하여 연중 내내 실내에서 수직으로 작물을 심습니다.
 - Plenty의 수직 농장 접근 방식은 단 2에이커에서 720에이커의 평평한 농장과 동일한 양의 과일과 채소를 생산할 수 있습니다.
 - Plenty는 AI, 드론, 로봇 및 IoT 센서를 포함한 새로운 기술 접근 방식을 사용하여 자본 시장에서 수십억 달러의 투자를 지원받는 수백 개의 ag-tech 신생 기업 중 하나입니다.
 - Plenty의 실내 온도 조절식 농장에는 천장에 매달린 수직으로 자라는 식물이 줄지어 있습니다. 태양을 모방한 LED 조명이 식물을 비춥니다. 로봇이 그들을 이리저리 움직입니다. AI는 물, 온도, 빛의 모든 변수를 관리합니다. AI는 더 나은 작물을 재배하는 방법을 지속적으로 학습하고 최적화합니다.
- 또한 수직 농장은 도시 지역에 위치하여 많은 운송 마일이 제거된 지역 생산 식품을 생산할 수 있습니다. 현지에서 생산되는 작물의 이점에는 운송 차량의 CO₂ 배출량 감소와 잠재적으로 소비자의 가격 인하가 있습니다.
- "COVID-19로 인한 공급망 붕괴와 올해 캘리포니아 산불과 같은 자연적 혼란은 예측 가능하고 내구성 있는 제품 공급의 필요성이 수직 농업에서만 올 수 있음을 보여줍니다."라고 Storey는 말했습니다



AI in Healthcare

Ag-tech Employing AI and Range of Tools With Dramatic Results



July 29, 2021

Hyperconnectivity, Not Well Understood, Ties All Smart Devices



Predictive Maintenance is a Killer AI App



Pandemic Has Spurred CIOs to Crystallize the IT Strategy

<https://www.aitrends.com/category/iot/>

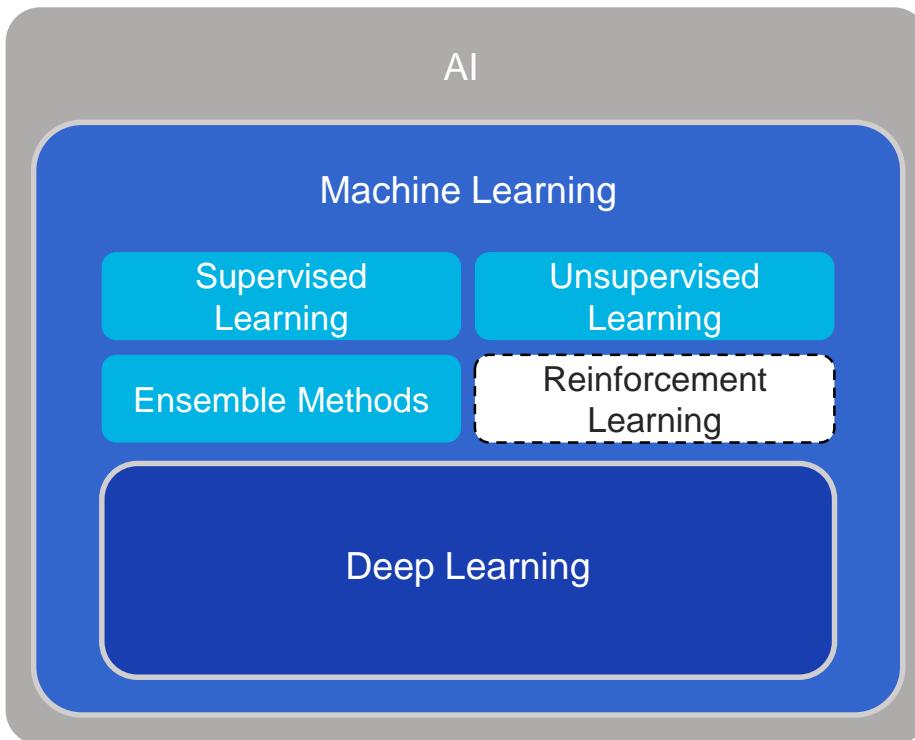
Unit 5.

Course Roadmap

- 5.1 Artificial Intelligence Course Roadmap
- 5.2 Category of Machine Learning Techniques (Full)

Artificial Intelligence Course Roadmap

- Course Coverage



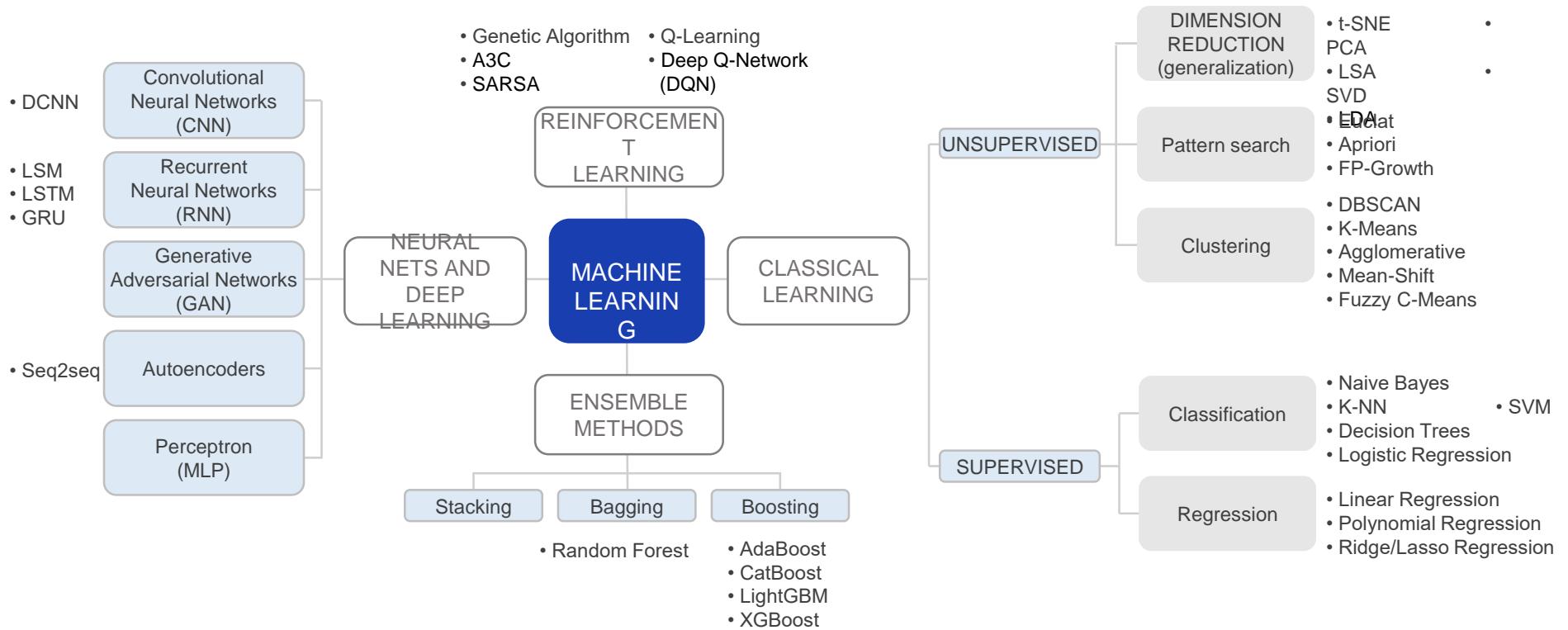
- ▶ Artificial Intelligence course in Samsung Innovation Campus will deal with the majority of key concepts and techniques of AI, except reinforcement learning.
- ▶ Since reinforcement learning is an advanced AI technique, it is not easy to learn in a short-term period. For those who have interest in the topic, it is highly recommended to start learning by yourself during the course or after completing the course.
- ▶ The course will be focused mainly on practical knowledge on how to code by yourself.

Unit 5.

Course Roadmap

- 5.1 Artificial Intelligence Course Roadmap
- 5.2 Category of Machine Learning Techniques (Full)

Category of Machine Learning Techniques (Full)



End of chapter Quiz

Quiz #0201 ~ #0204

Duration : 2.5 Hours

A photograph of a person working at a desk. They are wearing a light-colored sweater and holding a brown paper coffee cup with a black lid in their right hand. Their left hand is on a black computer keyboard. In the background, there are two computer monitors displaying code or text. On the desk, there is also a white electronic device, possibly a calculator or a small laptop, and some papers. The scene is lit from the side, creating shadows and highlights.

End of Document



Together for Tomorrow! Enabling People

Education for Future Generations

©2021 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.