

Machine Learning

# 머신러닝 소개

류영표 강사

youngpyoryu@dongguk.edu



Copyright © "Youngpyo Ryu" All Rights Reserved.

This document was created for the exclusive use of "Youngpyo Ryu".

It must not be passed on to third parties except with the explicit prior consent of "Youngpyo Ryu".



# 류영표

## Youngpyo Ryu

동국대학교 수학과/응용수학 석사수료

現 Upstage AI X 네이버 부스트 캠프 1기,2기 멘토

現 Innovation on Quantum & CT(IQCT) 이사

前 한국파스퇴르연구소 Image Mining 인턴(Deep learning)

前 (주)셈웨어(수학컨텐츠, 데이터 분석 개발 및 연구인턴)

### 강의 경력

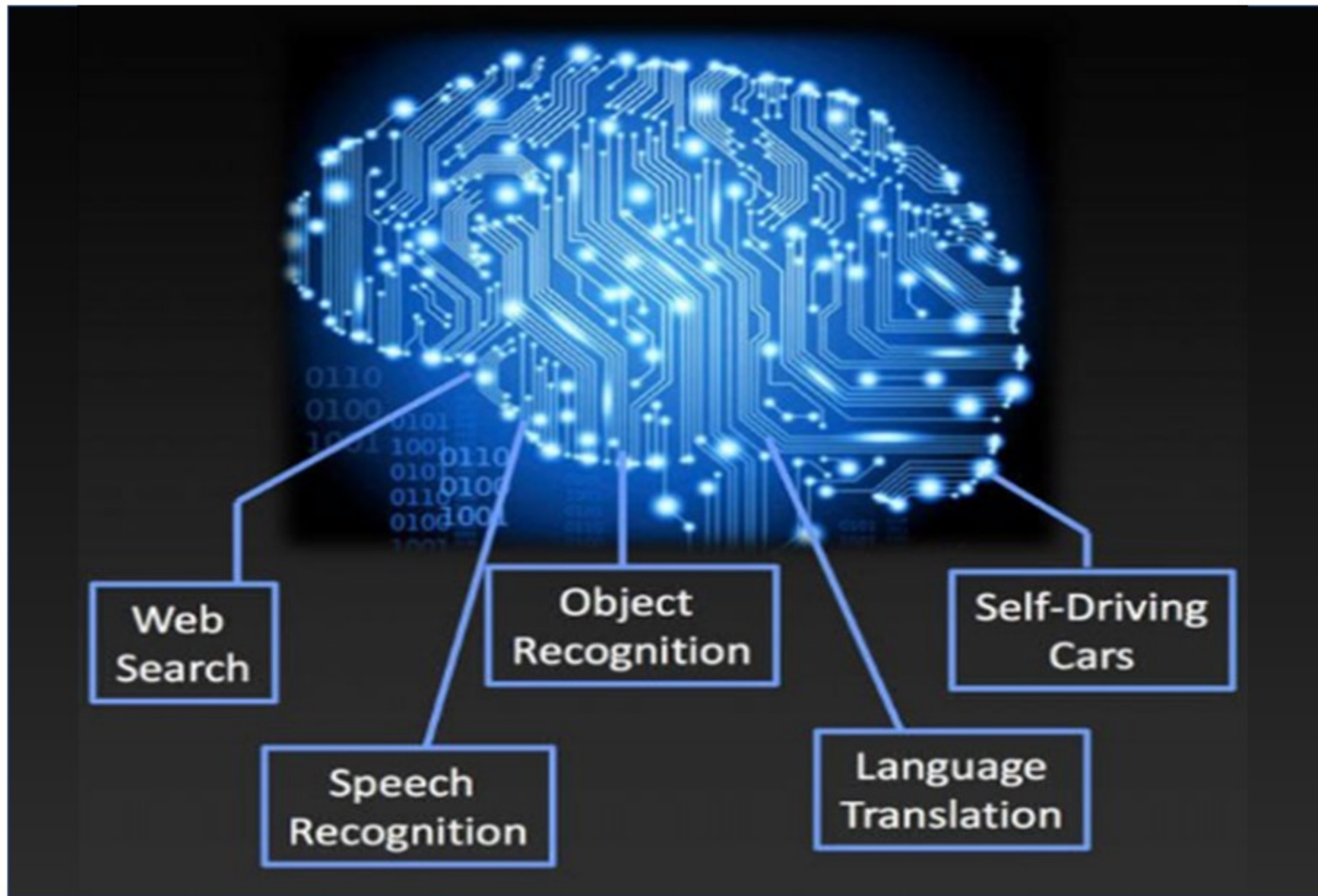
- 현대자동차 연구원 강의 (인공지능/머신러닝/딥러닝/강화학습)
- (재)윌튼블록체인 6일 과정 (파이썬기초, 크롤링,머신러닝)
- (주)모두의연구소 Aiffel 1기 퍼실리테이터(인공지능 교육)
- 인공지능 자연어처리(NLP) 기업데이터 분석 전문가 양성과정 멘토
- 공공데이터 청년 인턴 / SW공개개발자대회 멘토
- 고려대학교 선도대학 소속 30명 딥러닝 집중 강의
- 메가 IT 아카데미(파이썬, 빅데이터 강사)
- 이젠 종로 아카데미(파이썬, ADSP 강사)
- 최적화된 도구(R/파이썬)을 활용한 애널리스트 양성과정(국비과정) 강사
- 한화, 하나금융사 교육
- 인공지능 신뢰성 확보를 위한 실무 전문가 자문위원

### 주요 프로젝트 및 기타사항

- 제1회 인공지능(AI)기반 데이터사이언티스트  
전문가 양성과정 최우수상 수상(Q&A 챗봇)
- 인공지능(AI)기반 데이터사이언티스트 전문가 양성과정 1기 수료
- 제 1회 산업 수학 스터디 그룹 (질병에 영향을 미치는 유전자 정보 분석)
- 제 4,5회 산업 수학 스터디 그룹 (피부암, 유방암 분류)
- 빅데이터 여름학교 참석 (혼잡도를 최소화하는  
새로운 노선 건설 위치의 최적화 문제)



# 머신러닝 이란?



# 인공지능, 기계학습, 딥러닝 개념

Artificial Intelligence

## 인공지능

사고나 학습 등 인간이 가진  
지적 능력을 컴퓨터를 통해  
구현하는 기술



Machine Learning

## 머신러닝

컴퓨터가 스스로 학습하여  
인공지능의 성능을  
항상 시키는 기술 방법



Deep Learning

## 딥러닝

인간의 뉴런과 비슷한  
인공신경망 방식으로  
정보를 처리



# 통계 VS 머신러닝

통계 는 분포나 가정을 사용해서  
엄격한 규칙이 적용되는  
설문조사나 실험 계획에 사용됨

머신러닝 은 대용량 데이터의  
분석이나 패턴을 찾는 데 사용됨

# 머신러닝 VS 데이터 마이닝

**머신러닝**은 훈련 데이터를 통해 학습된 알려진 속성을 기반으로 한 예측으로 중점

**데이터마이닝**은 데이터의 미처 몰랐던 속성을 발견하는 것에 집중. 데이터베이스의 지식 발견 부분의 분석 절차에 해당됨

이들은 중복되는 부분이 있다. 데이터 마이닝에서 머신러닝은 필수가 아니지만 머신러닝에서는 데이터마이닝 필수라는 부분이 다르다.



# 머신러닝의 정의

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

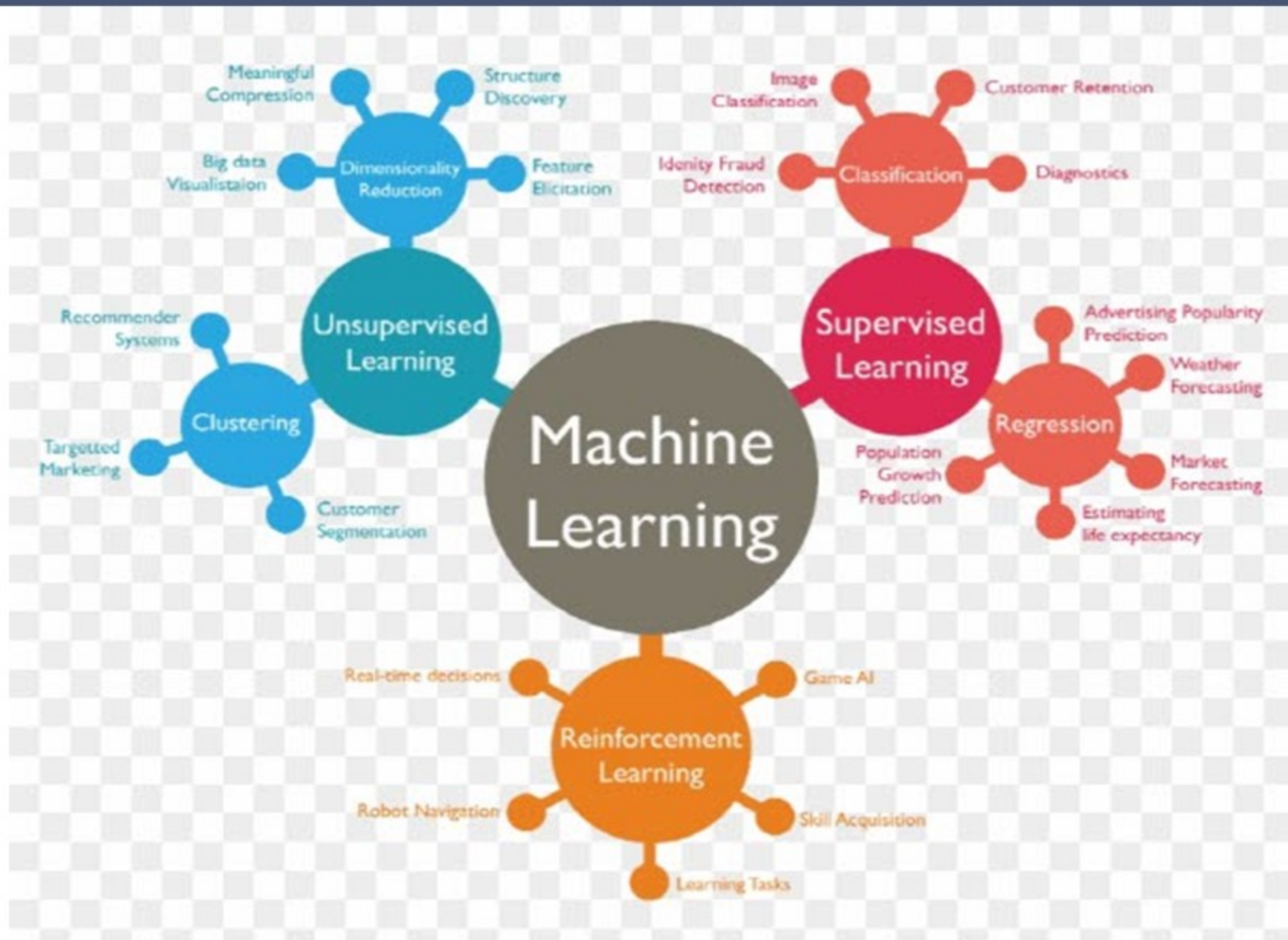
– TOM M. MITCHELL, Carnegie mellon University

테스크(T)에 대해 꾸준한 경험(E)을 통해 T에 대한 성능(P)을 높이는 것을 기계학습이라고 한다.

기계학습에서 가장 중요한 것은 E에 해당하는 데이터이다.

좋은 품질의 데이터를 많이 가지고 있다면 보다 높은 성능을 끌어 낼 수 있다.

# 머신러닝의 분류



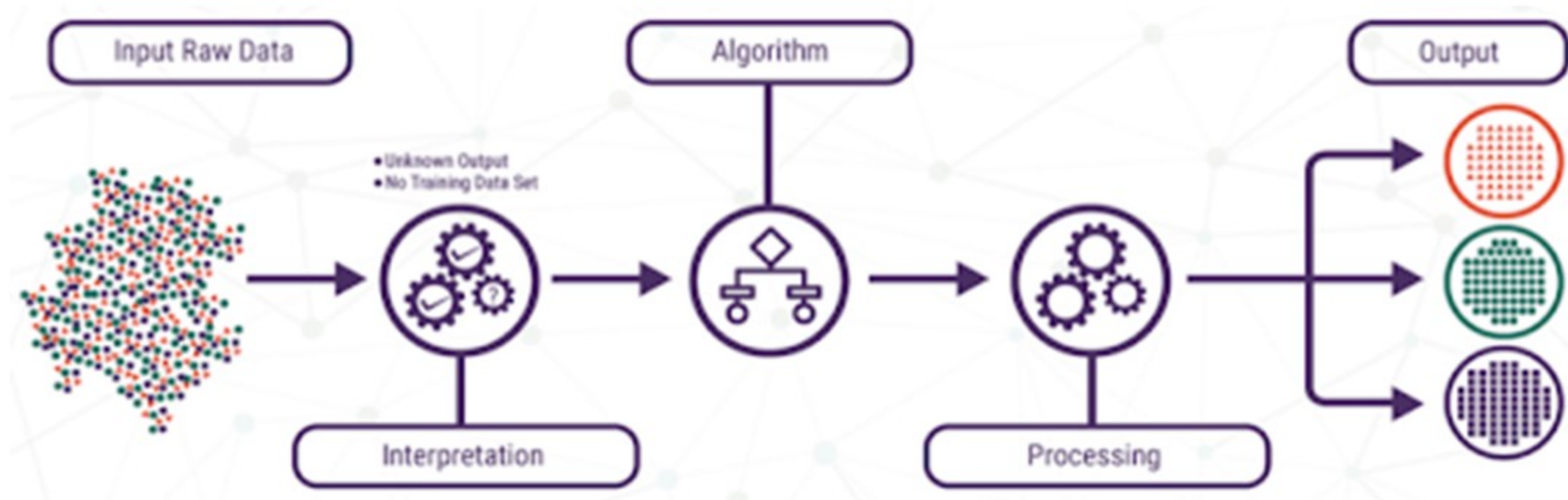


# Supervised Learning (지도학습)

- 정답을 이용하여 알고리즘을 학습하는 방법
- 정답은 label이라고 하고 target이라고도 함
- Supervised Learning 방법을 사용하기 위해서는 데이터를 수집 시 학습하고자 하는 정답을 반드시 같이 수집해야 됨
- Cost가 많이 듦 / labeled data가 많지 않음.



# Unsupervised learning(비지도 학습)



- 비정제 데이터를 입력하여 훈련데이터 없이 데이터의 특징 요약과 군집(Clustering) 수행
- 목표값을 정해주지 않아도 되고 사전 학습이 필요 없으므로 속도가 빠름.

# 지도학습과 비지도 학습 비교

구분	지도 학습	비지도 학습
사용이유	예측 모델 생성	고차원 데이터 분류
성능평가	교차 검증 수행	검증 방법 없음
입력정보	Labeled Data	Raw Data
유형	회귀 : $(x, y)$ 로 $f(x)=y$ 파악 분류 : 그룹별 특징 파악	군집 : 데이터끼리 묶음 패턴 인식 : 여러 그룹 인식
알고리즘	CNN, RNN, SVM, 의사결정 트리 등	K-Means, DBSCAN, 군집(Clustering) 등
장점	사람이 목표 값에 개입하여 정확도가 높음	목표 값을 정해주지 않아도 되므로 속도 빠름
단점	시간이 오래 걸리고 학습 데이터 양이 많음	학습 결과로 분류 기준과 군집 예측 불가
사례	패턴 인식, 질병진단, 주가 예측, 회귀 분석	스팸필터, 차원 축소, 데이터 마이닝, 지식발굴



# Semi-Supervised Learning (준지도학습)

- 한 범주로 목표 값이 표시된 데이터와 표시되지 않은 데이터를 모두 훈련에 사용하는 것
- 레이블이 달려있는 데이터와 레이블이 달려있지 않은 데이터(Unlabeled data)를 동시에 사용해서 더 좋은 모델을 만드는 것.

1. 자가 학습 (self-train)	2. 멀티뷰 학습 (Multi-view learning)	3. 셀프 인셈블링 (self-ensembling)	4. 트라이 학습 (tri-training)
라벨 된 데이터 셋을 추가하여 라벨 되지 않은 데이터에 대한 모델의 자체 예측 이용	데이터의 여러 측면에서 여러 모델을 학습	여러 다양한 설정의 단일 모델을 사용	기본적으로 3개의 모델을 만들고 2개 모델이 데이터 라벨에 합의할 때마다 이 라벨이 세 번째 모델에 추가 됨

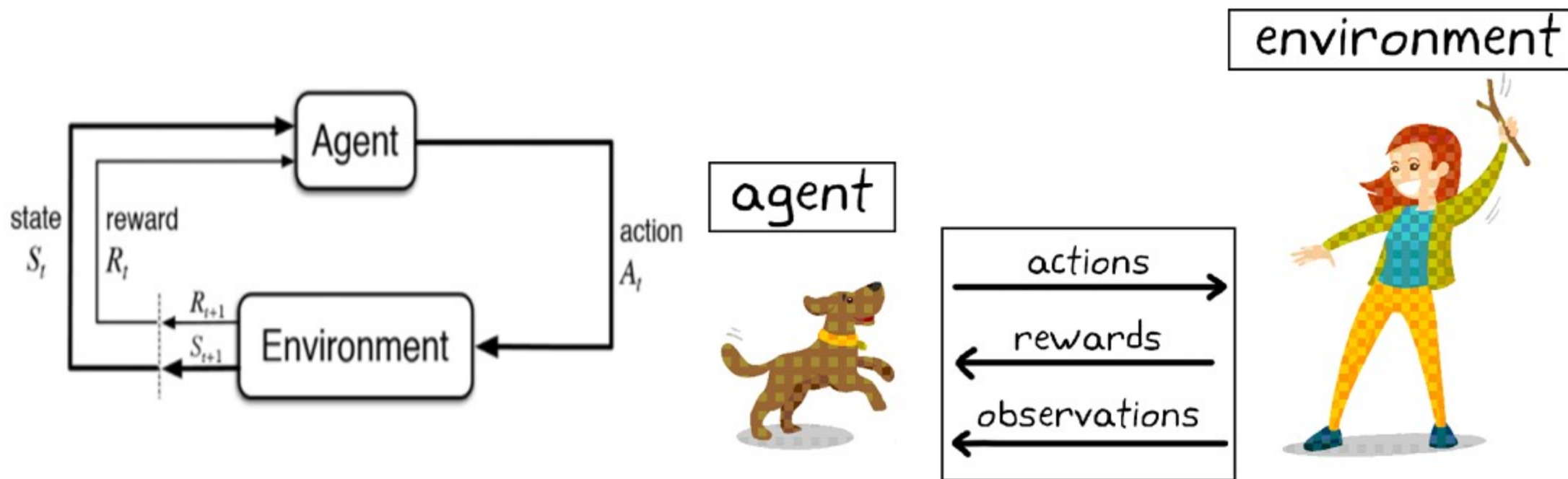
# Self-supervised learning (자기지도학습)

- Unsupervised Learning 연구 분야 중에 하나인 **Self-Supervised Learning**
- 삼성 AI 포럼에서 AI 분야의 전설적인 존재인 안 르쿤 교수는 Self-Supervised Learning이 미래의 먹거리이자 Real AI로 가기 위해선 이 분야가 발전해야 한다고 역설
- Supervision을 스스로 준다는 것을 의미하며, Unsupervised Learning 방식을 따라서 데이터셋은 오로지 레이블이 존재하지 않는(Unlabeled) 데이터만 사용
- 데이터 간의 특성을 Representation을 이용하는 Self supervised learning

Paper : Revisiting Self-Supervised Visual Representation Learning

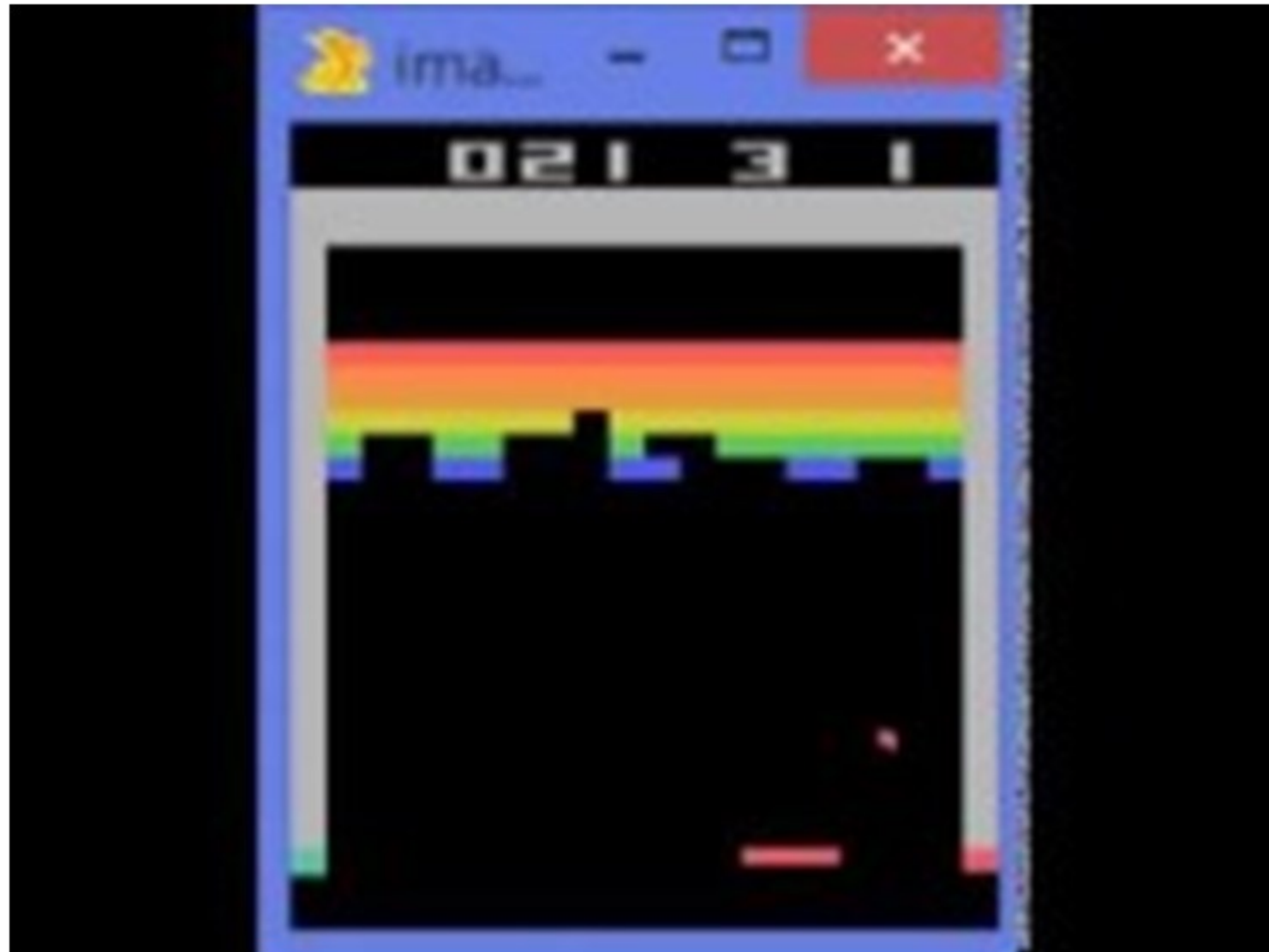
# 강화학습 (Reinforcement learning)

- 어떤 환경을 탐색하는 에이전트가 현재의 상태를 인식하여 어떤 행동을 취한다.  
그러면 그 에이전트는 환경으로부터 포상을 얻게 된다. 포상은 양수와 음수 둘 다 가능하다.  
강화 학습의 알고리즘은 그 에이전트가 앞으로 누적될 포상을 최대화하는 일련의 행동으로  
정의되는 정책을 찾는 방법

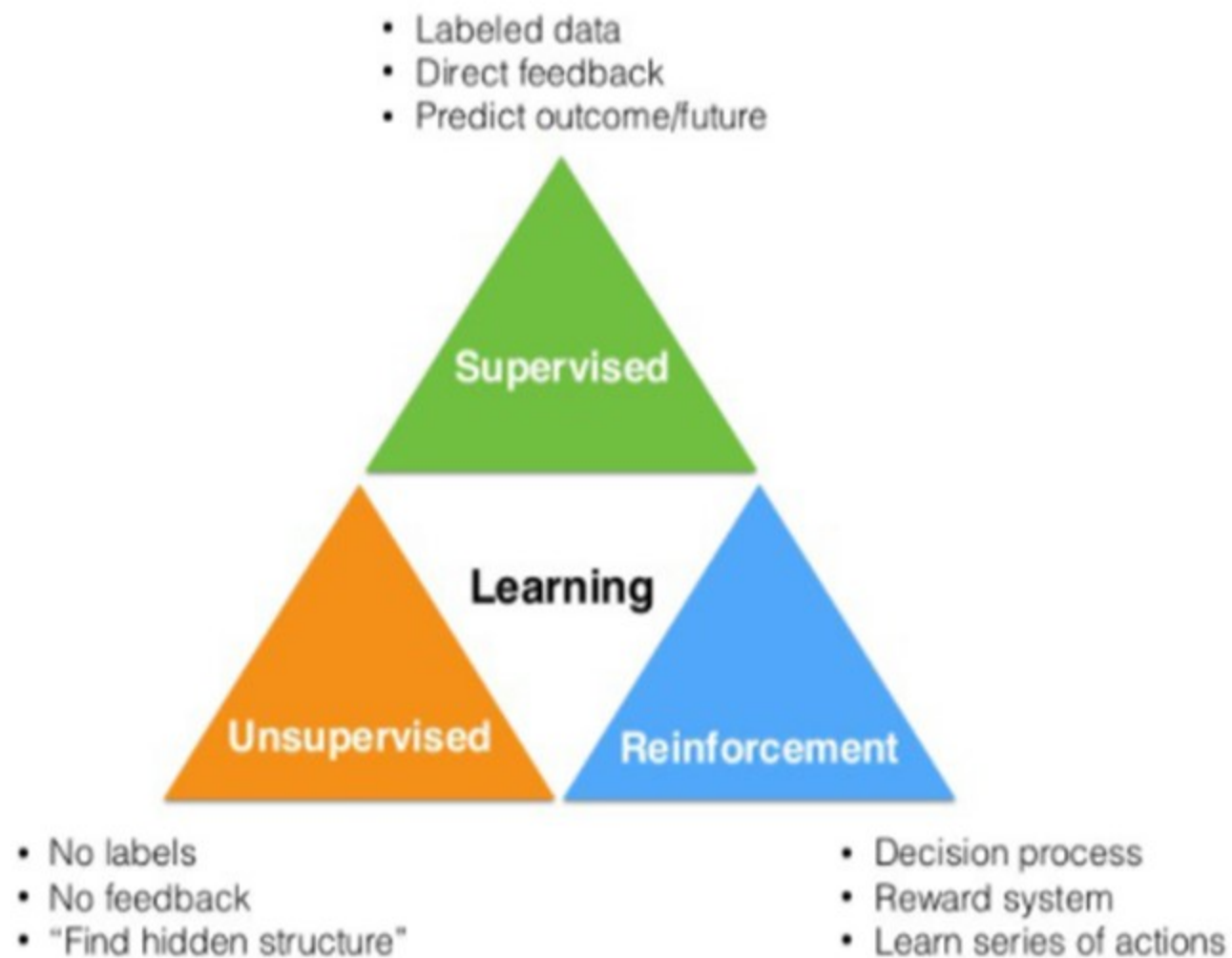




# Reinforcement learning



# Machine learning



# Few-shot learning

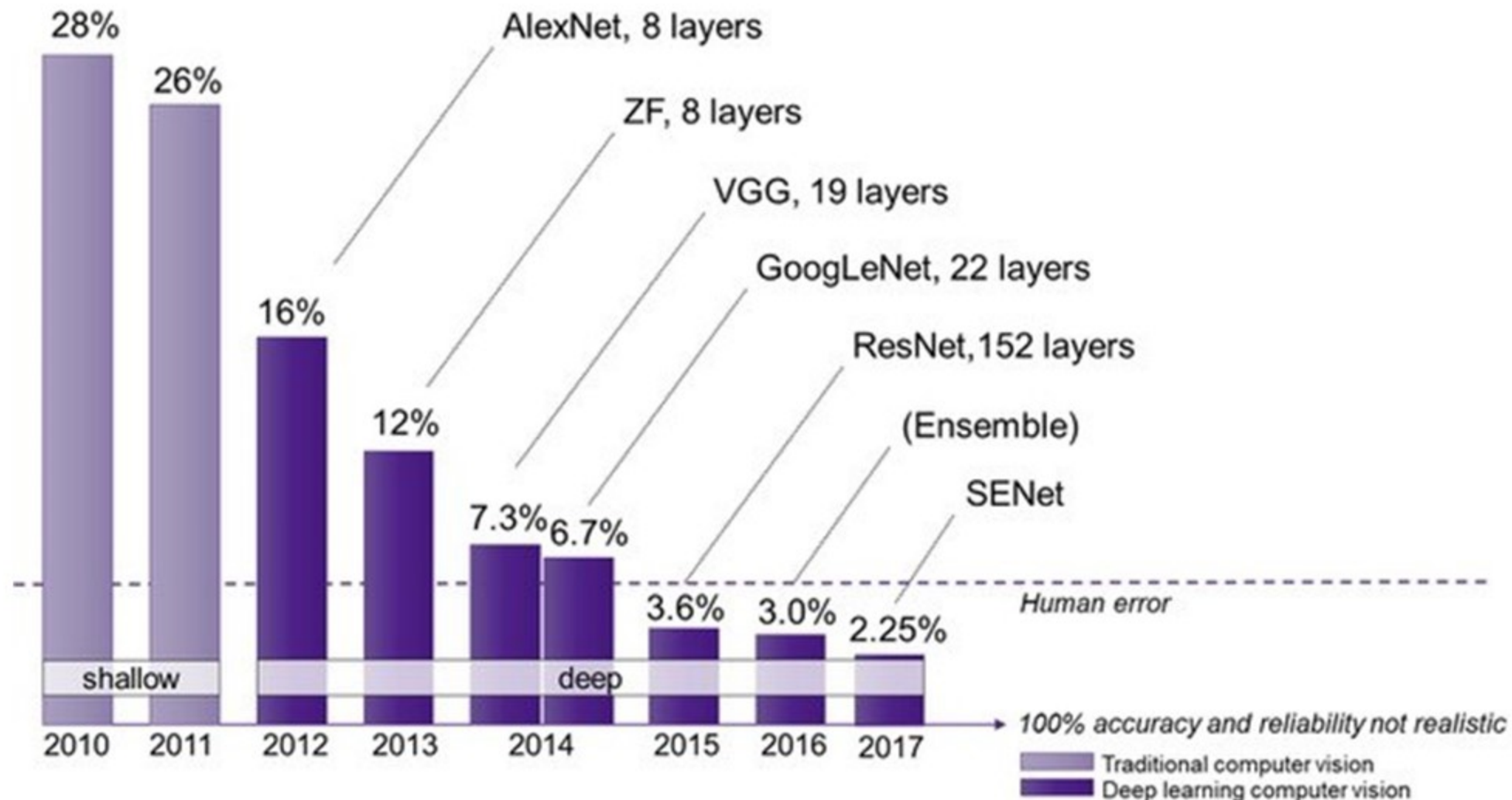
- 적은 수의 학습은 많은 양의 데이터를 사용하는 일반적인 관행과 달리 매우 적은 양의 훈련 데이터로 학습 모델에 공급하는 관행
- 이 기법은 주로 컴퓨터 비전 분야에서 활용되며, 개체 분류 모델을 사용하면 몇 가지 훈련 샘플이 없어도 적절한 결과를 얻을 수 있습니다.



[ 그림 1 ] 2-way 5-shot 이미지 분류 문제의 예시 © 카카오브레인



# High performance of Deep learning



# High performance of Deep learning



# High performance of Deep learning





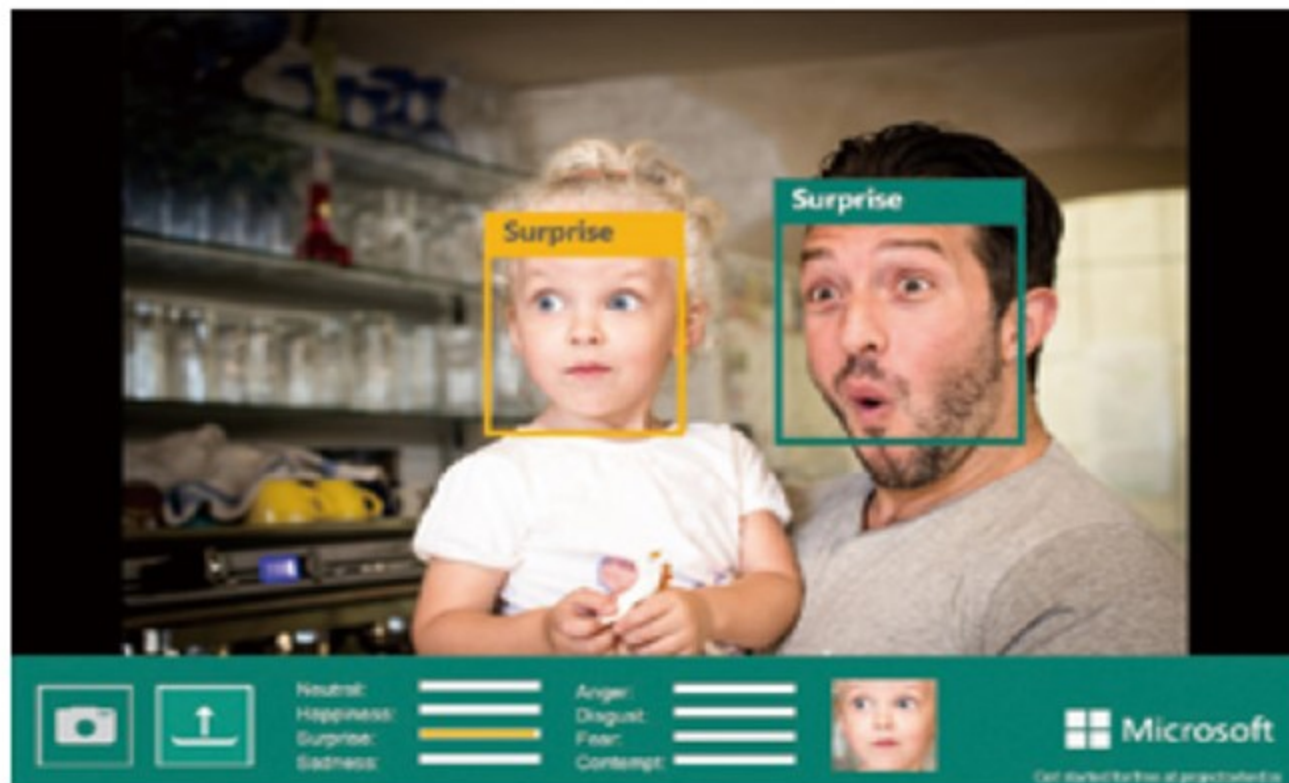
# High performance of Deep learning



사진에 담긴 사람의 8가지 감정을  
수치로 나타내는 서비스

- 분노(anger)
- 경멸(Contempt)
- 불쾌(Disgust)
- 공포(Fear)
- 행복(Happiness)
- 중립/무관심(Neutral)
- 슬픔(Sadness)
- 놀라움(Surprise) 등

8가지 감정들을 수치로 표시



[그림 8] 마이크로소프트(MS)가 만든 기계학습 프로젝트 옥스퍼드(Oxford). MS는 사진에 담긴 사람의 여덟 가지 감정을 수치로 나타내는 서비스를 공개했다.

# Thank you.

머신러닝 소개 / 류영표 강사  
youngpyoryu@dongguk.edu

Copyright © "Youngpyo Ryu" All Rights Reserved.  
This document was created for the exclusive use of "Youngpyo Ryu".  
It must not be passed on to third parties except with the explicit prior consent of "Youngpyo Ryu".