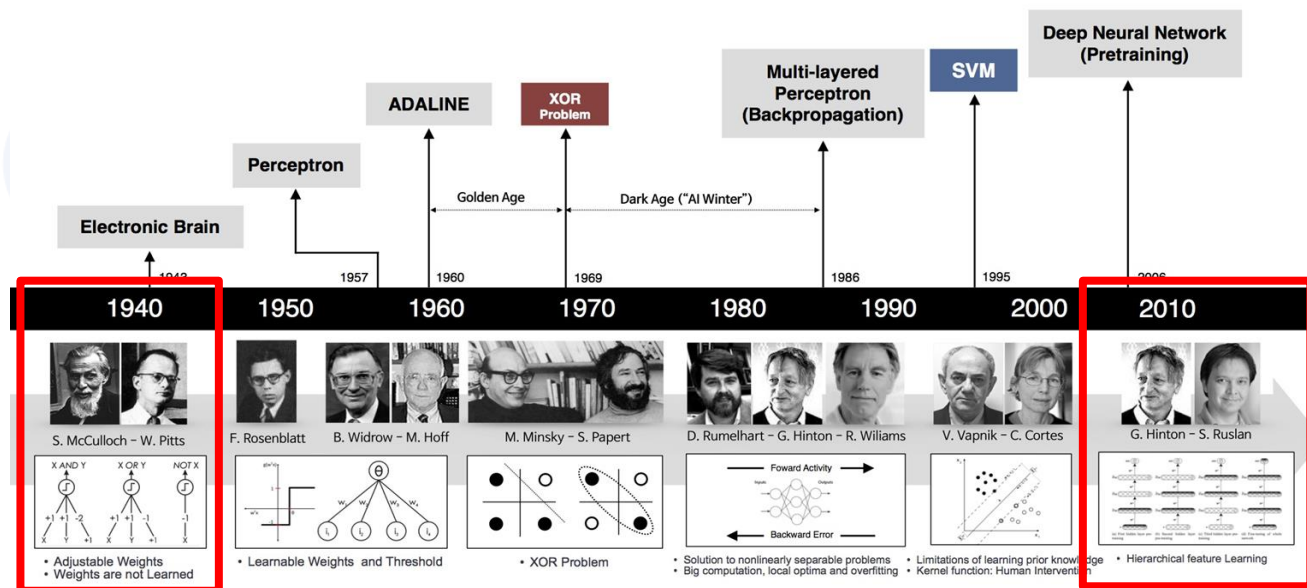


CONTENTS / 001

AI 개요

- 학습, 문제 해결, 패턴 인식 등 **인지 문제**를 해결하는 컴퓨터 공학 분야
- 1943년 처음으로 개념이 태동
- 최근 **빅데이터**와 **컴퓨팅 성능** 향상으로 인해 폭발적인 성장 중



- **Narrow AI**

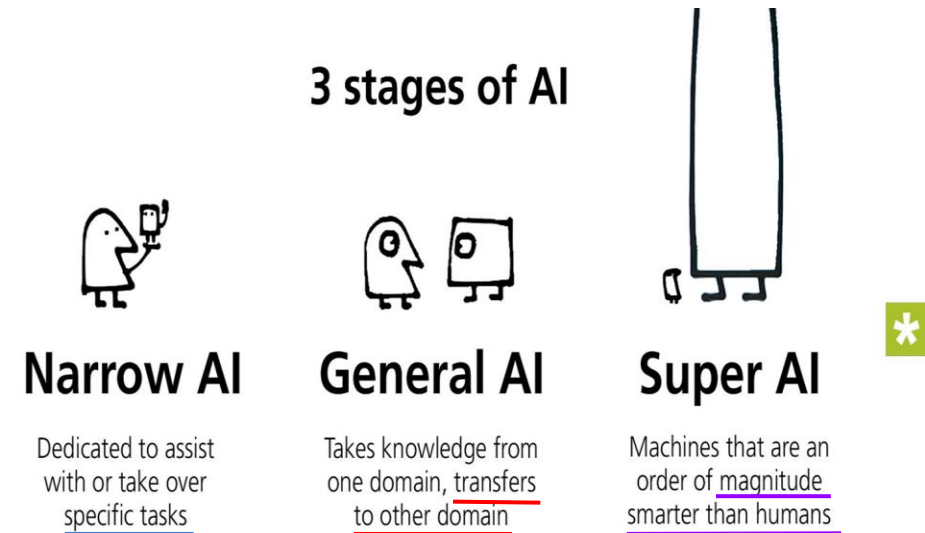
- 좁은 영역에서 **특정한 task**만 가능하지만 정확도나 속도는 인간을 넘어섬

- **General AI**

- 하나의 task에서 습득한 지식을 **여러 task**에 적용 가능

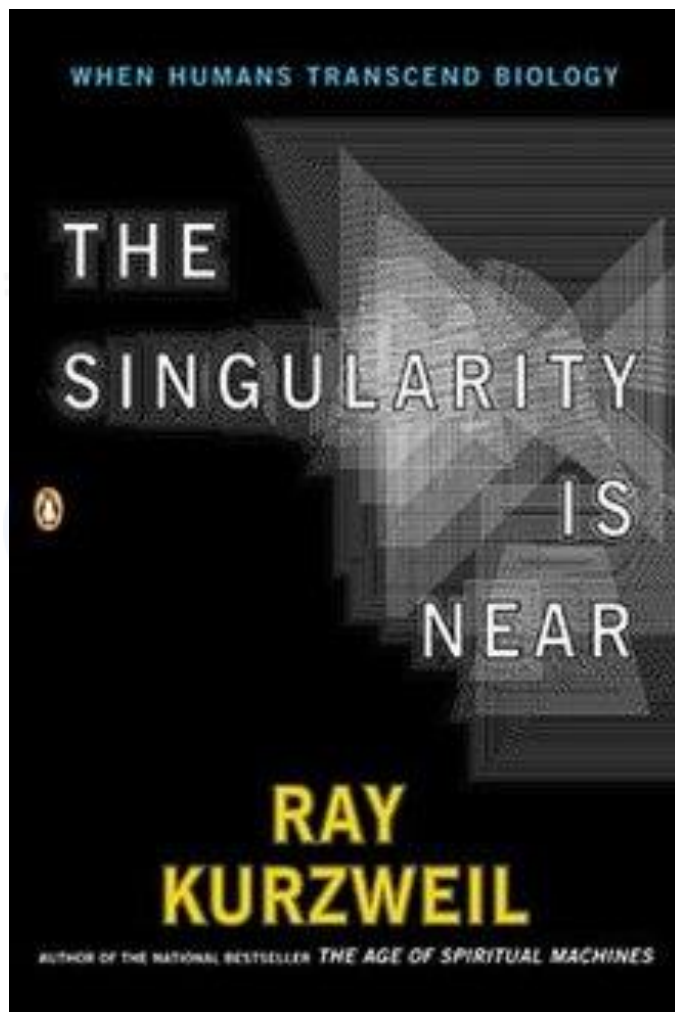
- **Super AI**

- **인간을 뛰어넘은** 성능을 보여주는 인공지능



- 레이커즈와일(The Singularity Is Near 의 저자)
 - 구글 기술 이사/ OCR, TTS 등 많은 기술 발명
 - "인공지능과 인간의 두뇌가 자연스럽게 하나가 될 것이라고 본다"
 - 하루에 영양제 250알, 다양한 처방 등 온갖 수단을 다 동원해서 불로에 집착



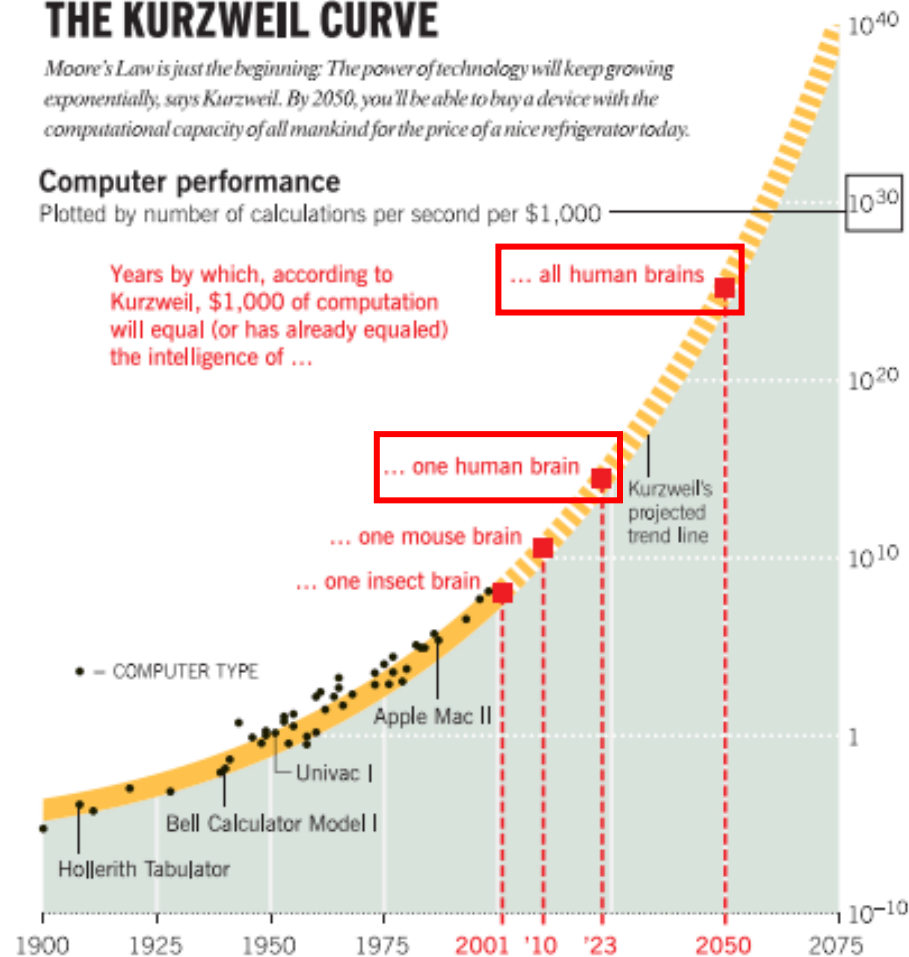


THE KURZWEIL CURVE

Moore's Law is just the beginning: The power of technology will keep growing exponentially, says Kurzweil. By 2050, you'll be able to buy a device with the computational capacity of all mankind for the price of a nice refrigerator today.

Computer performance


Plotted by number of calculations per second per \$1,000



SOURCE: DATA FROM RAY KURZWEIL

- 전기가 산업을 변화 시켰 듯이 인공지능도 같은 역할을 할 것이다

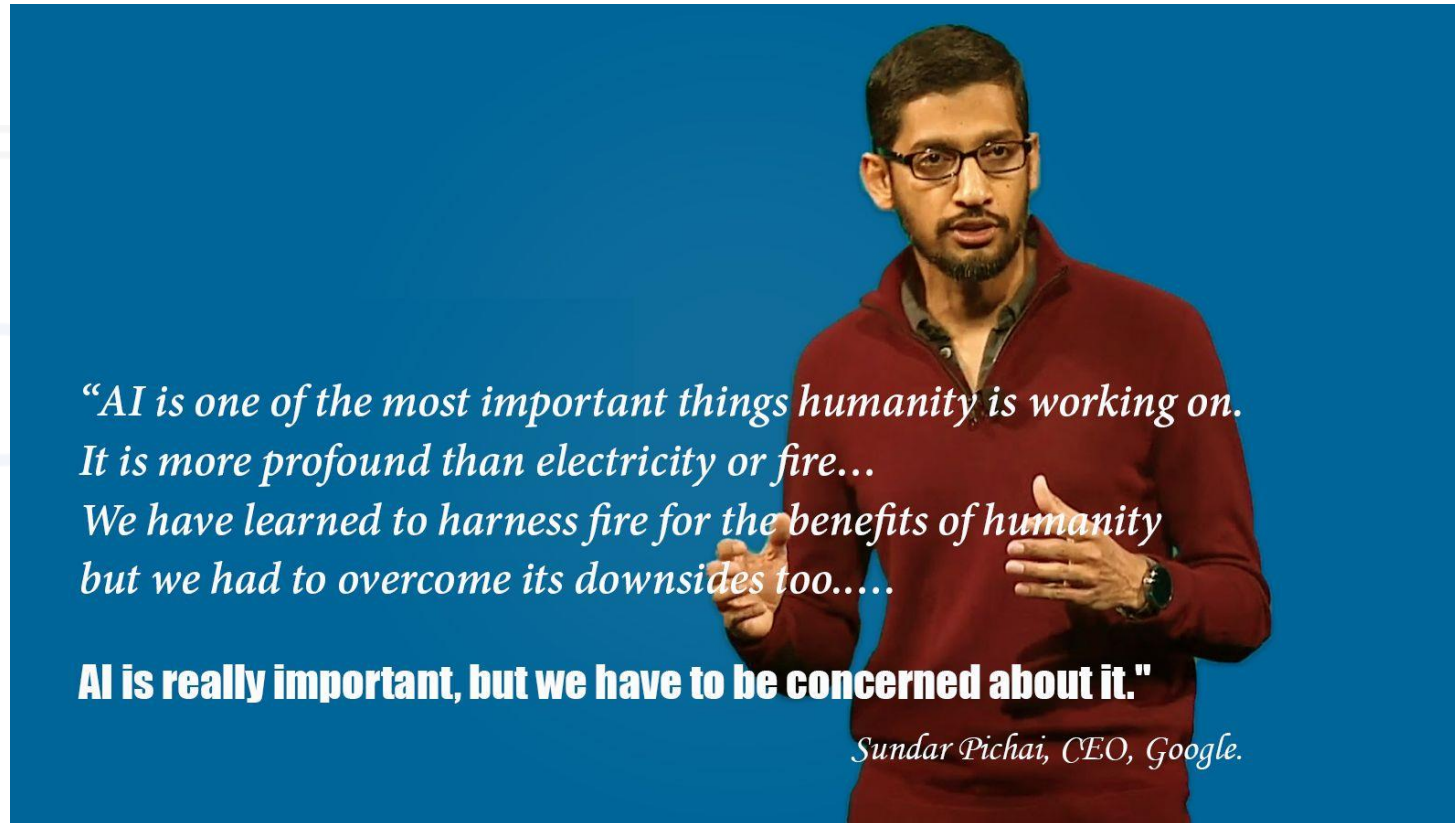
'AI IS THE NEW ELECTRICITY'



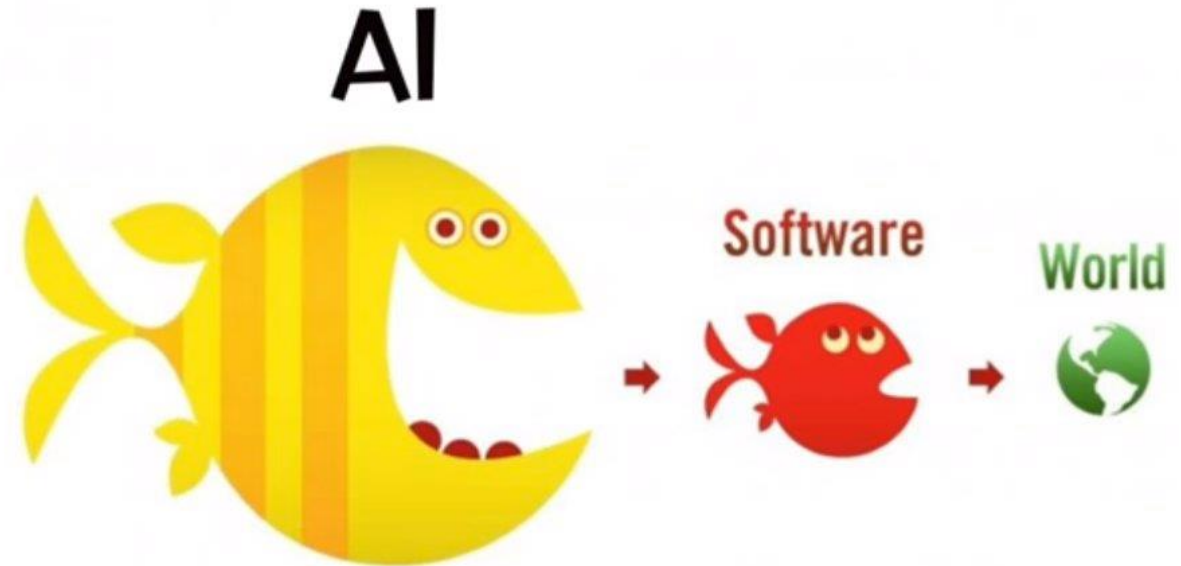
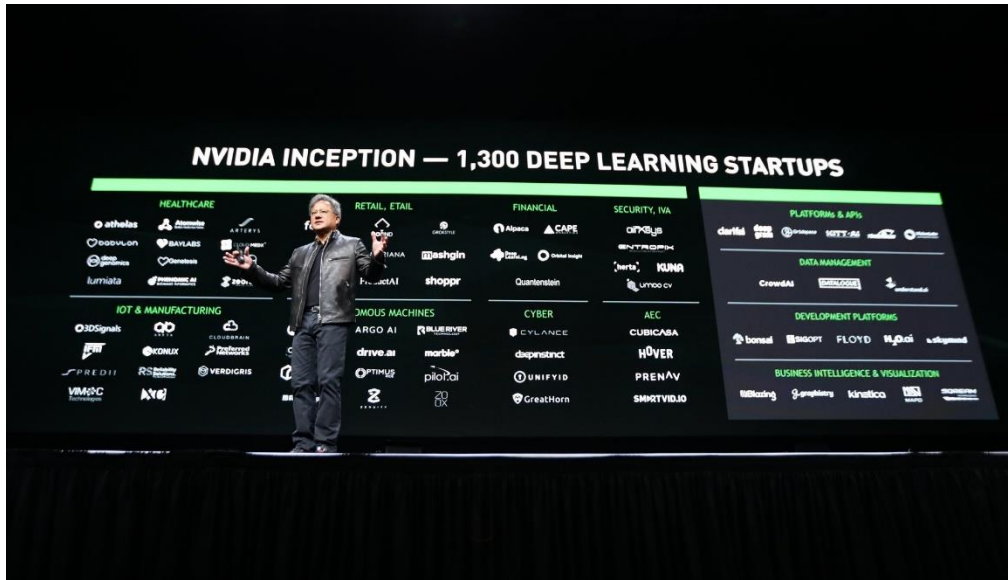
"Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI will transform in the next several years."

Andrew Ng
Former chief scientist at Baidu, Co-founder at Coursera

- AI는 불이나 전기의 발견보다도 더 중요하다



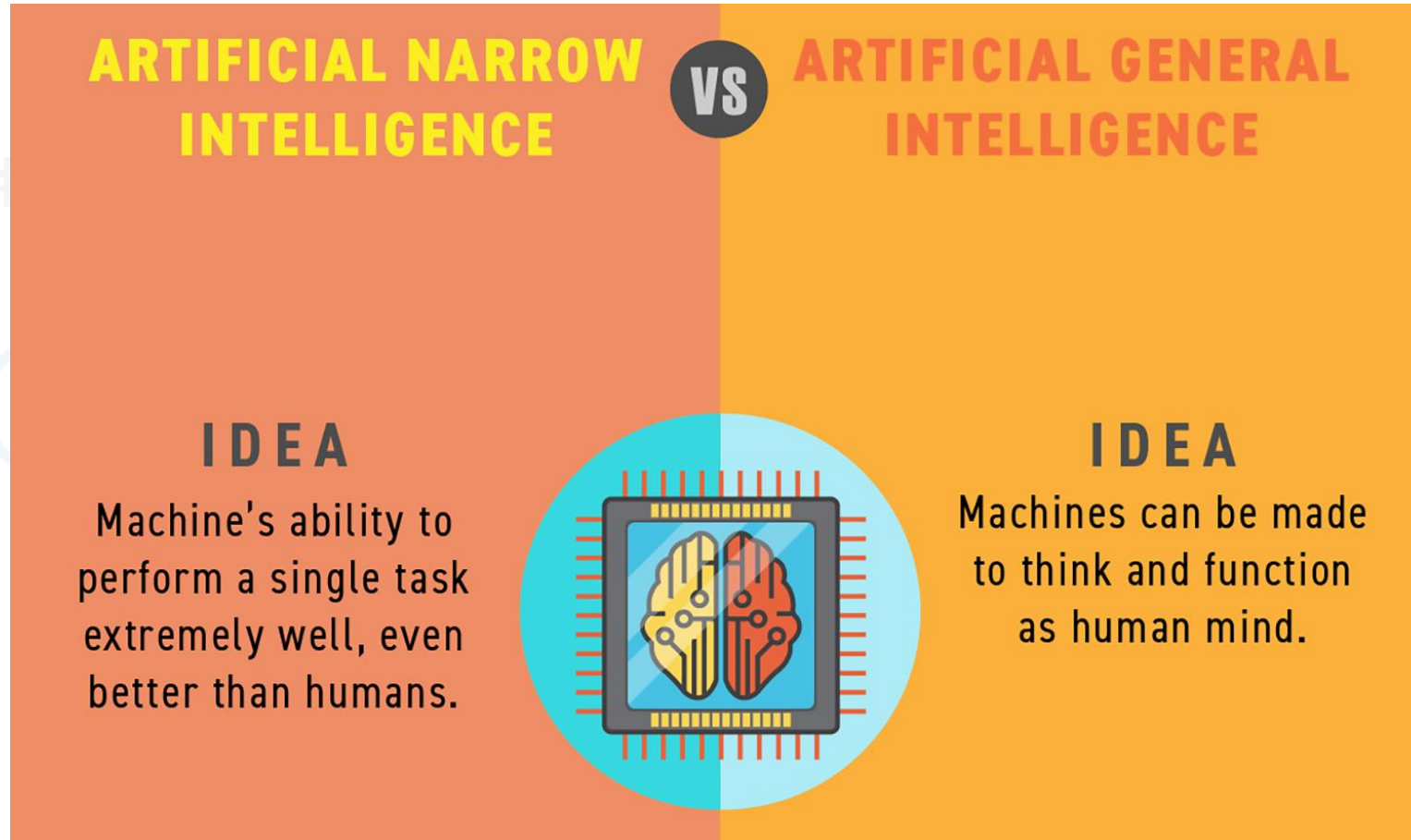
- SW is eating the world, but AI is going to eat SW



- 인류의 지성을 모두 합친 수준의 슈퍼 AI가 도래할 것이다.



- ANI를 넘어선 AGI의 시대로의 이동 중



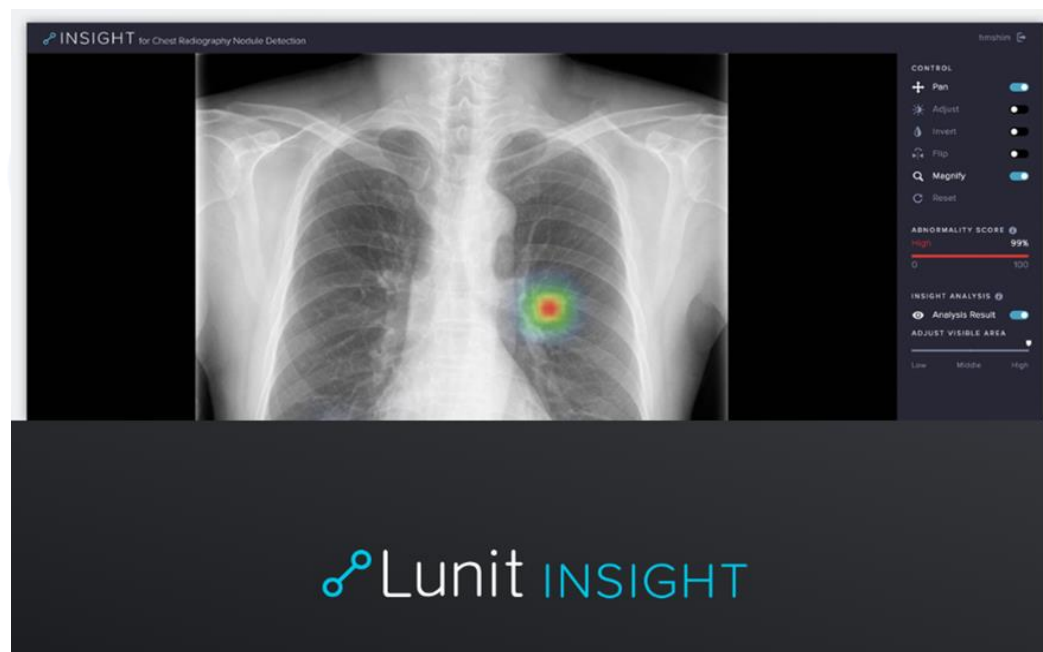








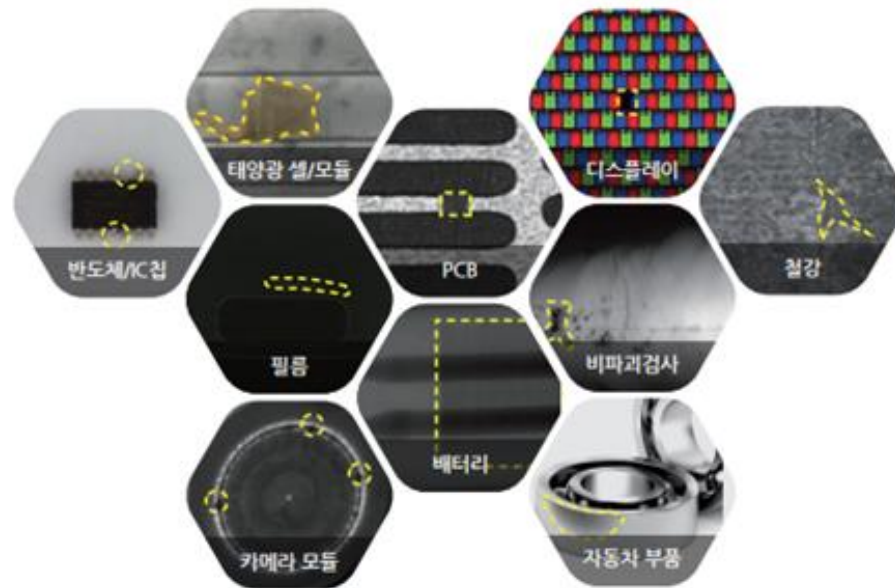
- 딥러닝 기반의 의료 영상 기반 진단 스타트업
- 폐 결절 예측 정확도 97%, 판독률 20% ↑
- 기흉 99%, 결핵 99%, 폐렴 98% 예측



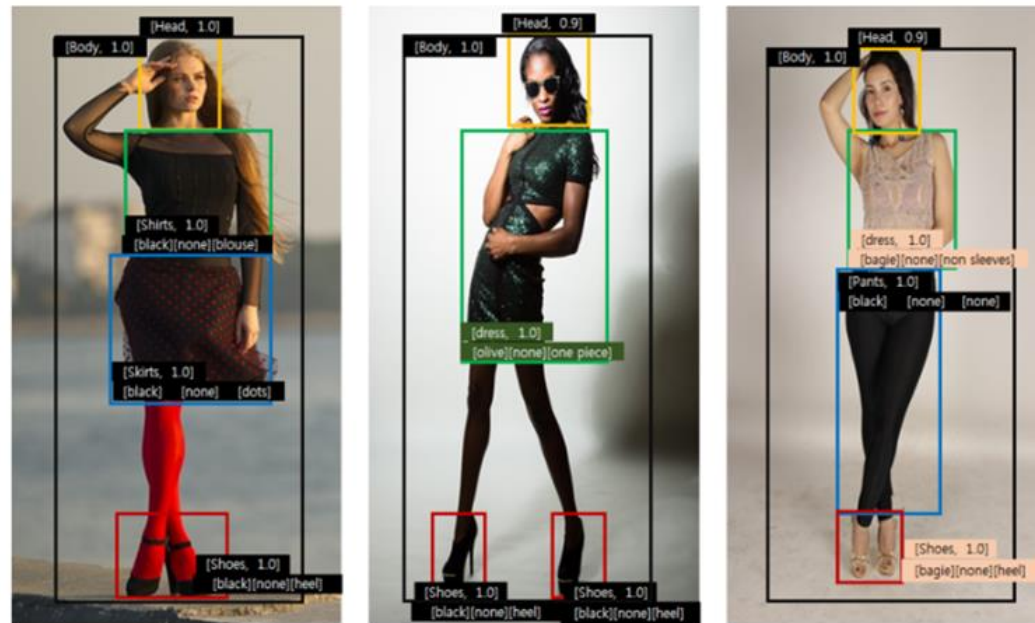
- 딥러닝 엔진 기반의 의료 영상 분석 솔루션 개발
- 수골 x-ray 이미지 데이터 사용하여 판독 정확도 향상, 판독 속도 개선
- 응급실 환자 데이터를 VunoNet 에 학습시켜 응급환자 중증도 예측 모델 개발



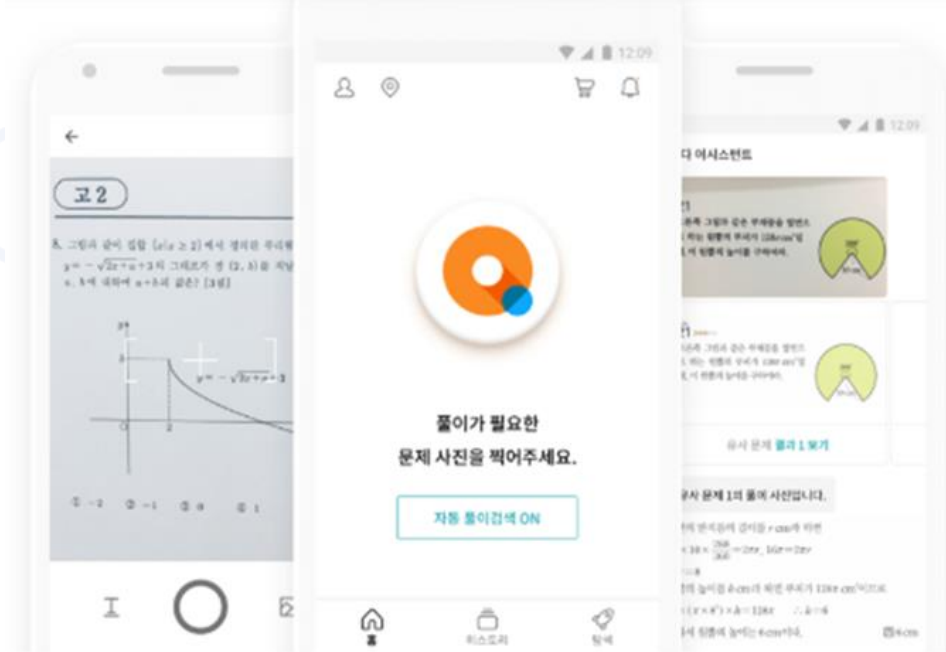
- 딥러닝 기반의 머신 비전 솔루션 'SuaKIT' 개발
- 딥러닝을 통한 이미지 인식 기반 불량 검출 서비스
 - PCB, 디스플레이, 필름, 반도체, 태양광, etc



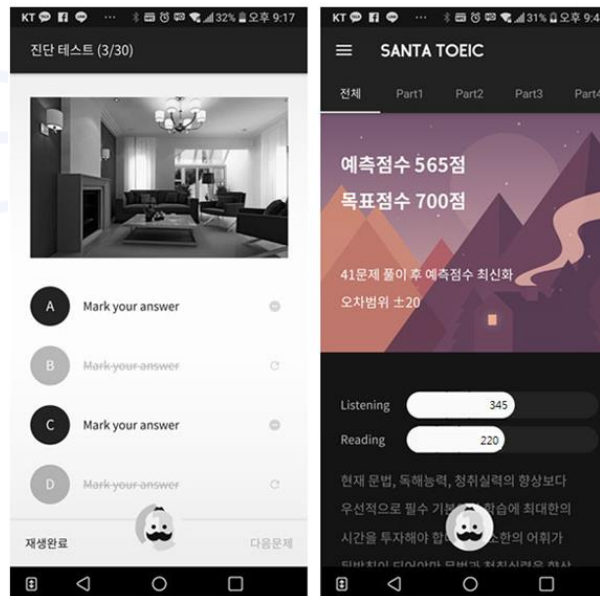
- 패션 이미지 전문 자동 Tagger 서비스 개발
- 패션 이미지를 딥러닝 기반으로 분석하여 tagging 해주는 솔루션
- 의류 카테고리화 80개, 의류 속성 900개 정의



- 딥러닝 기반 수학 문제 풀이 검색 플랫폼 '관다(QandA)' 개발
- 모르는 문제는 사진을 찍어 올리면 5초 내 해설을 줌
- 한국어와 수식 등 다양한 텍스트 인식을 위해 CNN 기반 자체 엔진 개발



- 머신러닝 기반 토익 학습 서비스 'SANTA TOEIC' 개발
- 객관식 시험 영역에 특화된 머신러닝 기술 연구, 개발
- 어떤 문제를 맞출지, 어떤 오답을 고르는 지를 90% 이상 예측

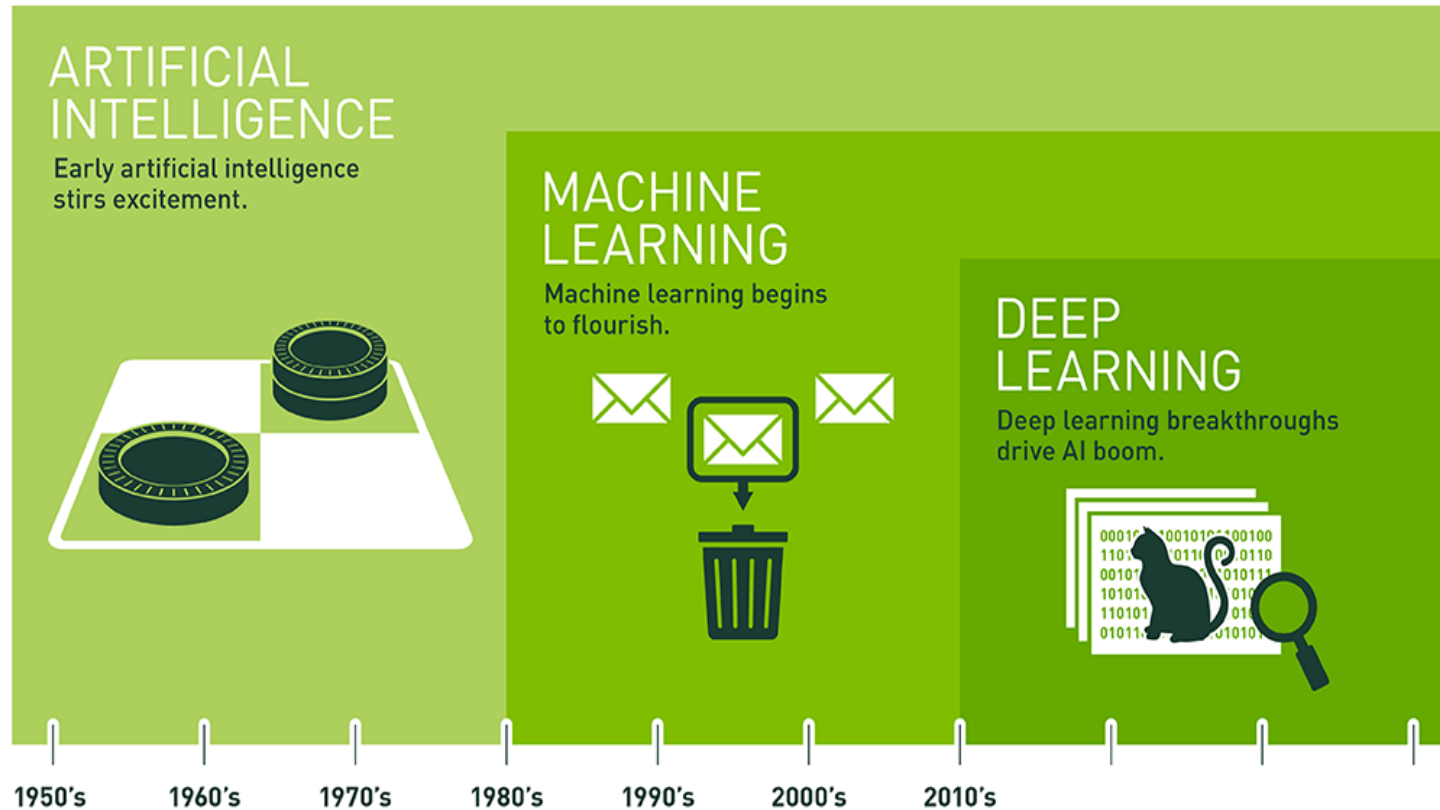


- 음성, 영상, 대화처리 기술을 이용한 대화형 인공지능 서비스 개발
- 자연어 분석을 통해 비대면 상담 챗봇 서비스 'Playchat' 개발
- 음성합성(TTS), GAN 기술을 이용하여 AI 뉴스앵커 서비스 개발



CONTENTS / 002

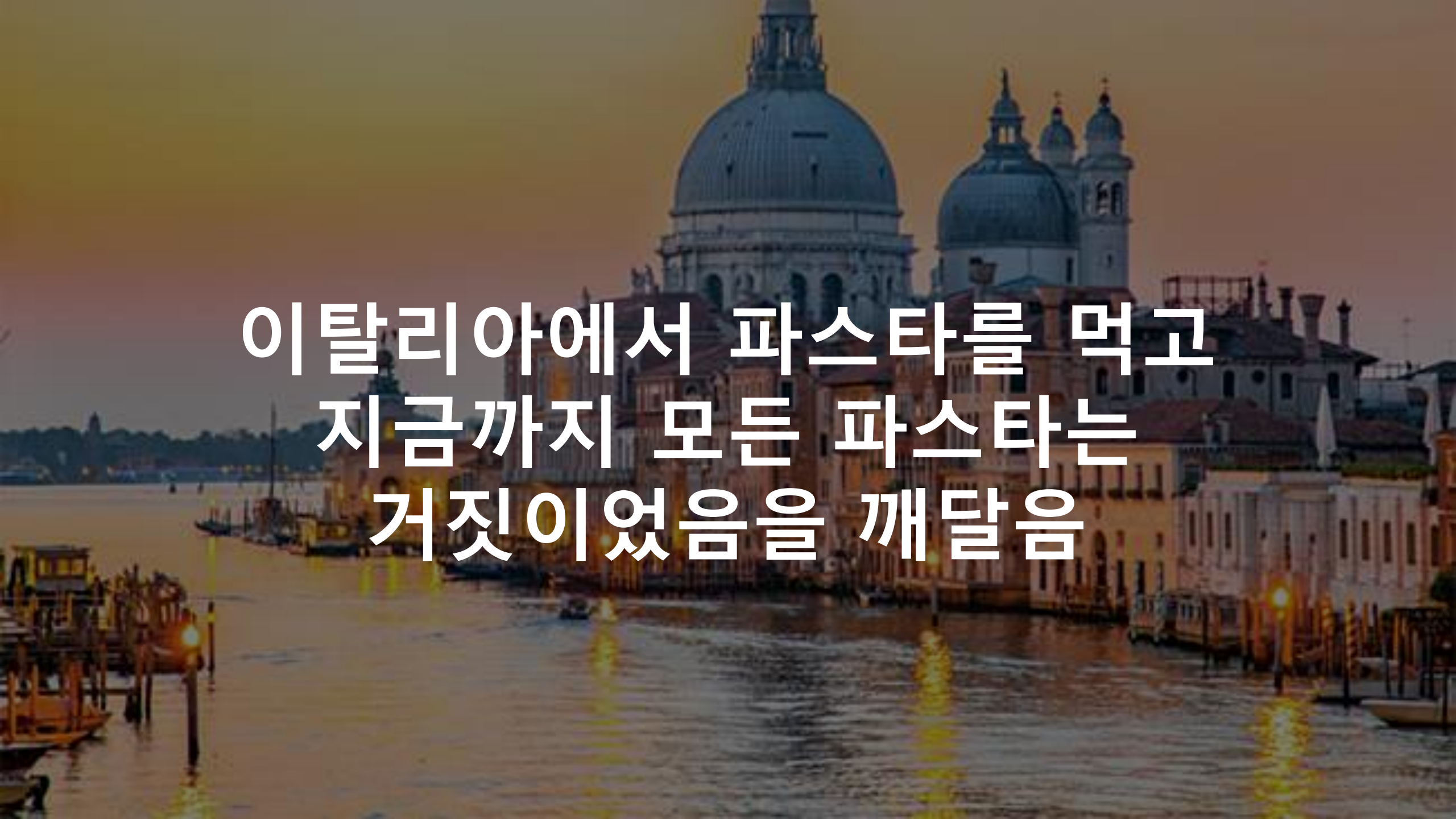
머신러닝 개요




Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

- 인공지능을 구현하는 구체적인 접근 방식 중 하나
- 기존의 방식은 코드를 통해서 수행 방법을 구체적으로 명시(심볼릭 AI)
- 머신러닝은 코드가 아닌 대량의 데이터로 부터 학습을 하고, 학습 된 모델을 통해 판단 또는 예측을 함

머신 러닝? 기계가 배운다 어떻게? 수많은 데이터로 부터



이탈리아에서 파스타를 먹고
지금까지 모든 파스타는
거짓이었음을 깨달음

A close-up photograph of a fork lifting a portion of spaghetti from a white plate. The spaghetti is coated in a light-colored sauce and topped with small pieces of meat. The background is softly blurred, showing a wooden surface and some kitchen items. The text is overlaid in the center of the image.

집에 돌아와
같은 맛을 내보려고 시도
레시피는 없고 맛의 기억만 존재

어떻게 비슷한 파스타를 만들 수 있을까?

#PROFESSIONAL IT EDU-PLATFORM

기억하는 맛을 재현할 때까지
끝 없는 반복과 시행 착오

요리하고 먹어보고 오차 측정하고 (많이 짜네)

개선해서 요리하고 먹어보고 오차 측정하고 (조금 짜네)

개선해서 요리하고 먹어보고 오차 측정하고 (느끼하네)

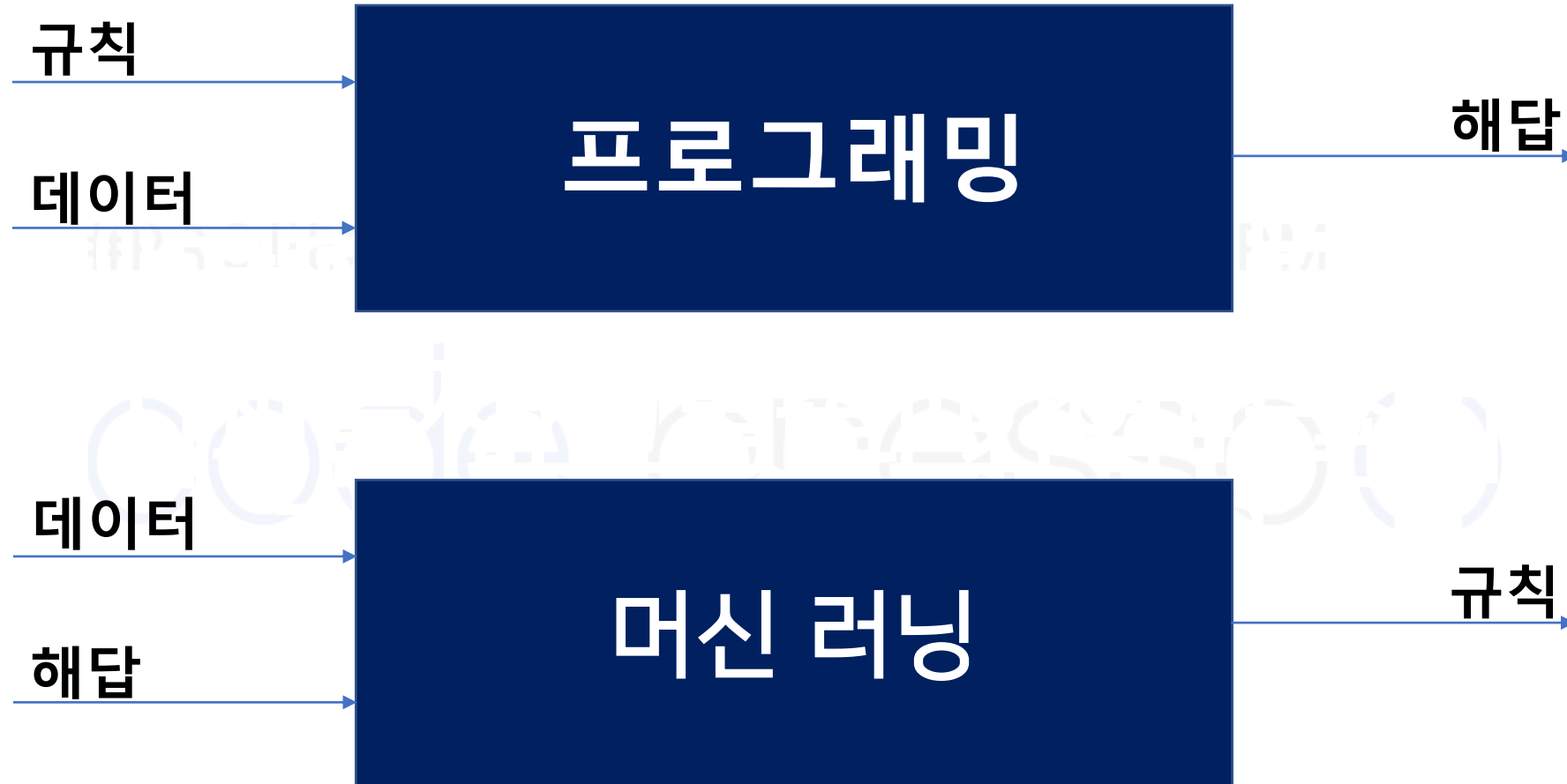
개선해서 요리하고 먹어보고 오차 측정하고 (감칠맛이 없네)

개선해서 요리하고 먹어보고 오차 측정하고

X 100



수백번의 시도 끝에
비슷한 맛을 만드는데 성공





- 규칙 기반
 - 대출 이란 단어가 들어가면
 - 링크가 삽입 되면
 - 광고 라는 단어가 들어가면
 - 욕설이 들어가면
 - 처음 받는 이메일 이면
 - 처음 받는 이메일 도메인 이면

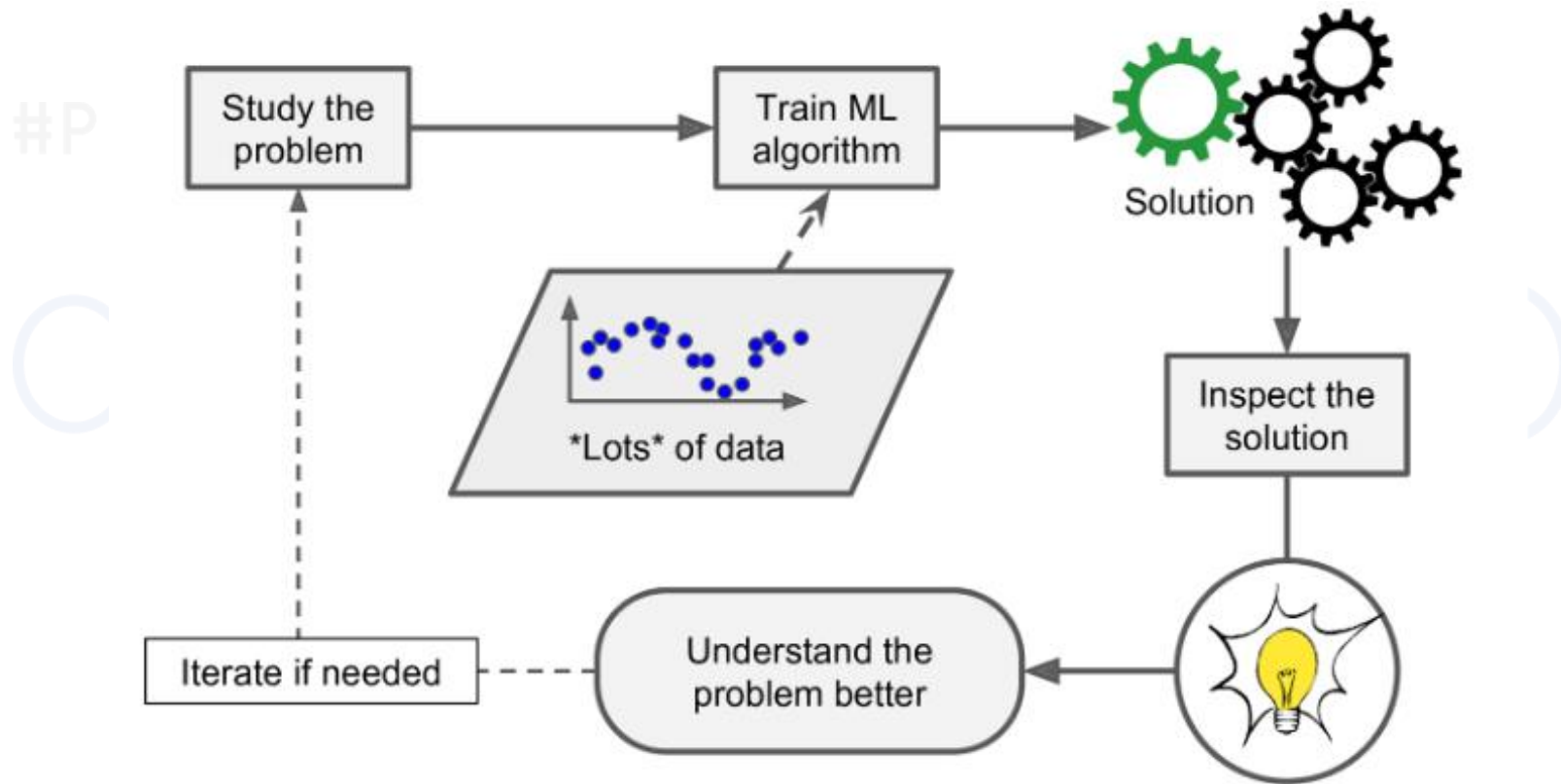
- 머신러닝

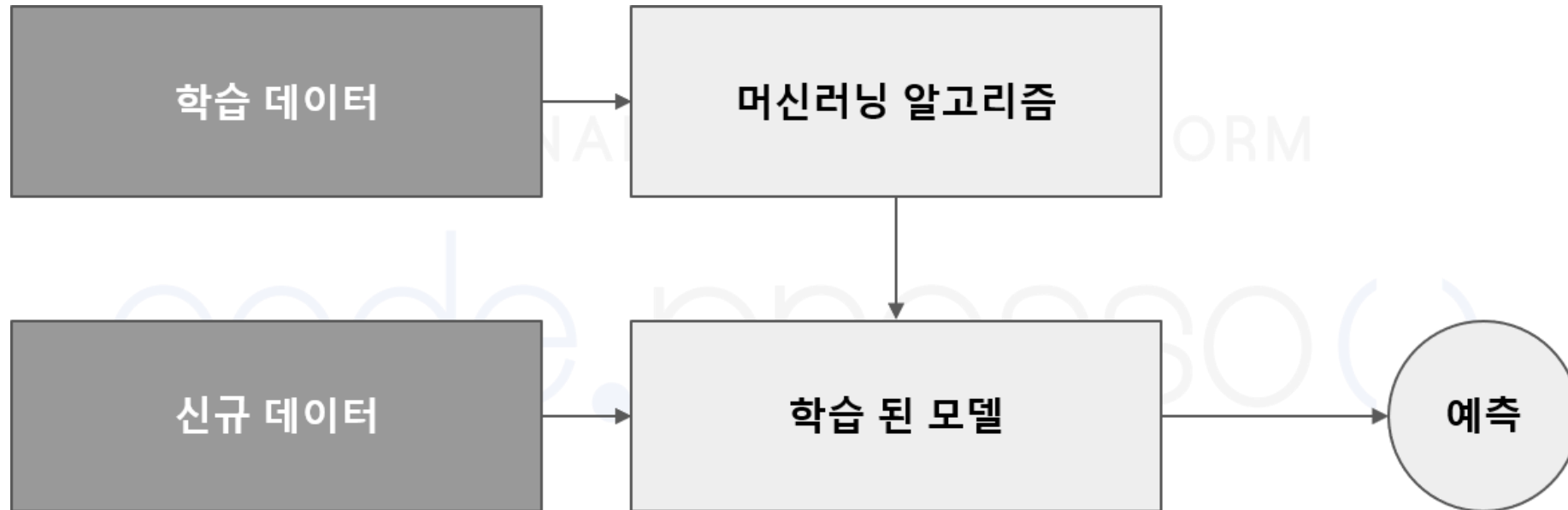
- 기계에게 수많은 데이터와 해답을 넣음
- 이 데이터는 스팸이야
- 이 데이터는 스팸 아니야
- 이 데이터는 스팸이야
- 이 데이터는 스팸이야
- 이 데이터는 스팸 아니야
- ...
- 기계가 무엇이 스팸이고 무엇이 스팸 아닌지 학습하게 됨

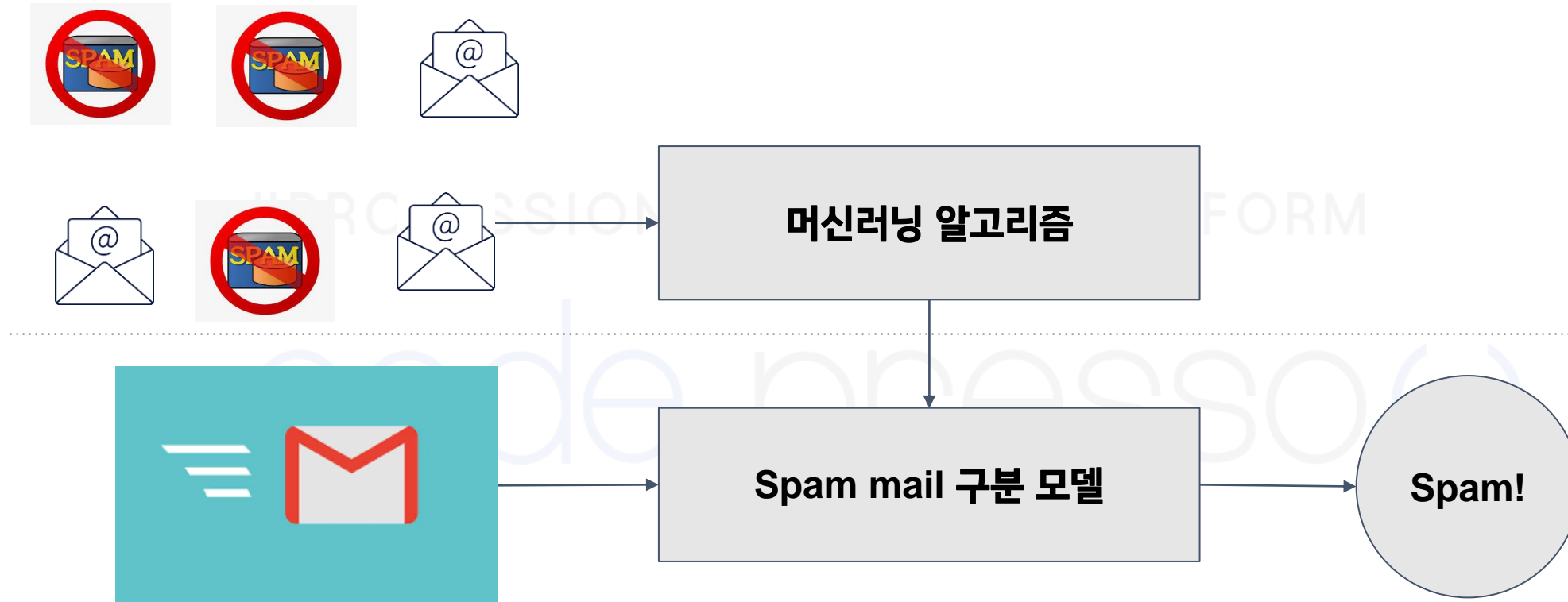
- 머신러닝 기법 사용의 장점

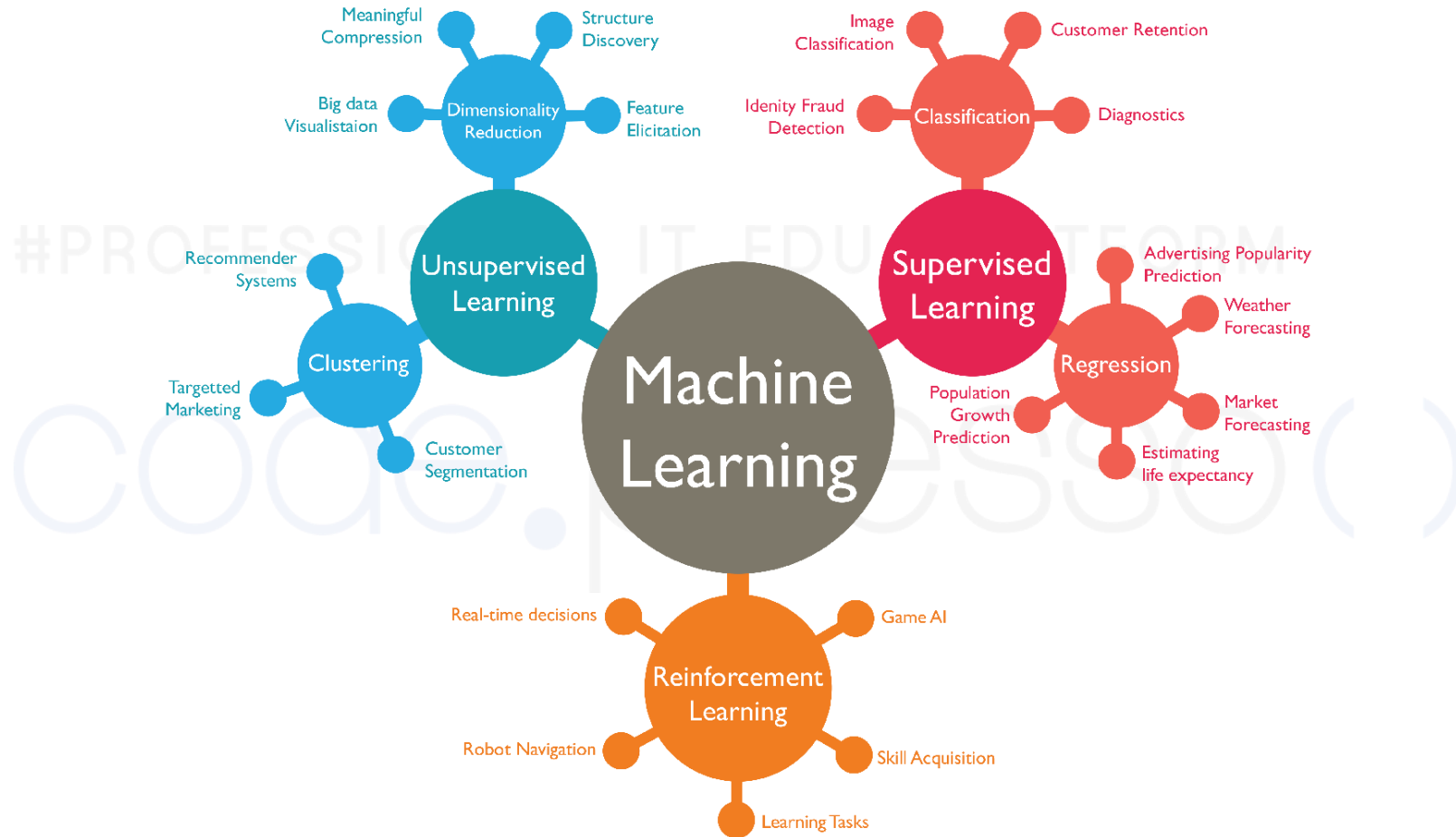
- 프로그램이 간단해 유지 보수가 쉽고, 대부분 정확도 더 높음
- 스팸 메일의 패턴 변경 시에도 별도의 작업을 하지 않아도 됨
- 전통적인 방식으로 해결하기에 복잡한 문제를 쉽게 해결 가능
- 전통적인 방식으로 해결 할 수 있는 알고리즘이 없는 경우에도 적용 가능
- 데이터 분석을 통해 전에 보이지 않던 패턴 발견 가능(데이터 마이닝)

- 데이터 마이닝 : 데이터 분석을 통해 전에 보이지 않던 패턴 발견



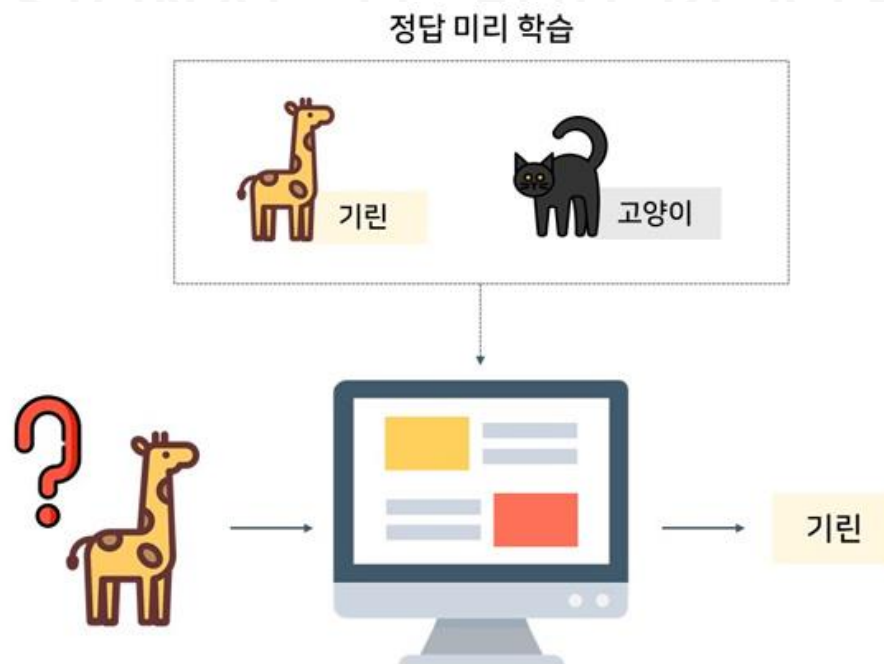






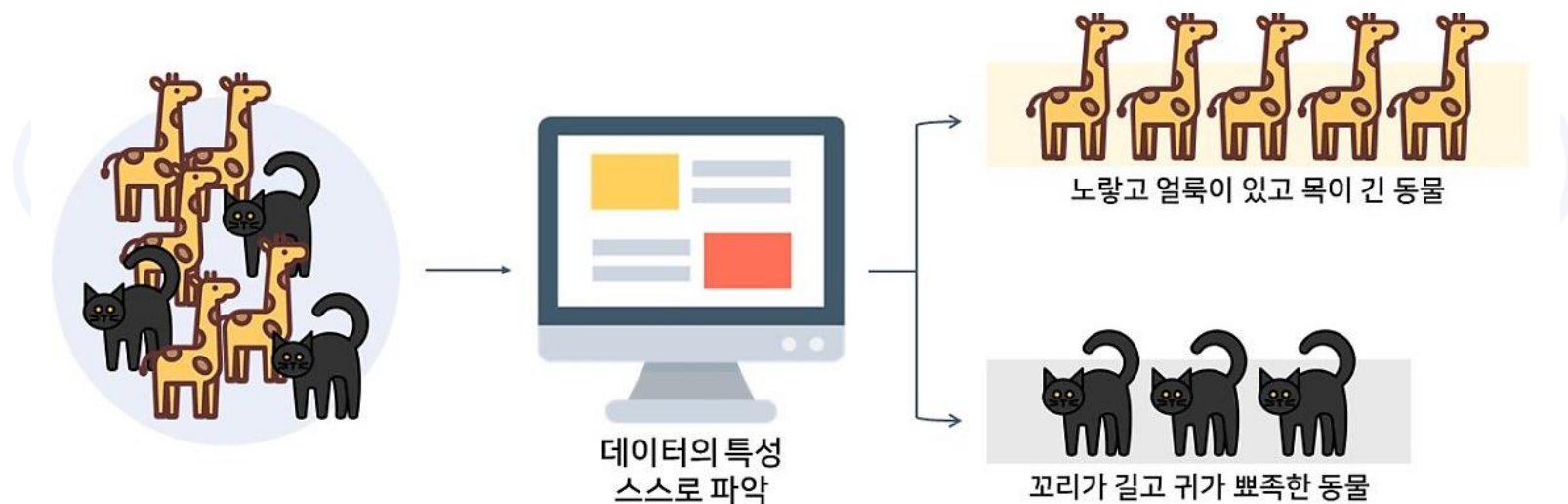
- 지도 학습 (Supervised Learning)

- 라벨이 붙여진 데이터 셋을 학습 후 신규 데이터에 라벨을 붙임
- 예) 이메일 스팸 필터링, 이미지 분류



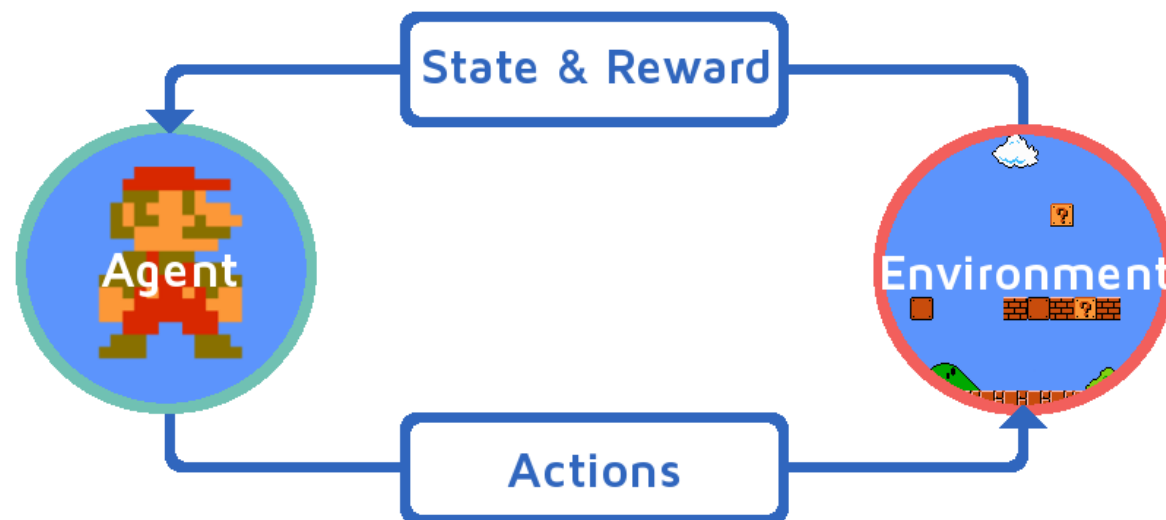
- 비지도 학습 (Unsupervised Learning)

- 라벨이 없는 데이터 들에서 패턴을 찾음
- 예) 유사 단어 클러스터링, 유사 이미지 클러스터링

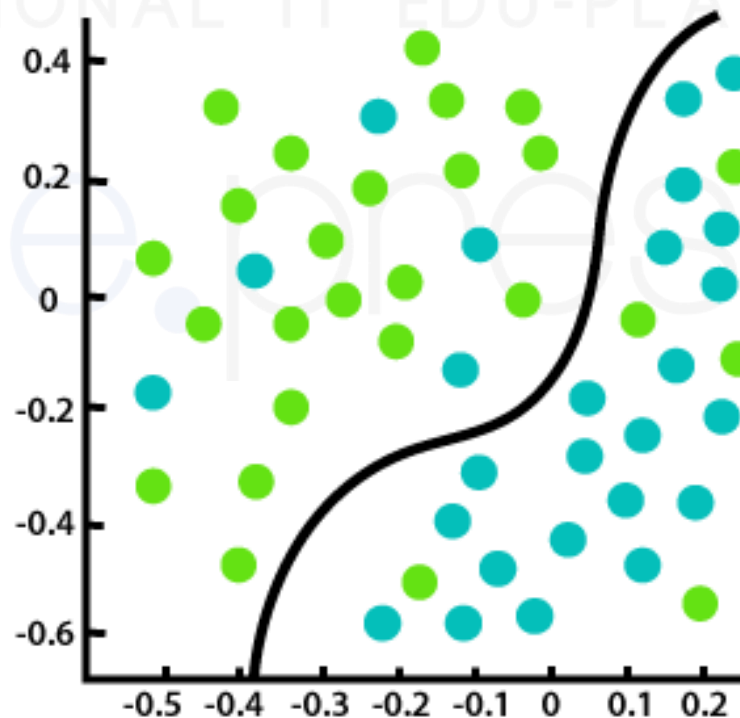


- 강화 학습 (Reinforcement Learning)

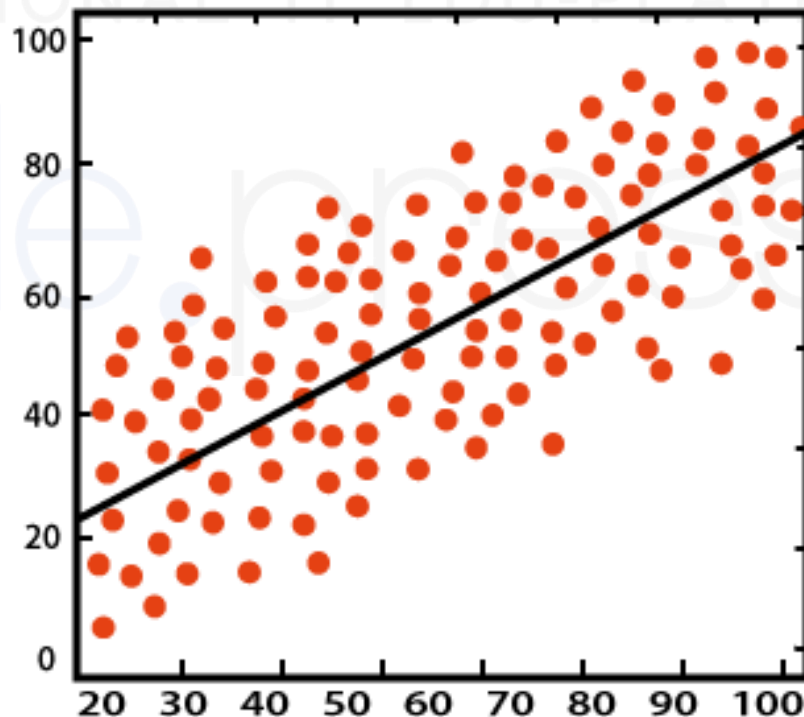
- 피드백 또는 보상을 통하여 학습을 진행
- 예) 체스게임, 알파고



- 분류 (Classification) 분석
 - 지도 학습
 - 문서 분류, 이미지 분류

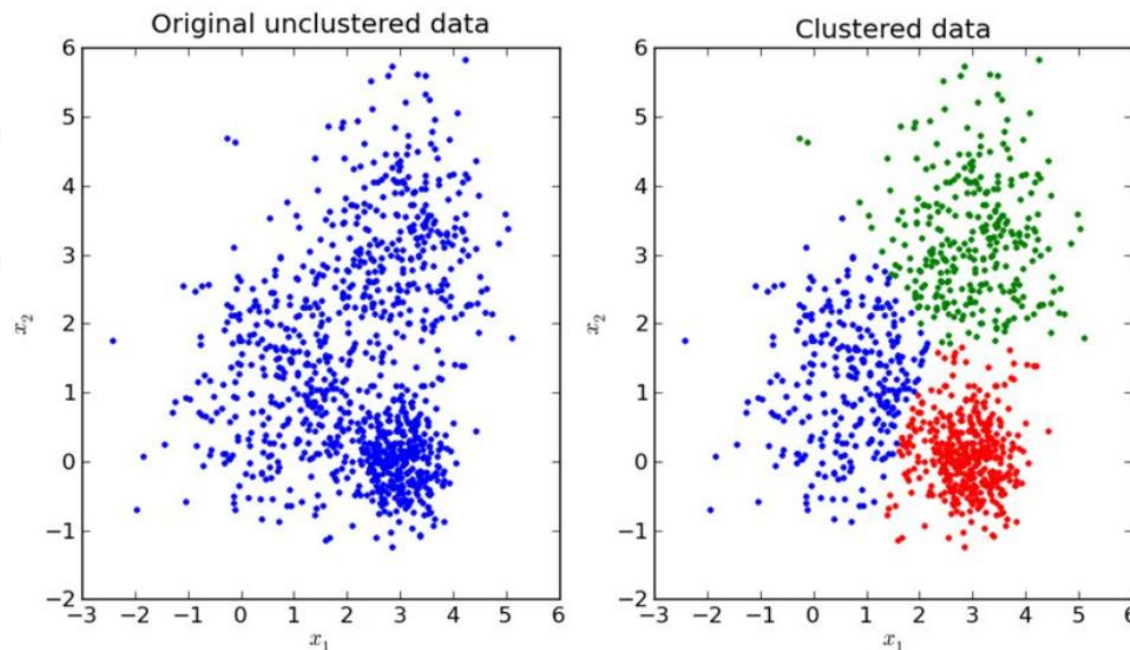


- 회귀 (Regression) 분석
 - 지도학습
 - 주식 가격 예측, 부동산 가격 예측



- 군집 (Clustering)

- 목적 : 라벨이 없는 데이터를 분류하는 것
- 데이터의 특징, 구조를 통해 유사한 특성을 가진 데이터끼리 그룹화 하는 것



CCTV 사진으로 화재를 감지하려면?

지도 학습

고객을 비슷한 군집으로 묶기 위해서는?

비지도 학습

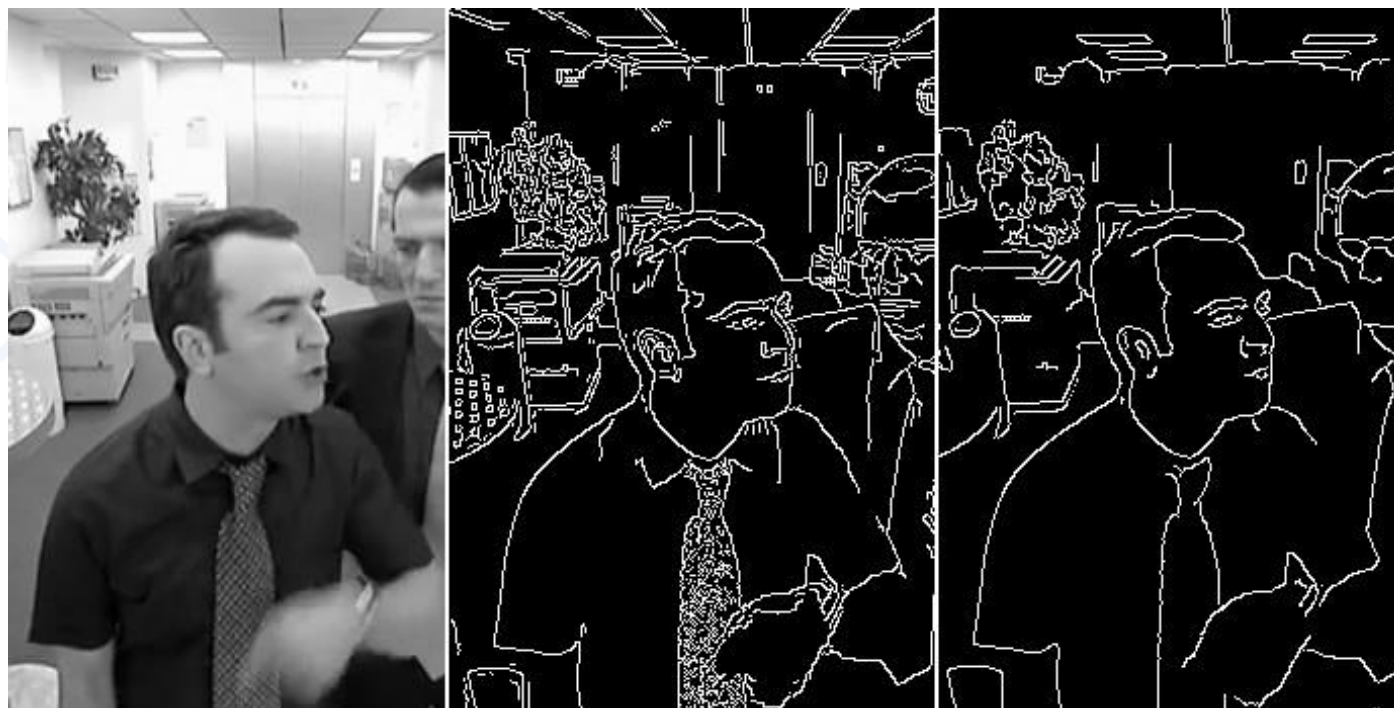
- 문제 정의 및 모델 선정
- 데이터 수집
- 데이터 전처리
- 특징 추출
- 학습
- 검증

- 해결하려는 문제를 명확하게 정의
- 머신 러닝으로 모든 문제를 해결할 수는 없음
- 해결할 문제에 가장 적절한 모델(알고리즘)을 선정

- 데이터의 양이 학습 된 모델의 품질을 결정
- 많은, 다양한 데이터를 수집해야 함
- 인공지능은 사람의 노가다가 필수적임

- **머신 러닝에서 가장 오래 걸리고 힘든 작업**
 - 사진의 경우 크기 조정, 밝기 조정, 중요한 부분 Crop 등의 작업
 - 음성의 경우 노이즈 제거 작업, 단위 시간 별로 자르는 작업
 - 자연어의 경우 오타 수정 등
- **지도 학습의 경우 데이터 Labeling 작업**
 - 수십, 수백만 장의 이미지에 Labeling

- 데이터 중 주요한 특징만 추출하는 작업
 - ex) Edge Detection



- 전처리가 완료된 학습 데이터 셋으로 모델을 학습
- 최적의 값을 찾을 때까지 충분한 데이터로 모델을 학습
- 모델 학습이 완료 후 검증 데이터 셋을 사용하여 모델을 검증
- 다양한 검증 지표들 존재

머신 러닝이란

#PROFESSIONAL IT EDU-PLATFORM
데이터로부터 기계가 학습을 하는
알고리즘