# 삼성전자 Citizen Developer ML 이해 및 구현

## 오리엔테이션

변치웅(fitror.com)
byun0419@gmail.com

# 강사 소개

## 변치웅 강사

- 고려대학교 경제학과 학사
- (전) 삼성생명, 스타트업 유쾌한형제, KTH 근무
- 패스트캠퍼스 데이터 사이언스 스쿨 5기 수료
- (현) fitror.com 대표, 패스트캠퍼스 강사

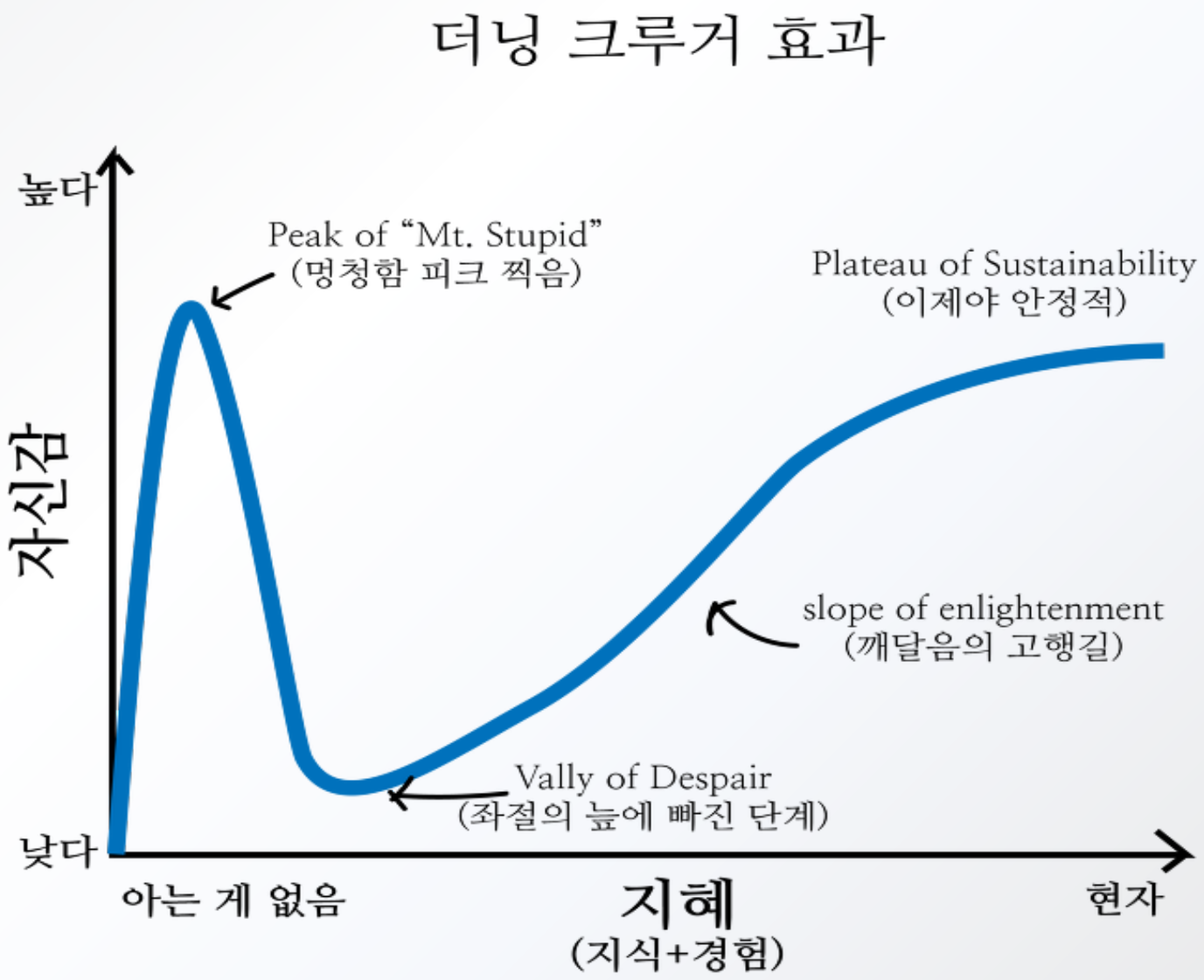
## 1. Doing

- 딥러닝 알고리즘을 활용한 주식 자동화 트레이딩 시스템 개발
- fitror, bitror 프로젝트 진행 중
- 패스트캠퍼스 기업 강의

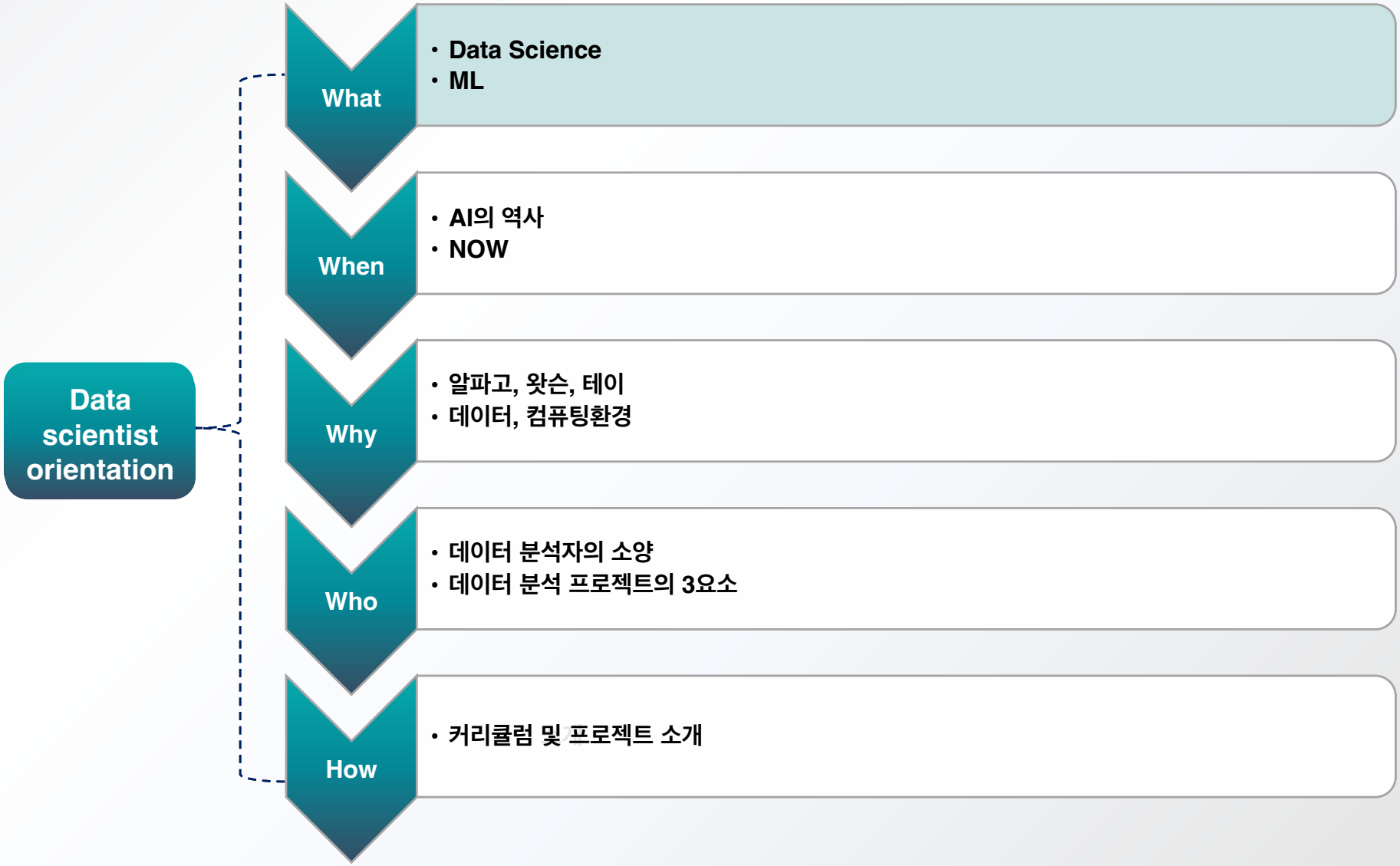
## 2. Done

- 삼성전자, 삼성중공업, 현대모비스 AIM OJT 2, 3, 4, 5기, LG전자, 에쓰오일, SK
- 신한카드, 미래에셋증권, 미래에셋생명, KB금융그룹, KB국민은행, 신한금융그룹, 롯데마트
- SK에너지솔루션, SK이노베이션, 현대자동차, 축산물품질평가원, K-digital training AI과정
- 상기 기업체 프로젝트 약 300여건 컨설팅

# 나는 어디인가?


더닝 크루거 효과
높다
자신감
낮다
Peak of "Mt. Stupid"
(멍청함 피크 찍음)
Plateau of Sustainability
(이제야 안정적)
slope of enlightenment
(깨달음의 고행길)
Vally of Despair
(좌절의 늪에 빠진 단계)
아는 게 없음
지혜
(지식+경험)
현자

# 오리엔테이션 목차


Data scientist orientation
What
• Data Science
• ML
When
• AI의 역사
• NOW
Why
• 알파고, 왓슨, 테이
• 데이터, 컴퓨팅환경
Who
• 데이터 분석자의 소양
• 데이터 분석 프로젝트의 3요소
How
• 커리큘럼 및 프로젝트 소개

# Data Science



## 1. 데이터 과학

- 데이터를 대상으로 실험하고 연구하는 학문
- 데이터를 기반으로 실험과 연구를 통해 정보를 생산하는 기술

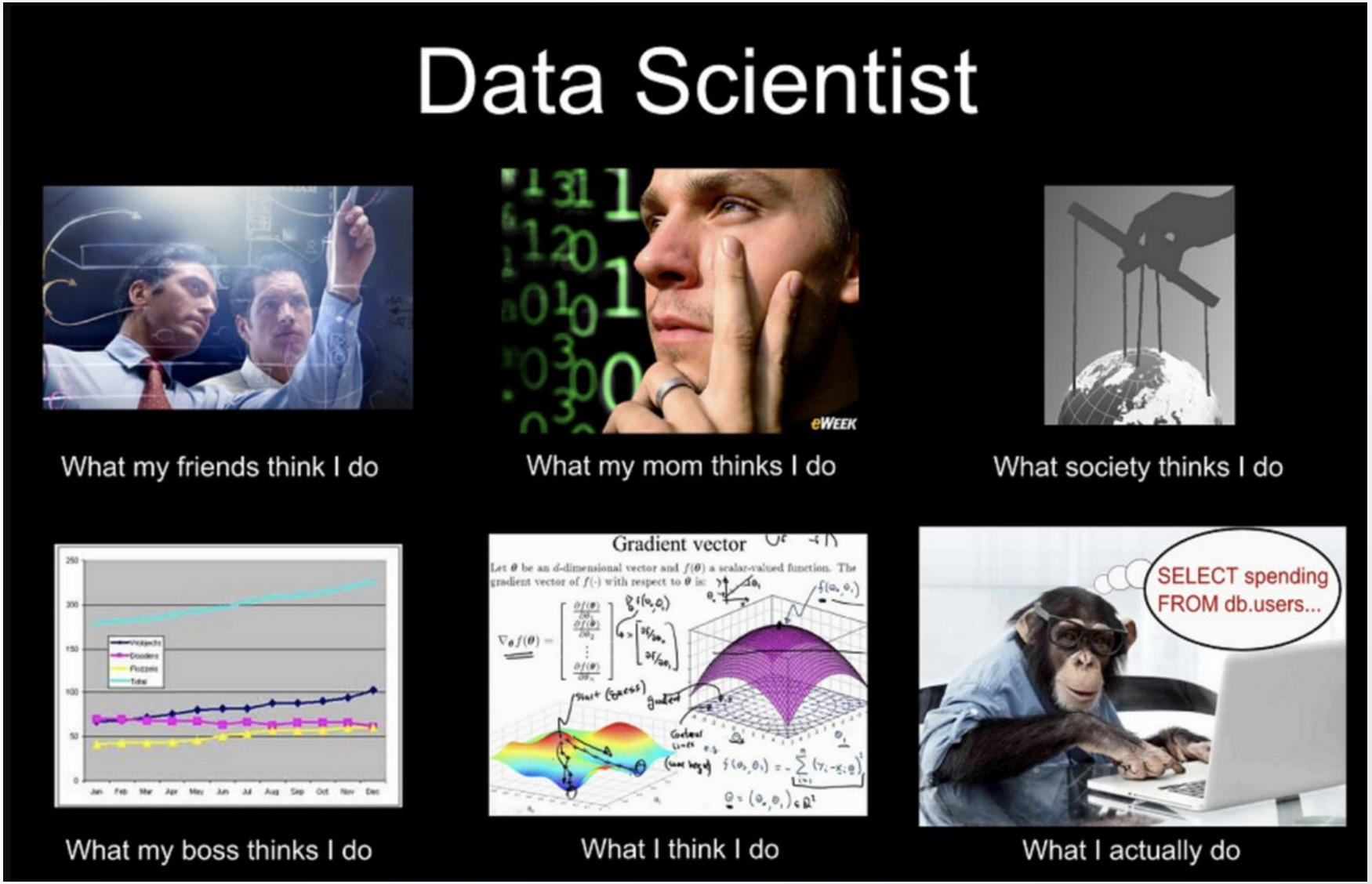
## 2. 사회과학? 자연과학? 데이터과학!!

- 데이터는 인류가 문명을 통해 이룩한 사회과학, 자연과학의 토대로 누적 및 생성
- 데이터를 대상으로 하기에 특정영역에 머무르지 않고 다양한 필드에서 활용

## 3. Data scientist

- 데이터를 다루고 이를 바탕으로 실험과 연구를 통해 정보를 생산하는 자
- 실상은?

# 실상은?




출처: https://twitter.com/kirkdborne/status/1031575412185460736

# 1. Machine learning

- 인공지능의 한 분야
- 기계가 사람처럼 학습을 할 수 있도록 하는 연구분야
- “환경과의 상호작용을 통해서 축적되는 경험적인 데이터를 바탕으로 모델을 자동으로 구축하고 스스로 성능을 향상 시키는 시스템” (Mitchell, 1977)

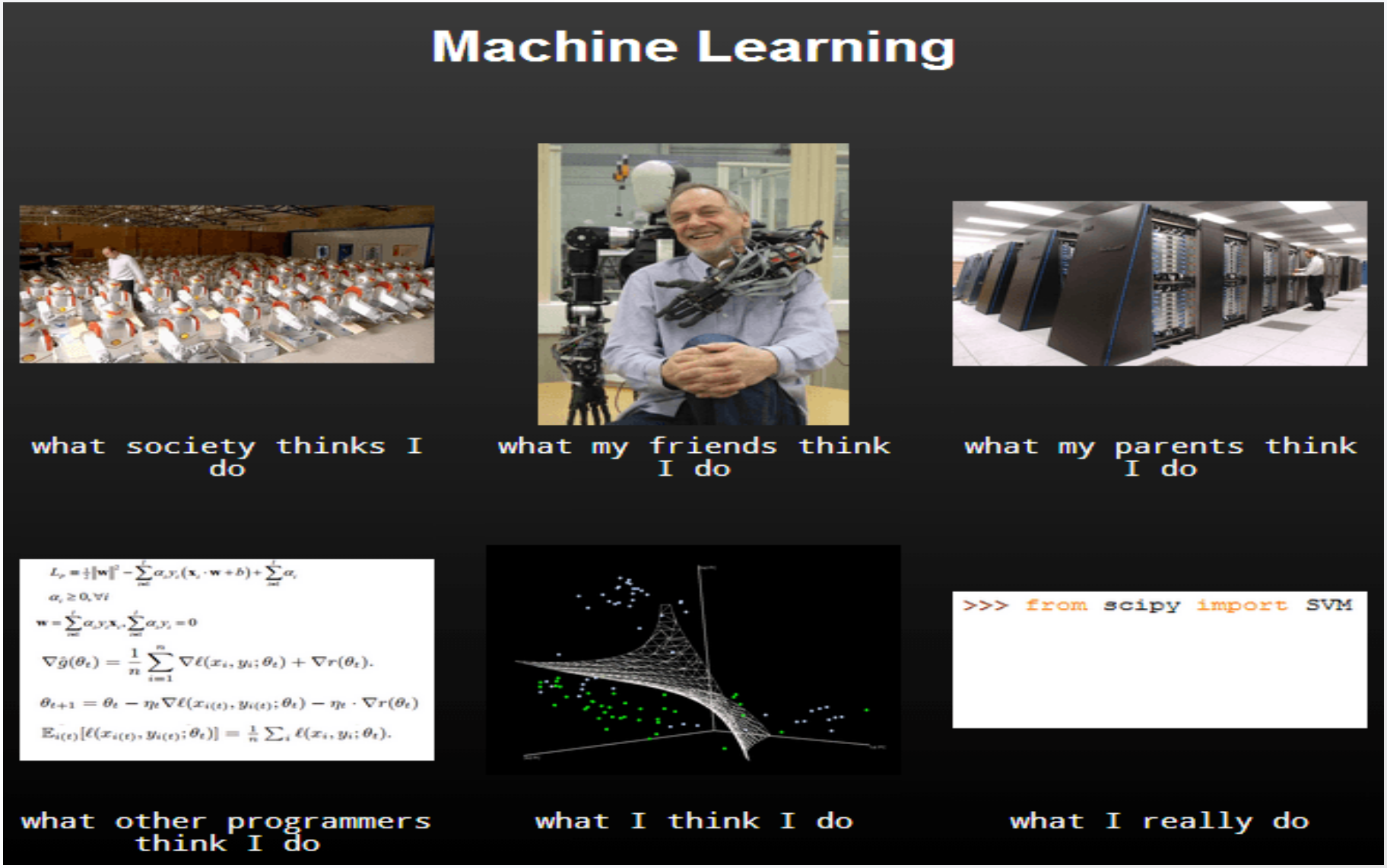
# 2. 데이터와 정보(data, information)

- 프로그램을 운용할 수 있는 형태로 숫자/기호화 된 자료
- 관찰이나 측정을 통해 수집된 데이터를 실제 문제에 도움이 될 수 있도록 해석하고 정리한 지식

# 3. 목적

- 데이터를 통해 정보를 생산하는 것.
- 단순 반복 작업 및 사람이 수행하기에 귀찮고 힘든 일을 컴퓨터가 수행
- 특히, 사람의 지적능력을 요구하는 작업을 대신 수행

# 오해가 있더라



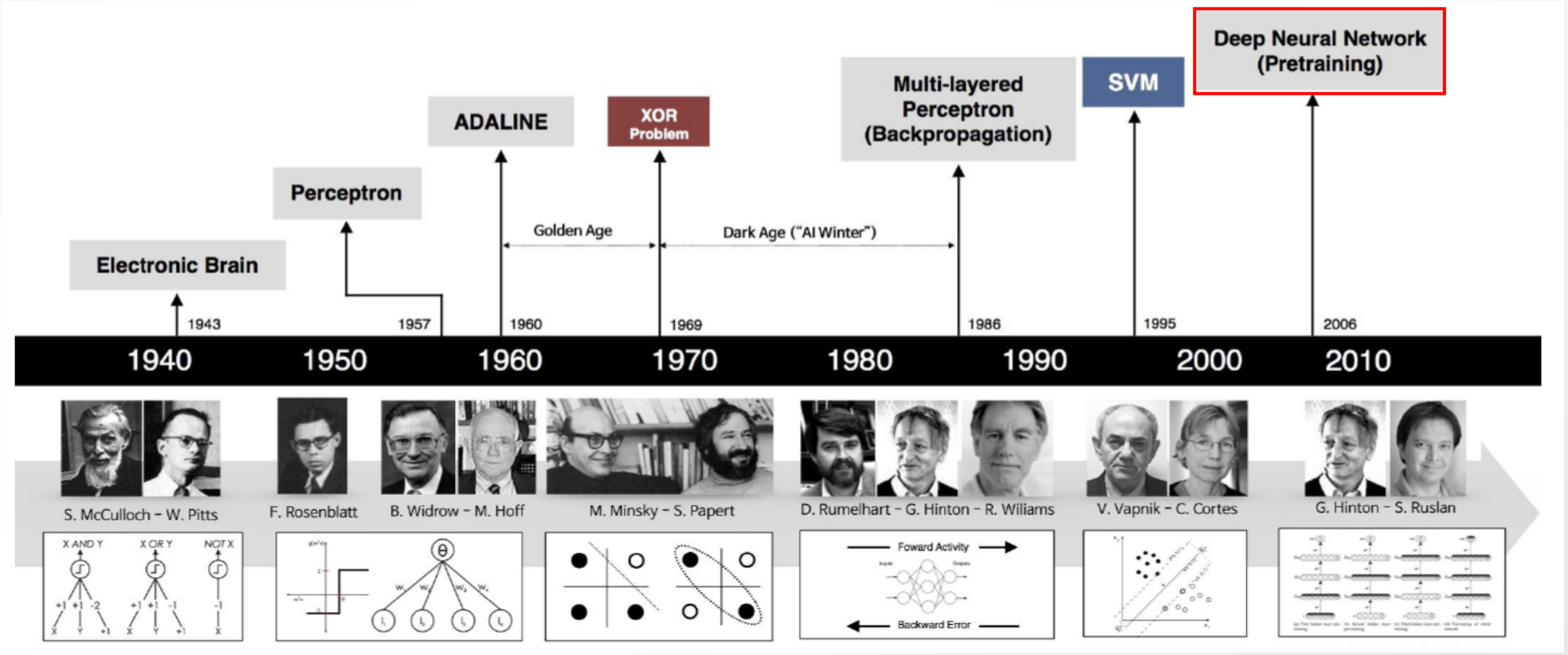

출처: https://twitter.com/kirkdborne/status/1031575412185460736

# 오리엔테이션 목차


Data scientist orientation
What
• Data Science
• ML
When
• AI의 역사
• NOW
Why
• 알파고, 왓슨, 테이
• 데이터, 컴퓨팅환경
Who
• 데이터 분석자의 소양
• 데이터 분석 프로젝트의 3요소
How
• 커리큘럼 및 프로젝트 소개

# 시작은 1940년대



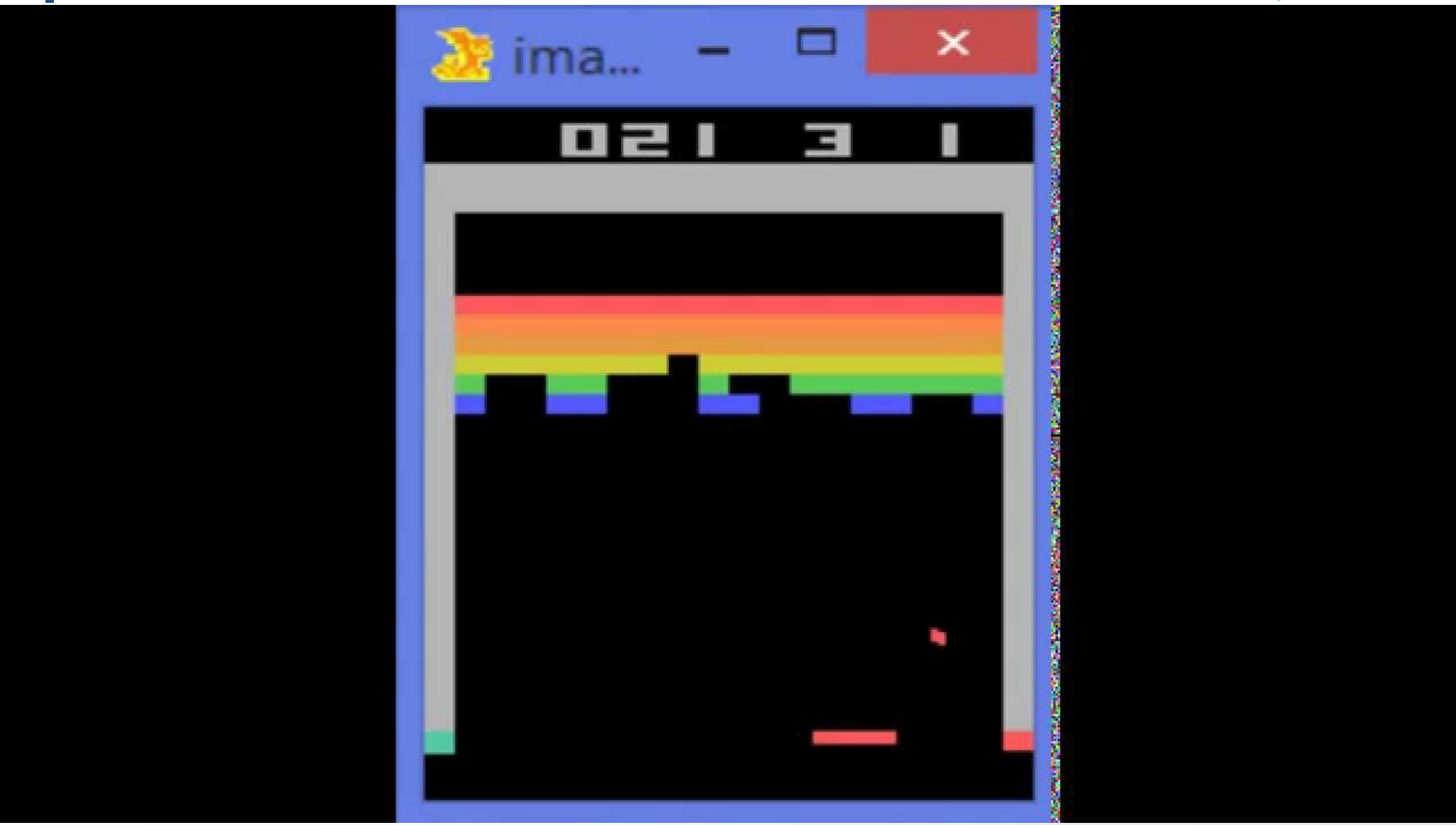



출처:http://hochul.net/blog/data_analysis_machine_learning_basic_1s/

XOR 시각화 : https://playground.tensorflow.org/ 힐튼교수님 논문 : https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en

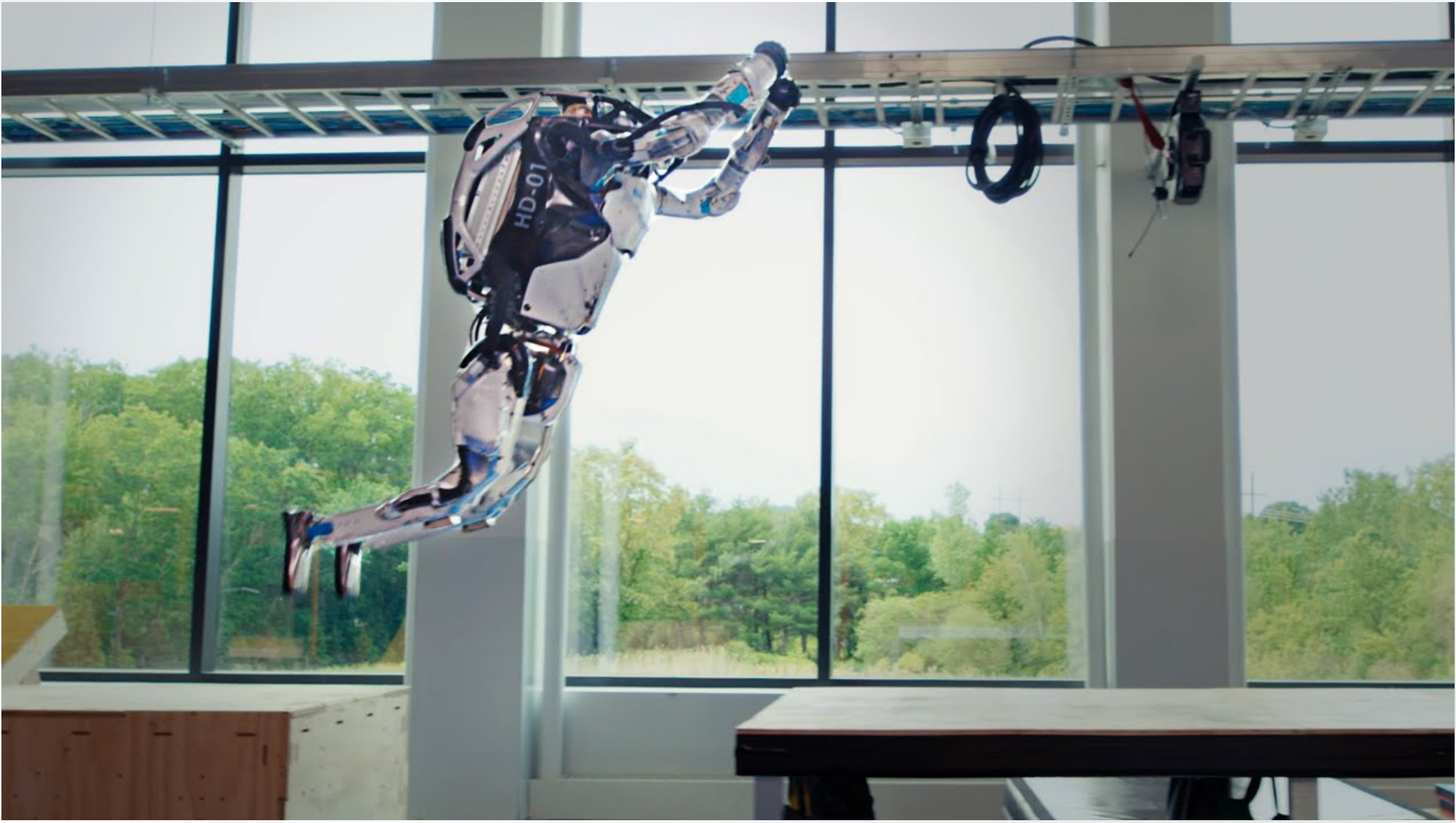
# Now





# Now





# 오리엔테이션 목차


Data scientist orientation
What
• Data Science
• ML
When
• AI의 역사
• NOW
Why
• 알파고, 왓슨, 테이
• 데이터, 컴퓨팅환경
Who
• 데이터 분석자의 소양
• 데이터 분석 프로젝트의 3요소
How
• 커리큘럼 및 프로젝트 소개


AlphaGo
Deep Learning

# 시대가 되었다

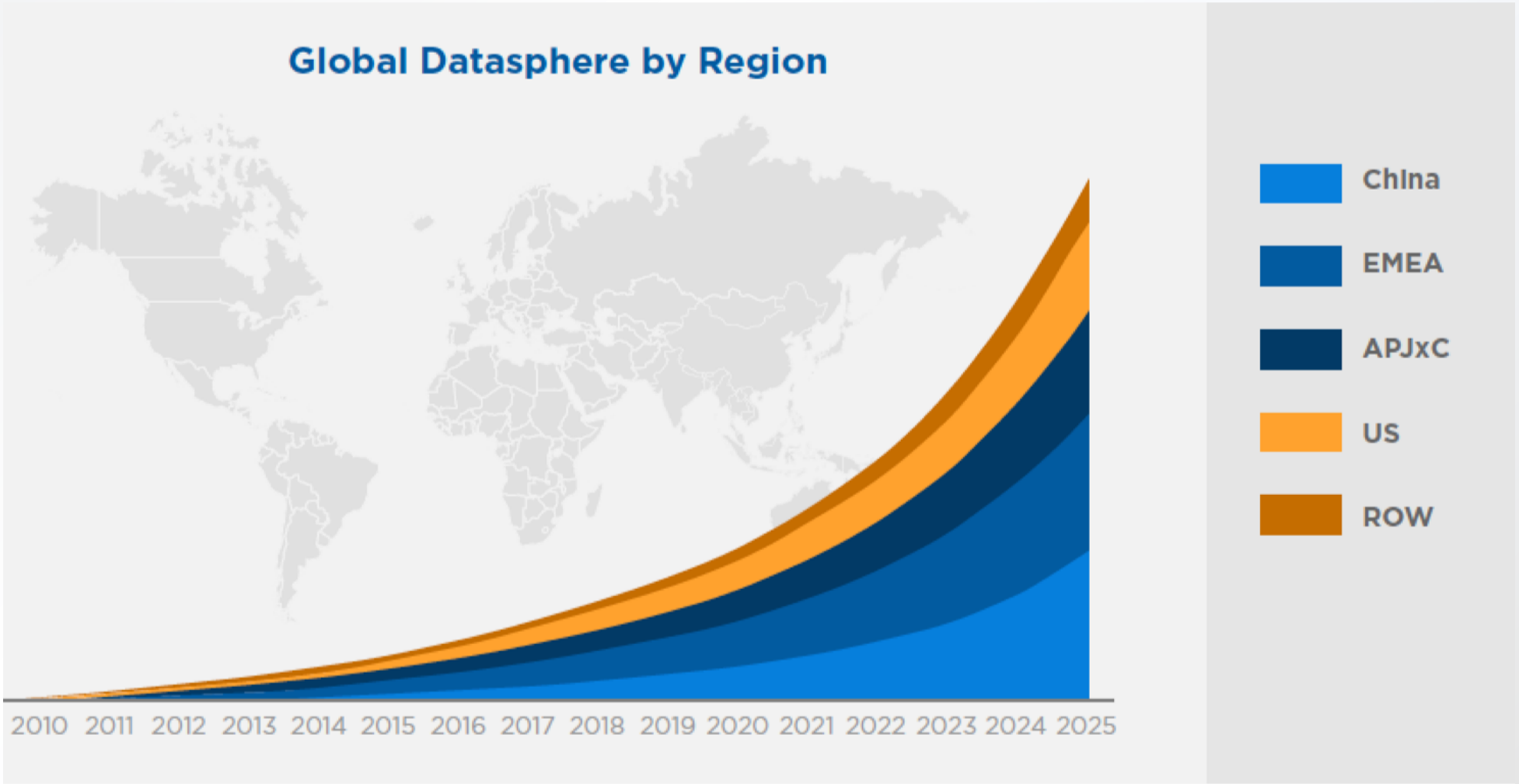


## 1. 데이터의 폭발적인 생산

- 지수적으로 늘어나는 데이터의 생산량
- 온라인 기반 활동이 늘어나며 데이터의 생산 및 누적이 쉬워짐
- 한국의 경우 IT 인프라 기준 데이터 생산 세계 6위
- 실제 데이터를 기반으로 생산활동은? (SSG.com 일일 데이터 4GB)

# 시대가 되었다






출처:https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/

# 시대가 되었다

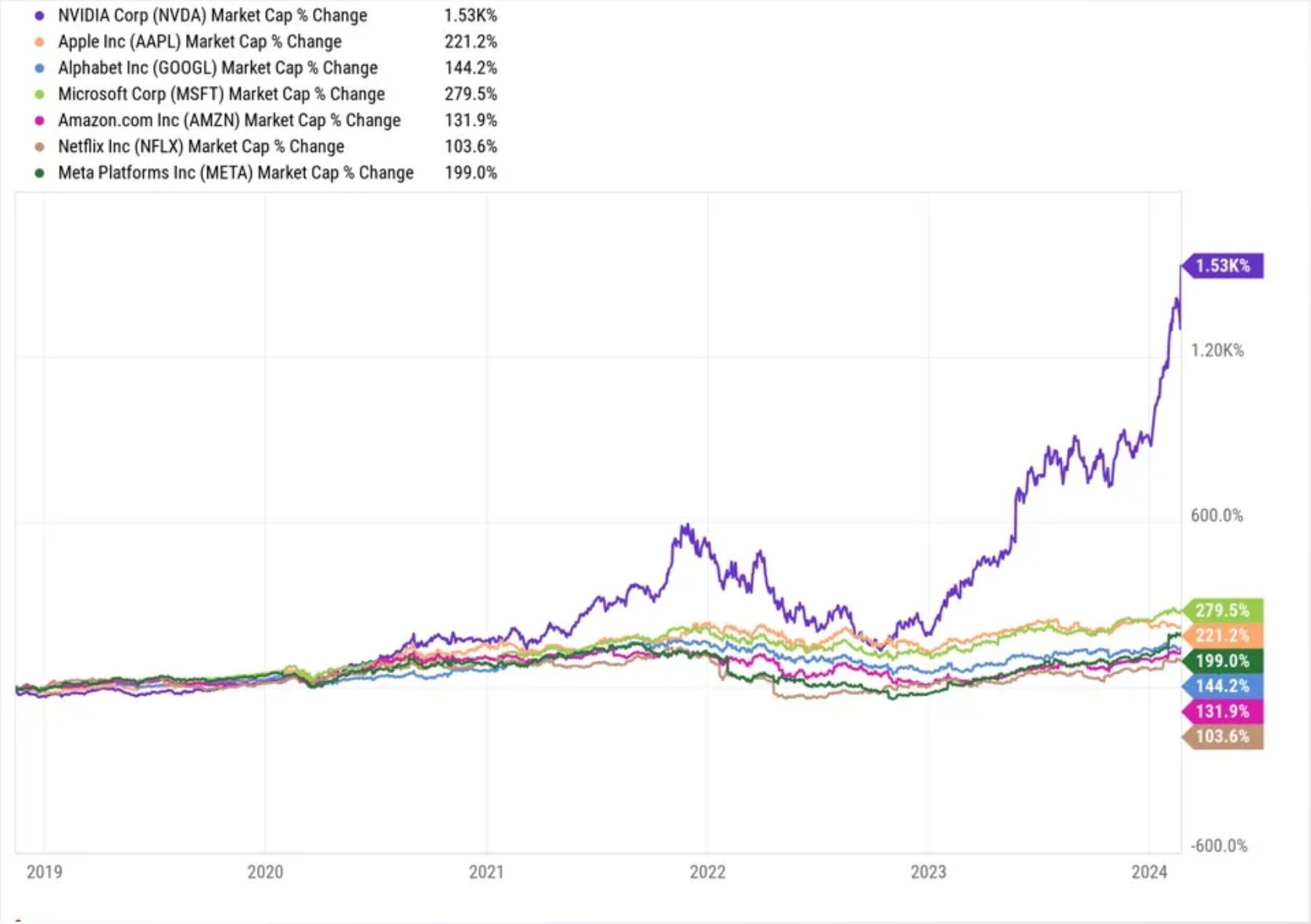


## 2. 컴퓨팅 환경

- CPU 및 GPU 기반 연산속도 증가
- 기존 컴퓨팅 환경으로 오랜 기간이 걸리던 작업들이 가능해짐
- Python 등 오픈소스 기반의 프로그래밍 언어의 대중화
- 시스템 반도체 수요 폭증

# 시대가 되었다




NVIDIA Corp (NVDA) Market Cap % Change 1.53K%
Apple Inc (AAPL) Market Cap % Change 221.2%
Alphabet Inc (GOOGL) Market Cap % Change 144.2%
Microsoft Corp (MSFT) Market Cap % Change 279.5%
Amazon.com Inc (AMZN) Market Cap % Change 131.9%
Netflix Inc (NFLX) Market Cap % Change 103.6%
Meta Platforms Inc (META) Market Cap % Change 199.0%
1.53K%
1.20K%
600.0%
279.5%
221.2%
199.0%
144.2%
131.9%
103.6%
-600.0%
2019
2020
2021
2022
2023
2024

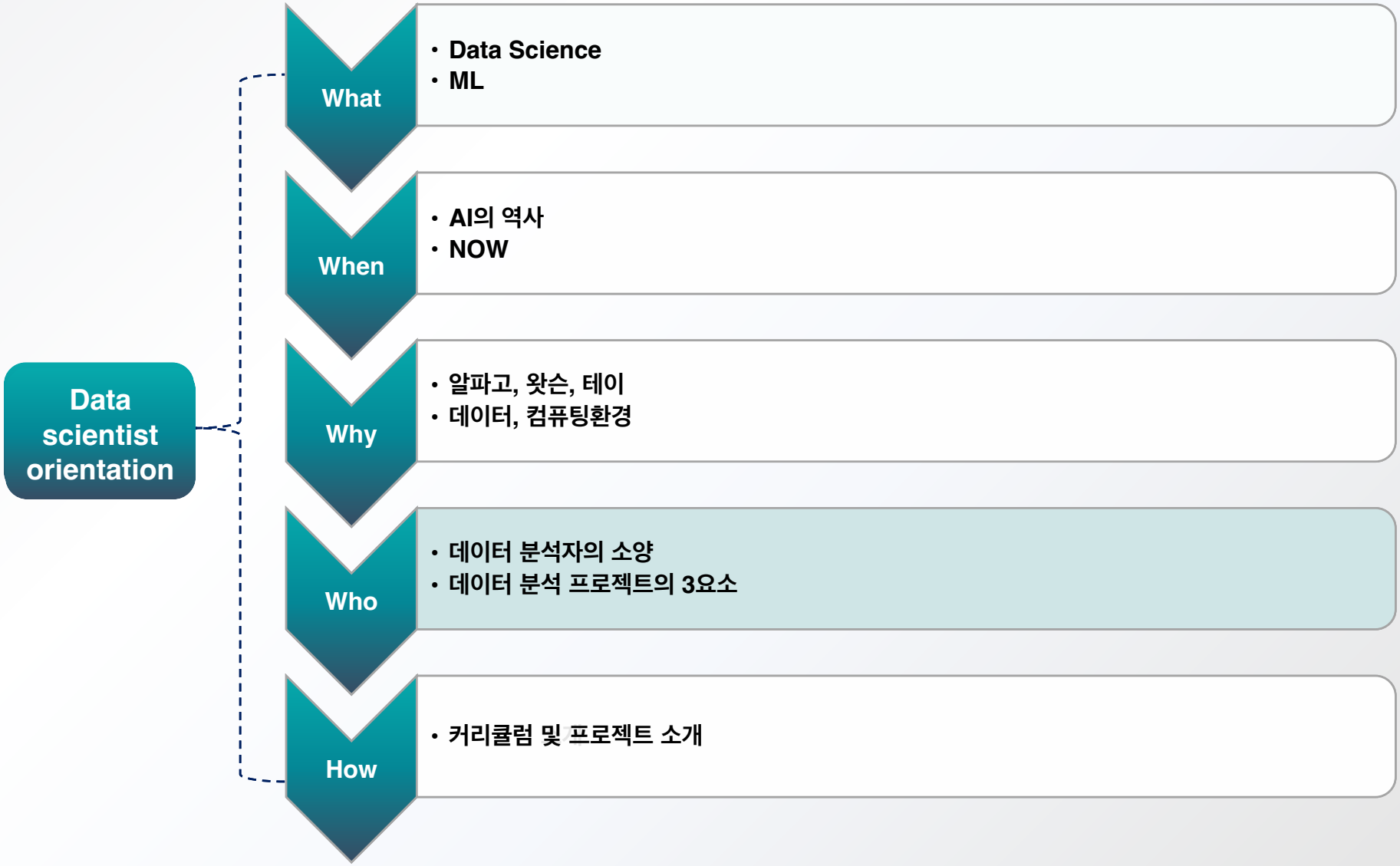
# 시대가 되었다



## 3. 우리가 해야 할 일은?

- 경쟁사가 혹은 타 부서에서 머신러닝/딥러닝을 활용한 생산툴을 개발하였다.
- 이 소식을 알게 된 회사 사장님, 팀장님들의 반응은?
- 우리가 살아남기 위해 해야 할 일은?

# 오리엔테이션 목차


Data scientist orientation
What
• Data Science
• ML
When
• AI의 역사
• NOW
Why
• 알파고, 왓슨, 테이
• 데이터, 컴퓨팅환경
Who
• 데이터 분석자의 소양
• 데이터 분석 프로젝트의 3요소
How
• 커리큘럼 및 프로젝트 소개

# 누군가는 하게된다



## 1. 내부전문가

- 회사의 사정을 잘 알고있는 내부전문가
- 프로젝트 수행에 필요한 사내 프로세스 및 지원, 협조 체계 구성이 가능한 사람
- 의사결정에 대한 권한 혹은 이에 준하는 권한을 가진 사람

### Skills

- 현업에 필요한 혹은 해결해야 할 문제 인식, 프로젝트 기획의 기초
- 직면한 문제가 무엇인지, 개선할 점에 대한 정확한 파악이 필요
- 문제해결을 위한 협조체계, 지원요청, 예산, 성과측정

# 누군가는 하게된다



## 2. 데이터 분석가

- 문제해결을 위한 방법을 찾을 수 있는 데이터 사이언티스트
- 데이터에 기반한 프로그래밍, 알고리즘에 능한 사람
- 다양한 데이터를 다루며 도메인 지식 외 지원을 할 수 있는 사람

### Skills

- 데이터를 바탕으로 문제해결을 위한 방법을 찾는 능력
- 수학, 통계, 프로그래밍, 알고리즘 활용
- 유연한 문제해결 능력

# 누군가는 하게된다



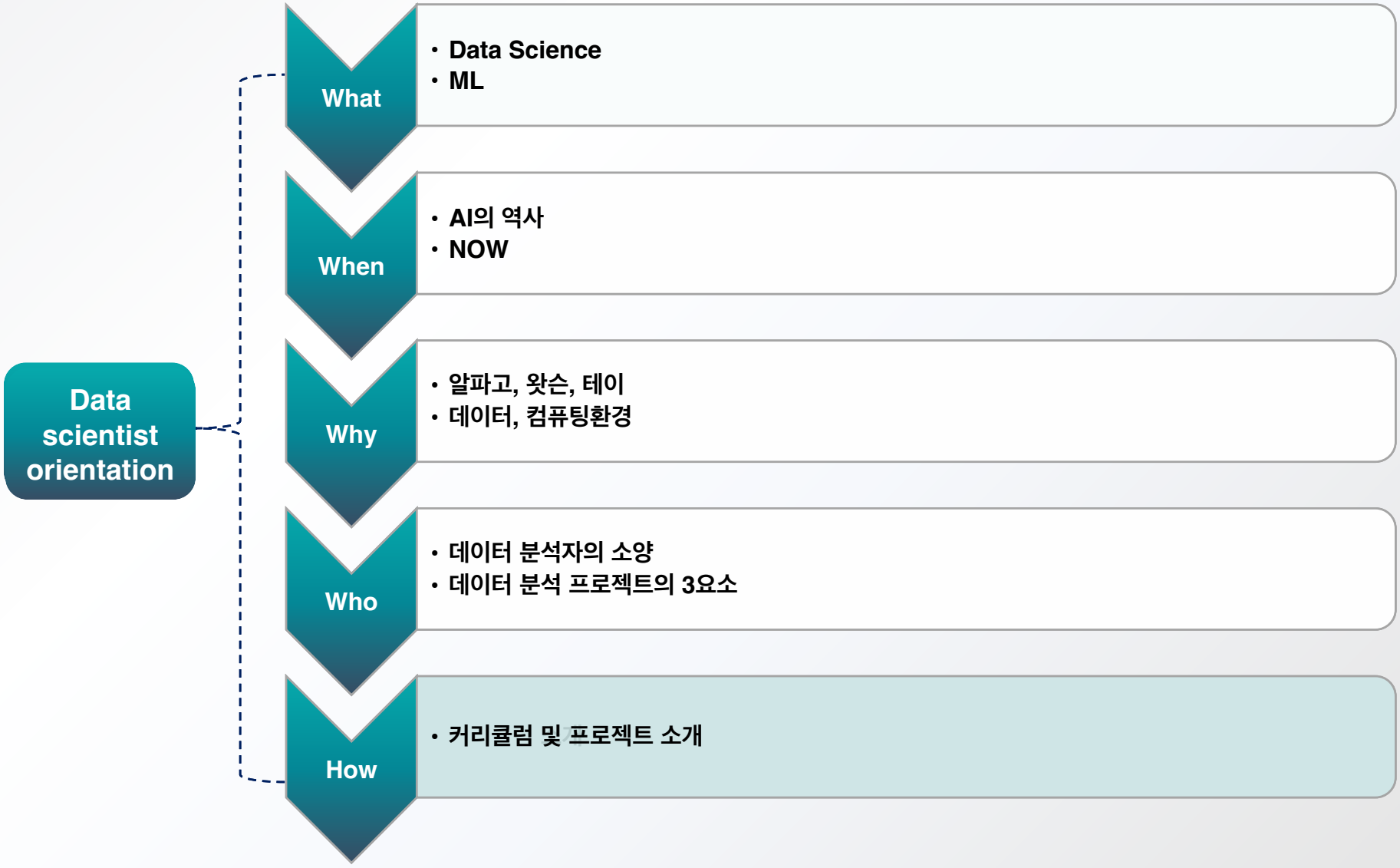
## 3. 외부전문가

- 내부인원만으로 프로젝트 진행 시 발생할 수 있는 문제점을 알 수 있는 사람
- 프로젝트의 기획부터 평가까지 전 과정을 매니징하는 사람
- 프로젝트 매니징 + 의사결정권자의 설득 지원이 가능한 사람

### Skills

- 데이터 과학자
- 수학, 통계, 프로그래밍, 알고리즘의 이해가 뛰어난 사람
- 다양한 프로젝트 진행 경험
- 프로젝트 매니징이 가능한 사람

# 오리엔테이션 목차


Data scientist orientation
What
• Data Science
• ML
When
• AI의 역사
• NOW
Why
• 알파고, 왓슨, 테이
• 데이터, 컴퓨팅환경
Who
• 데이터 분석자의 소양
• 데이터 분석 프로젝트의 3요소
How
• 커리큘럼 및 프로젝트 소개

# 커리큘럼 소개



## 1. 머신러닝 개요

- 머신러닝 모델의 종류 및 구분
- 지도학습, 비지도학습, prediction, classification, clustering
- 머신러닝 모델링을 위한 데이터 플로우 프로세스를 학습
- train, test, model, fit, predict

## 2. 지도학습

- label이 있는 데이터에 적용가능한 지도학습 모델을 학습
- linear regression, logistic regression, Decision tree, Random forest, xgboost

## 3. 머신러닝을 위한 전처리

- 머신러닝 모델링을 위한 전처리 방법을 학습
- 결측치 처리, 컬럼데이터 변환, 범주형데이터 처리, 스케일링, 클래스불균형