Introduction to Data Science

Why Change?

- **IBM**
  - (1960 – 2000) From hardware, software company
  - (2000 – present) To a **service** company (mainly, consulting)
- If organizations don't successfully change and innovate, ***they die***

What is Intelligence?

Capacity to learn and solve problems

Specifically:

- Ability to solve novel problems
- Ability to act rationally
- Ability to act like humans

❑ **Ability to interact with the real world**

- To perceive, understand, and act
- Eg., Speech Recognition and Synthesis
- Eg. Image Understanding

❑ **Reasoning and Planning**

- Modelling the external world, given input
- Solving new problems, planning, and making decisions
- Ability to deal with unexpected problems, uncertainties

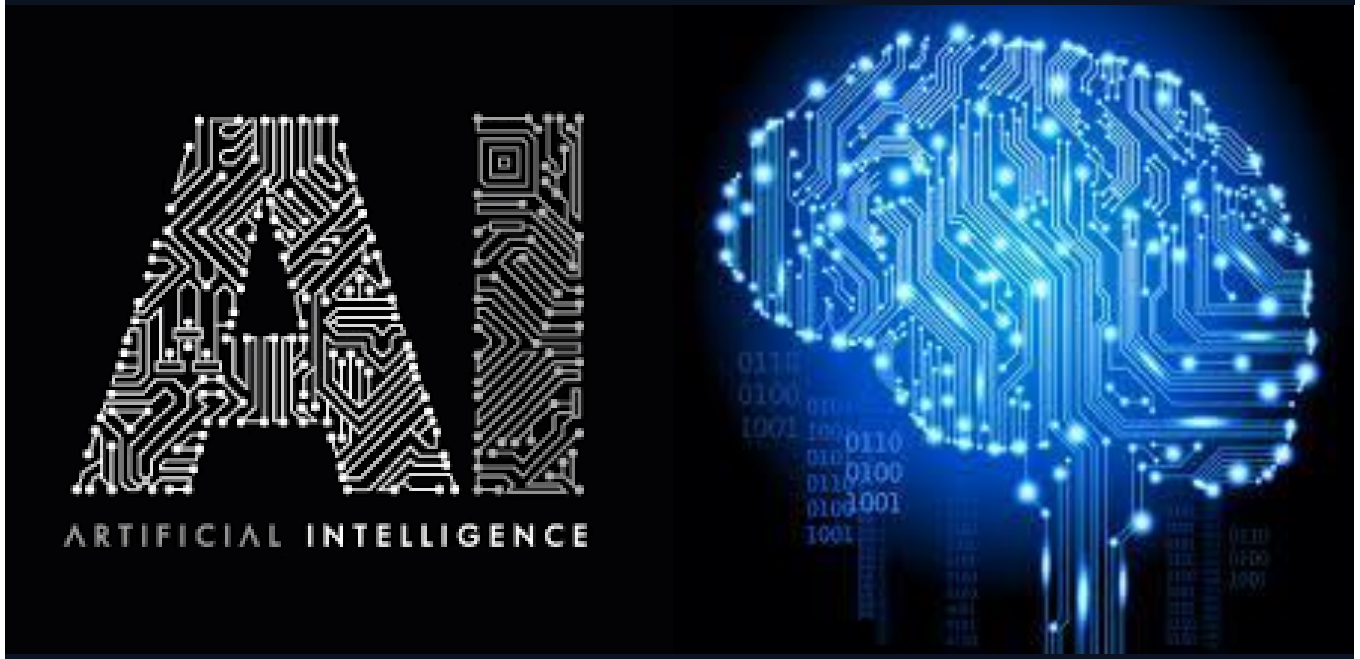
❑ **Learning and Adaptation**

- Humans are continuously learning and adapting
- Eg., A baby learning to categorize and recognize animals

Steering machines, towards understanding human intelligence

Artificial Intelligence (AI) is a popular branch of Computer Science concerned with building ***intelligent*** smart machines capable of performing intelligent tasks

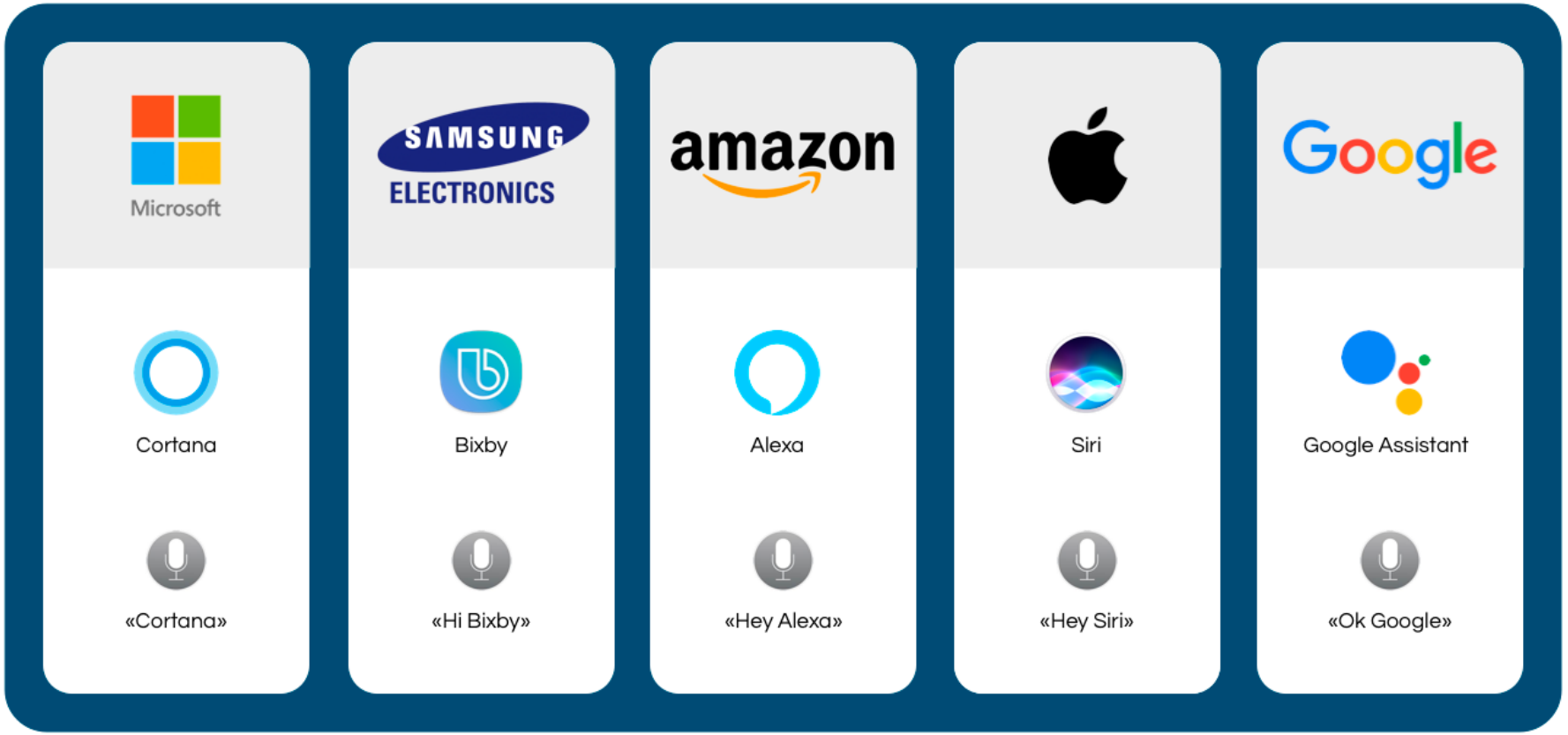
AI is a bridge between art and science



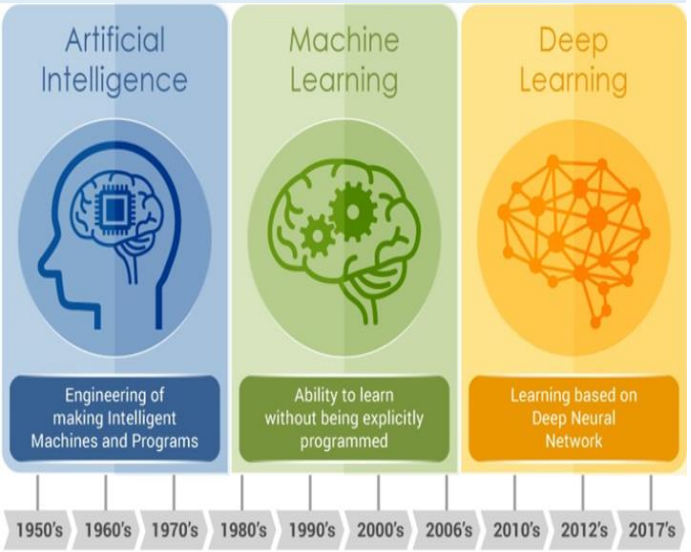
When a machine mimics the cognitive functions that humans ought to carryout

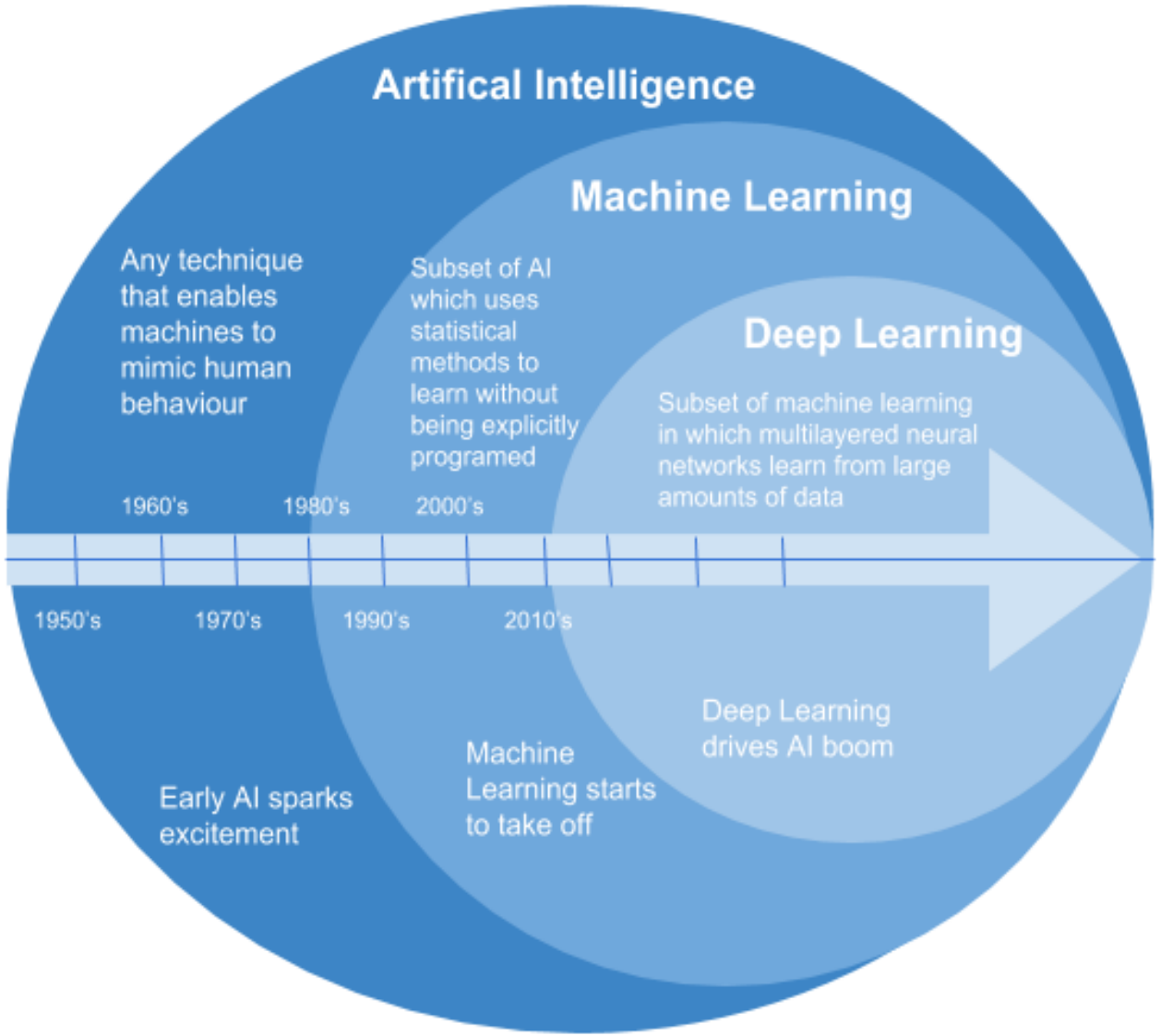
Real-Time Artificial Intelligence

Virtual Assistants



Evolution of AI





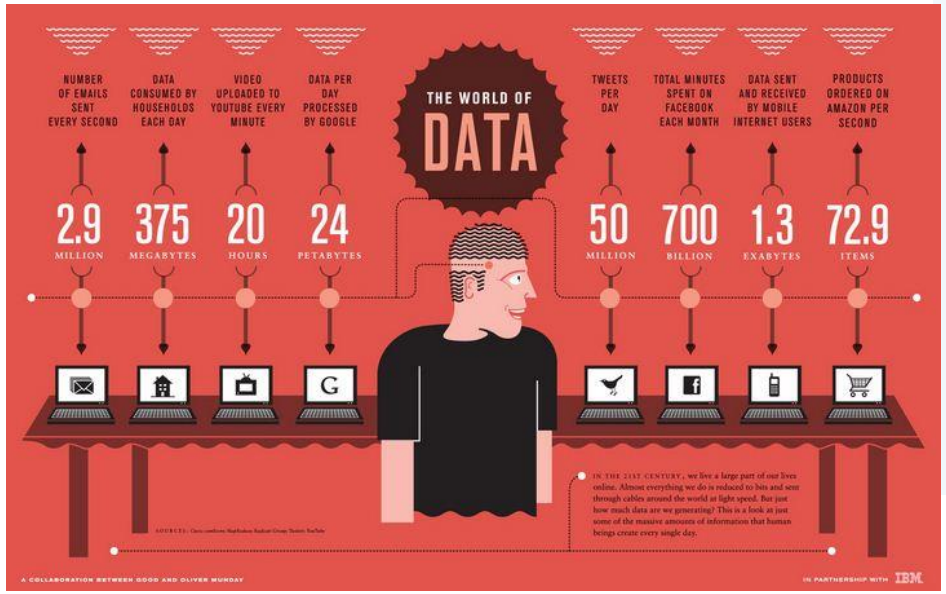
The science and engineering of making intelligent machines, especially intelligent computer programs

Data All Around

- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - Financial transactions, bank/credit transactions
  - Online trading and purchasing
  - Social Network

How Much Data Do We have?

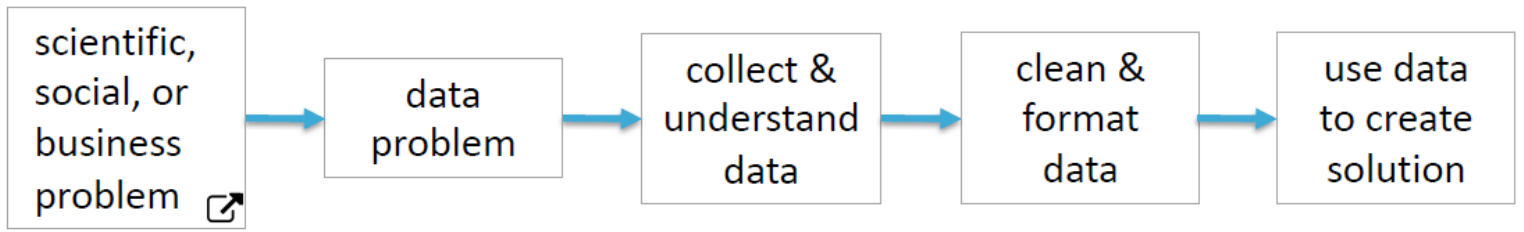
- Google processes 20 PB a day (1,048,576 Gigabytes.)
- Facebook has 60 TB of daily logs
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)
- 1000 genomes project: 200 TB



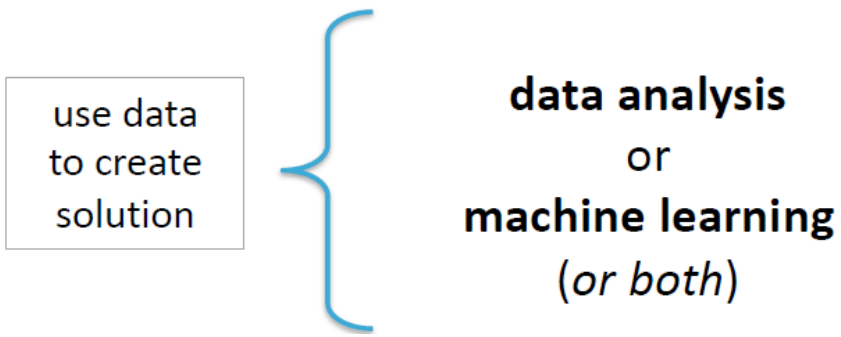
Data Science

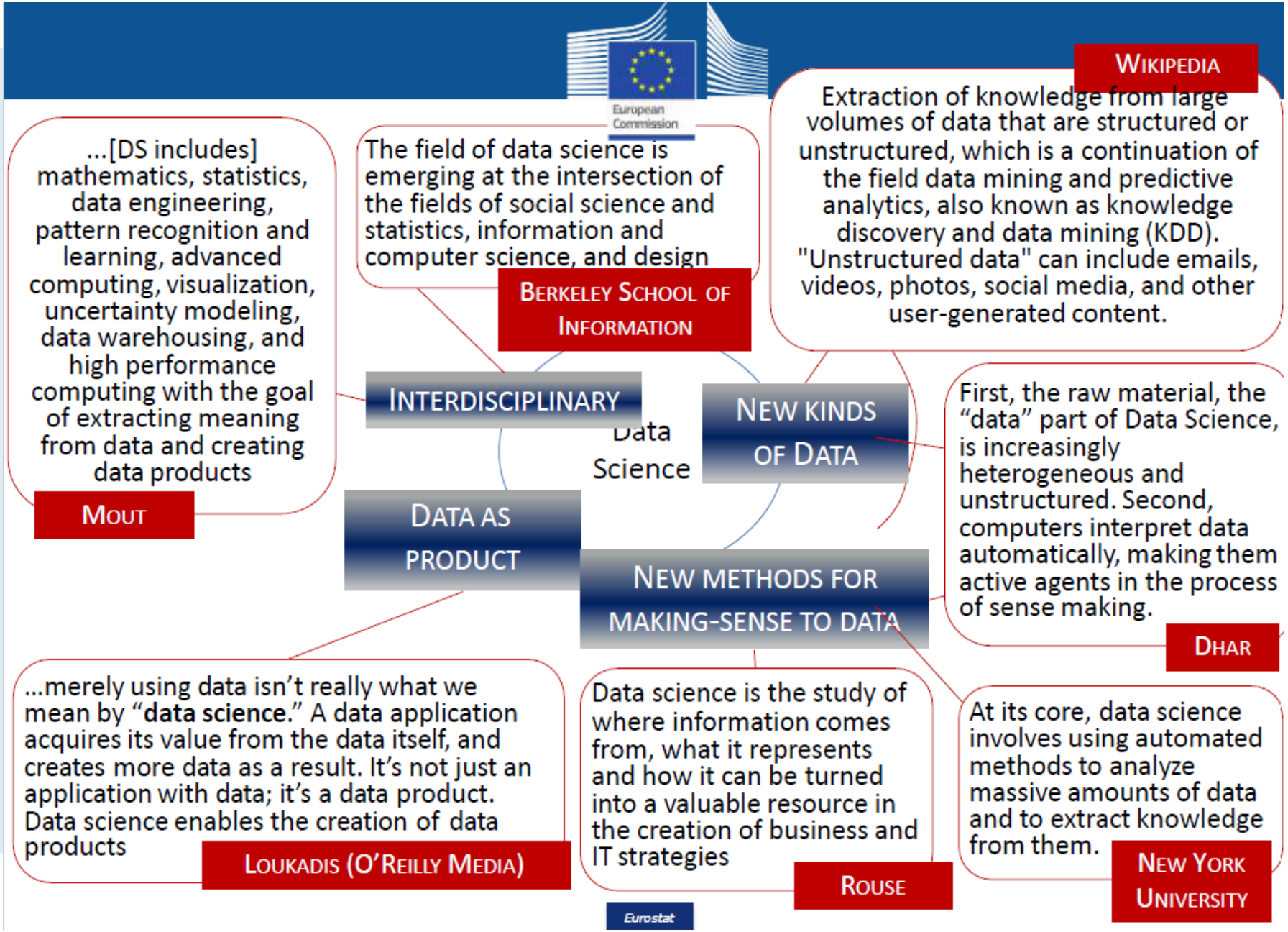
WHAT IS DATA SCIENCE?

...solving problems with data...

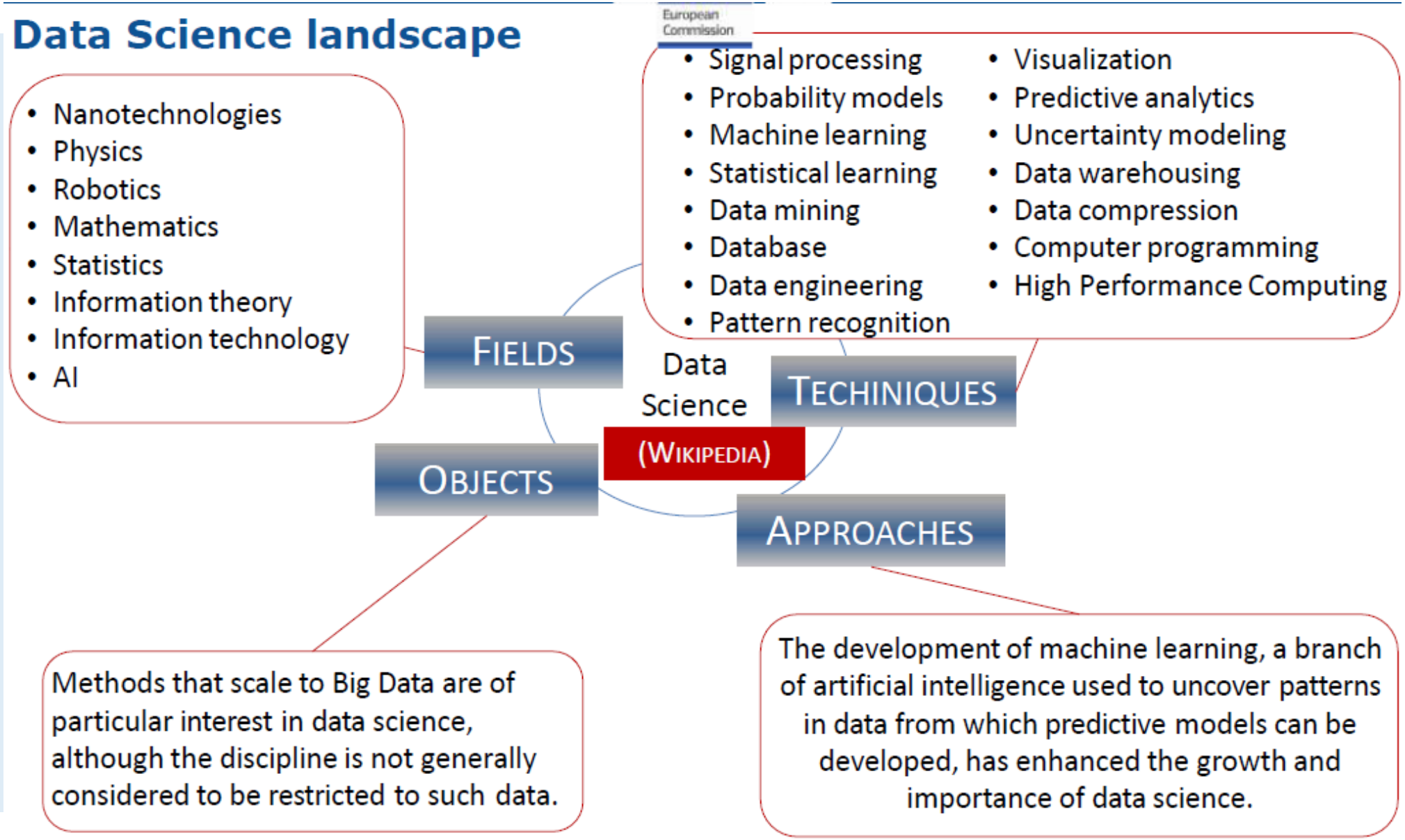


...which step is most challenging?




European Commission
...[DS includes] mathematics, statistics, data engineering, pattern recognition and learning, advanced computing, visualization, uncertainty modeling, data warehousing, and high performance computing with the goal of extracting meaning from data and creating data products
MOUT
The field of data science is emerging at the intersection of the fields of social science and statistics, information and computer science, and design
BERKELEY SCHOOL OF INFORMATION
WIKIPEDIA
Extraction of knowledge from large volumes of data that are structured or unstructured, which is a continuation of the field data mining and predictive analytics, also known as knowledge discovery and data mining (KDD). "Unstructured data" can include emails, videos, photos, social media, and other user-generated content.
INTERDISCIPLINARY
Data Science
NEW KINDS OF DATA
First, the raw material, the "data" part of Data Science, is increasingly heterogeneous and unstructured. Second, computers interpret data automatically, making them active agents in the process of sense making.
DHAR
DATA AS PRODUCT
NEW METHODS FOR MAKING-SENSE TO DATA
...merely using data isn't really what we mean by "**data science**." A data application acquires its value from the data itself, and creates more data as a result. It's not just an application with data; it's a data product. Data science enables the creation of data products
LOUKADIS (O'REILLY MEDIA)
Data science is the study of where information comes from, what it represents and how it can be turned into a valuable resource in the creation of business and IT strategies
ROUSE
At its core, data science involves using automated methods to analyze massive amounts of data and to extract knowledge from them.
NEW YORK UNIVERSITY
Eurostat

Data Science landscape

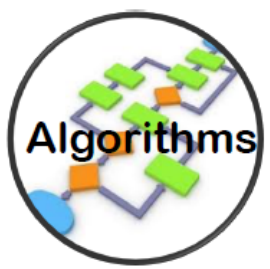


WHAT IS DATA ANALYSIS?

...using data to discover useful information...

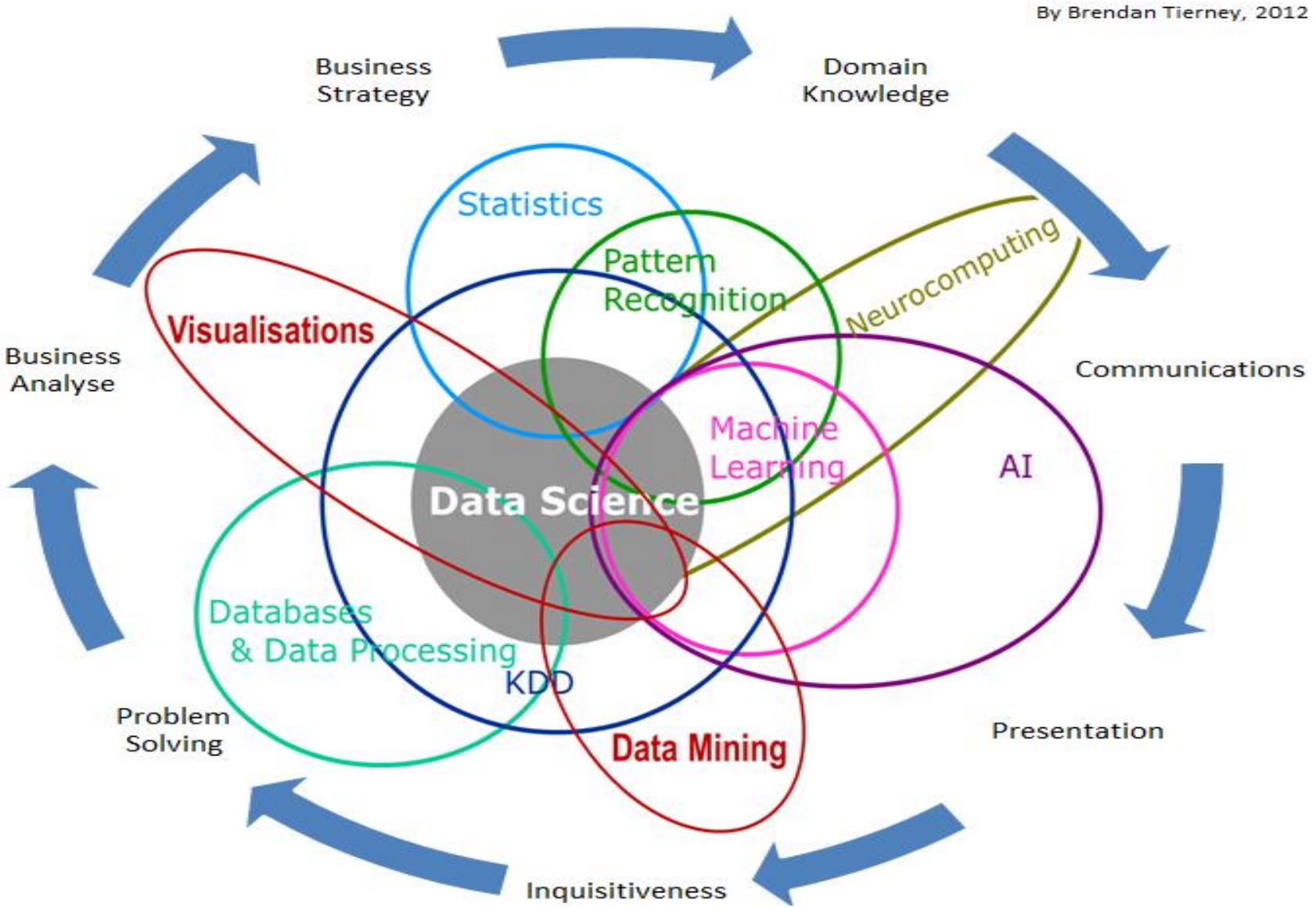
- **data**: anything you can *measure* or *record*

- **statistics**: summarize (and visualize) *main characteristics* of the data

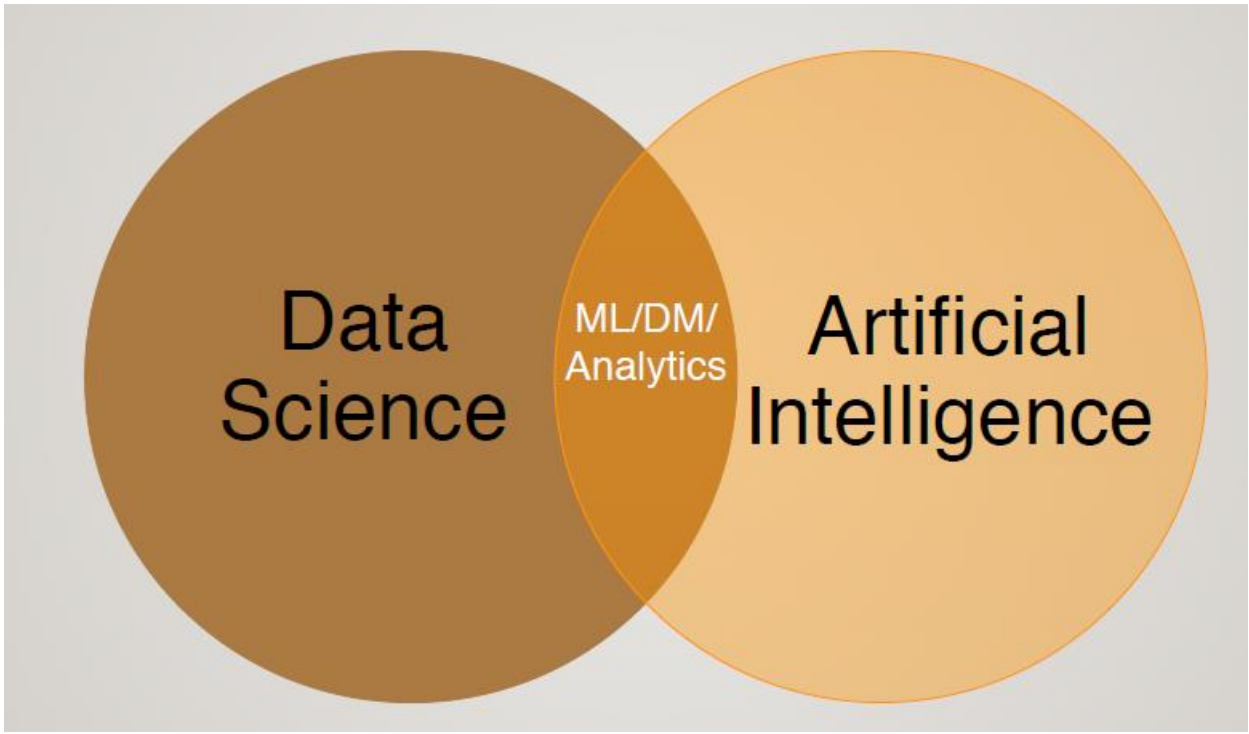


- **algorithms**: apply algorithms to find *patterns* in the data

Data Science Is Multidisciplinary

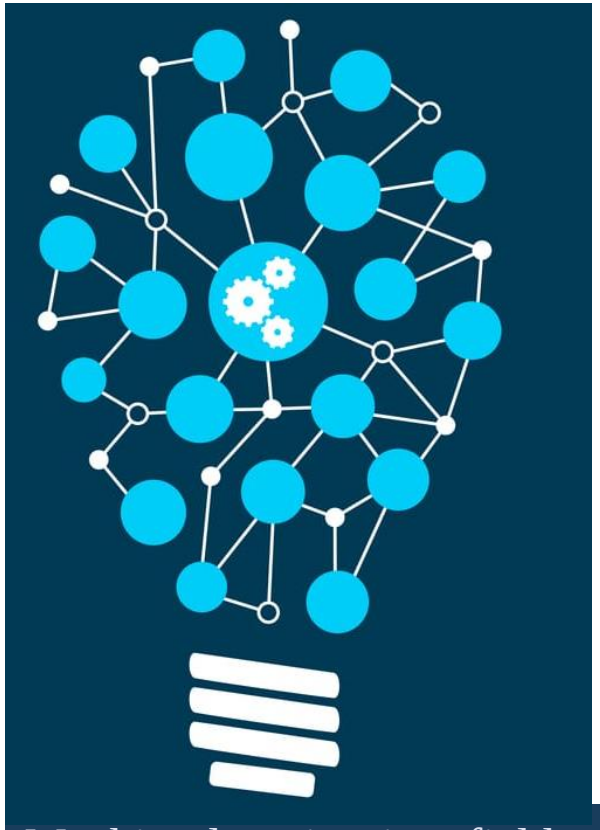


Data Science & Artificial Intelligence (AI)



Data science produces **insights**.
Machine learning produces **predictions**

Machine Learning



Machine learning is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can effectively generalize and thus perform tasks without explicit instructions.

Real-Time

facebook

If a member frequently "likes" a friend's posts, the news feed will automatically start showing more of that friend's activity, earlier in the feed.

Fine Art

Machine learning algorithms have helped reveal previously unrecognized influences between artists.

NETFLIX

Netflix predicts the ratings an individual will give a movie, which they haven't even watched yet, based on previous movie ratings made by them.

Anybody can ask a question → Anybody can answer → The best answers are voted up and rise to the top