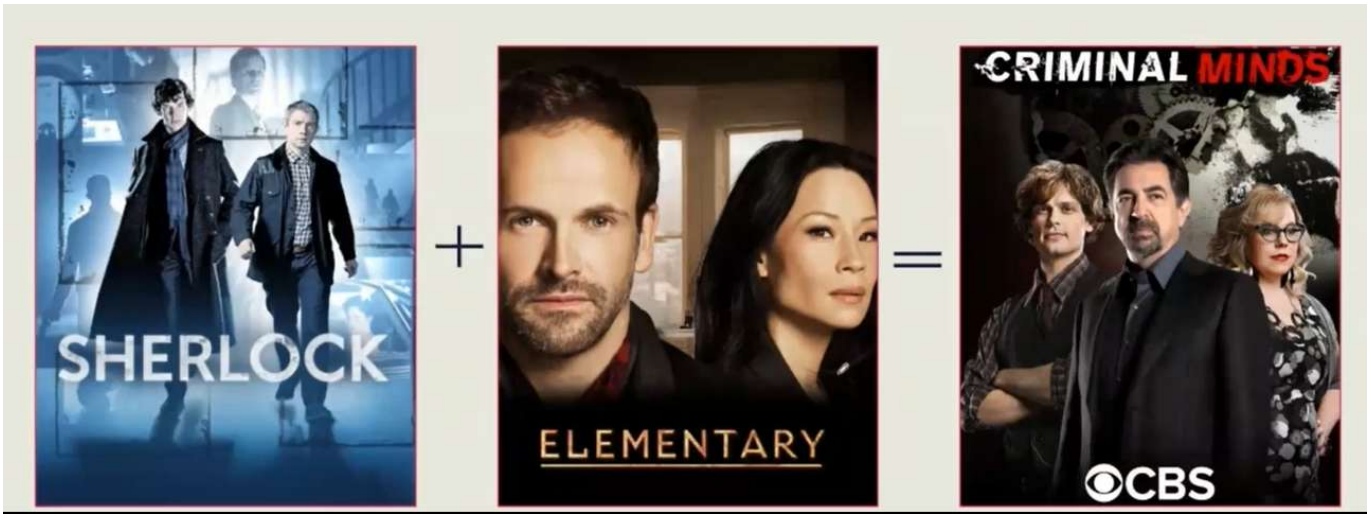
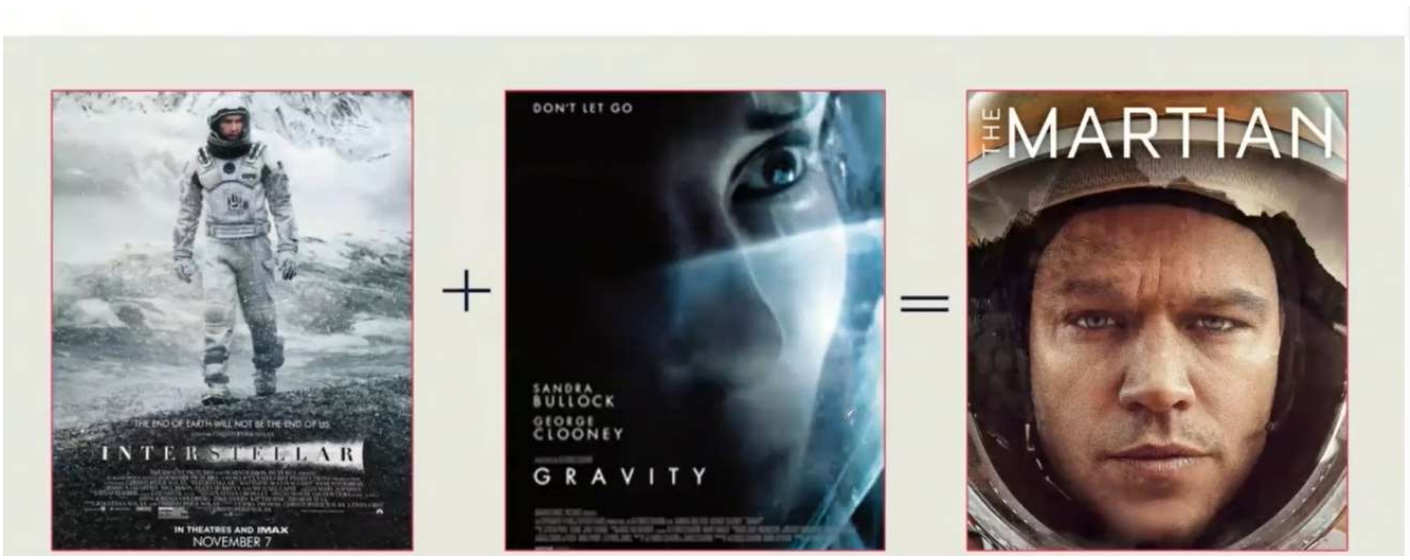
Learning

Tuesday, January 6, 2026 2:19 PM





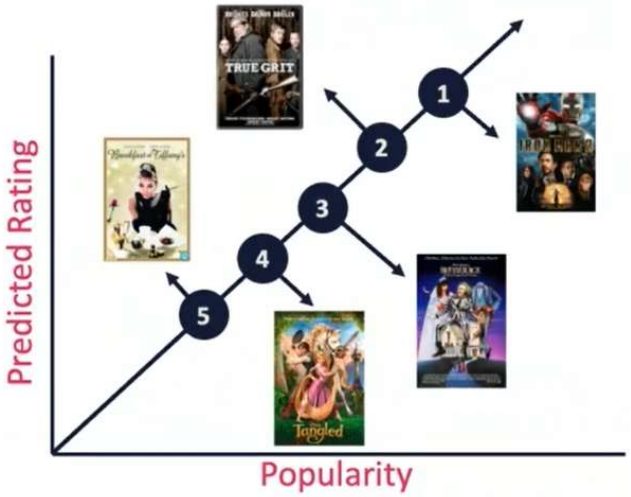
Netflix has adapted personalization algorithms in such a way that now 75% of what people watch is from some sort of recommendation.

The Top 10 rows of Netflix are their best guess at the ten titles you are most likely to enjoy. The recommendations are not only meant for the user, but also everyone in the user's **household**.

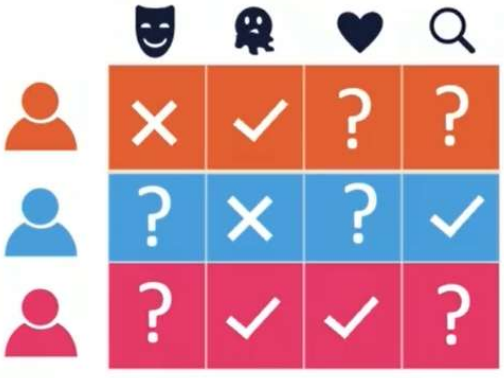
In Netflix, the poster of a movie recommendation changes based on a user's tastes and preferences.

A user who likes genre "Romance" will most likely watch Good Will Hunting with the lead actors featured in the poster; whereas a user who likes genre "Comedy", will most likely watch Good Will Hunting if a well-known comedy actor is featured in the poster.

A simple scoring approach for the ranking function could be a linear combination of popularity and predicted rating



Usually, recommendation engines consist of a utility matrix, which is a probability model that places a score on the relationship between a user and a movie type in order to predict their preferences.



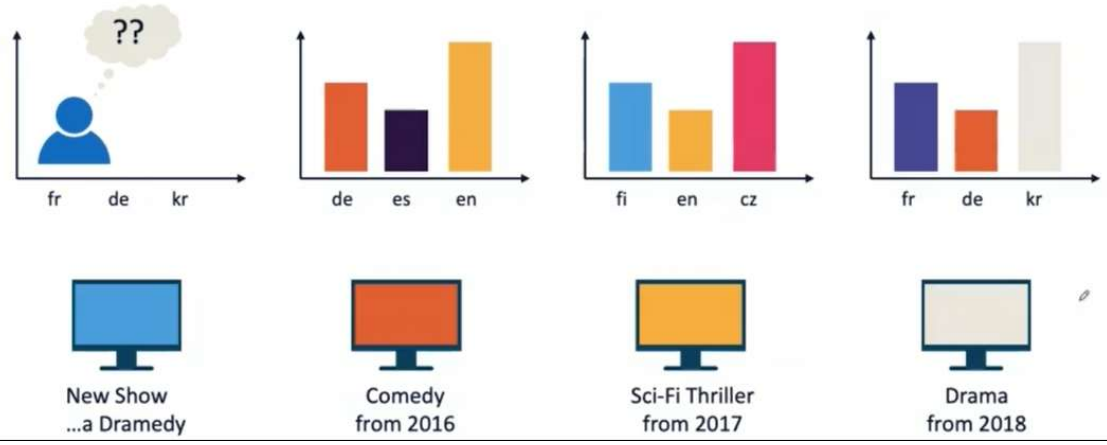
U_1 U_2

According to this matrix:

- User 1 would be given the recommendation of horror movies
- User 2 would be given the recommendation of lighter films with comedy and drama
- User 3 would be given the recommendation of more niche content like romantic thrillers

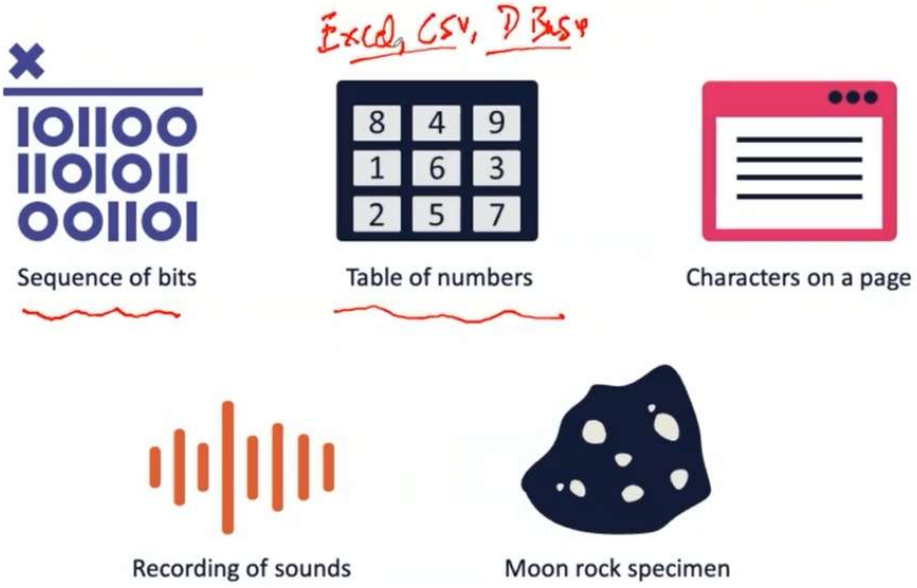
GLOBALLY?

Netflix predicts the per-language consumption for each show *k* months before it is released using historical viewing trends.



DATA!

WHAT IS DATA?



WHAT IS DATA?

Data ≠ AI

Sequence of bits

Excel, CSV, DBase

Table of numbers

FB, Twitter,

Characters on a page

Recording of sounds

Moon rock specimen

TYPES OF DATA

Structured data

Semi-structured data

Unstructured data

Graph data

TYPES OF DATA

AI

Approach 1

Structured data

Dbase, Excel, CSV

Semi-structured data

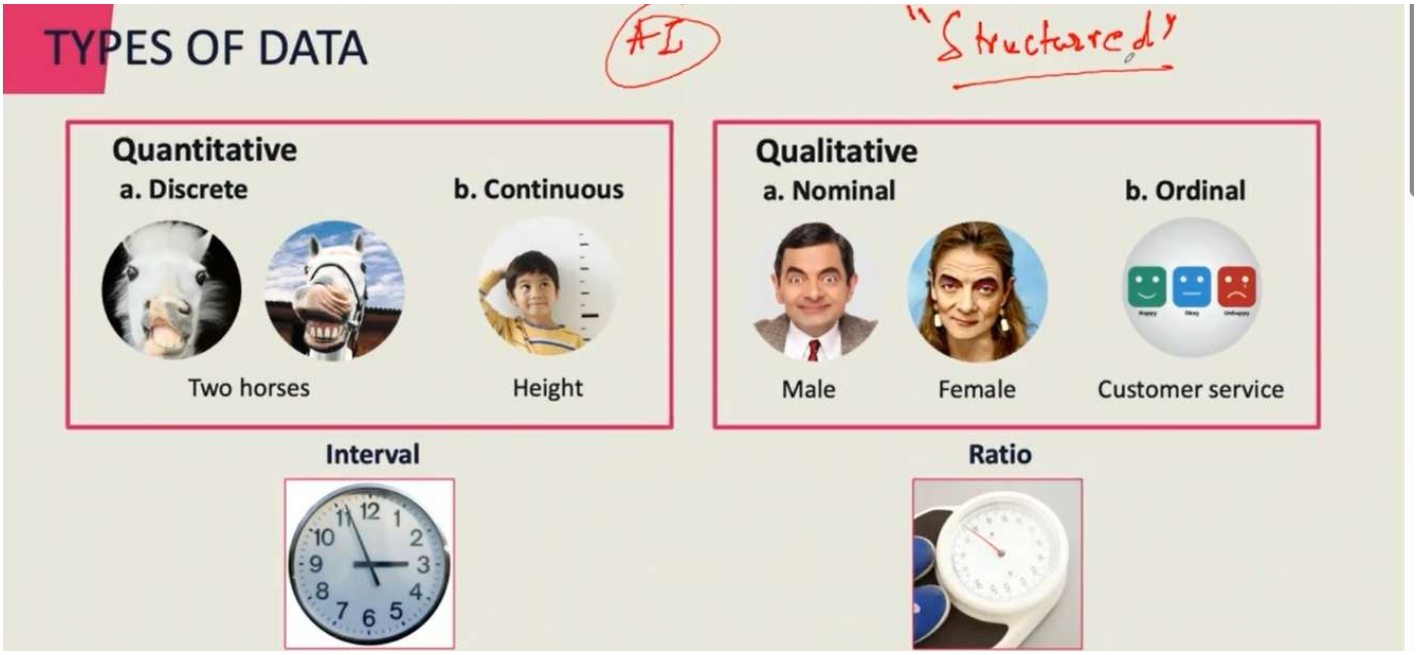
Log Files,

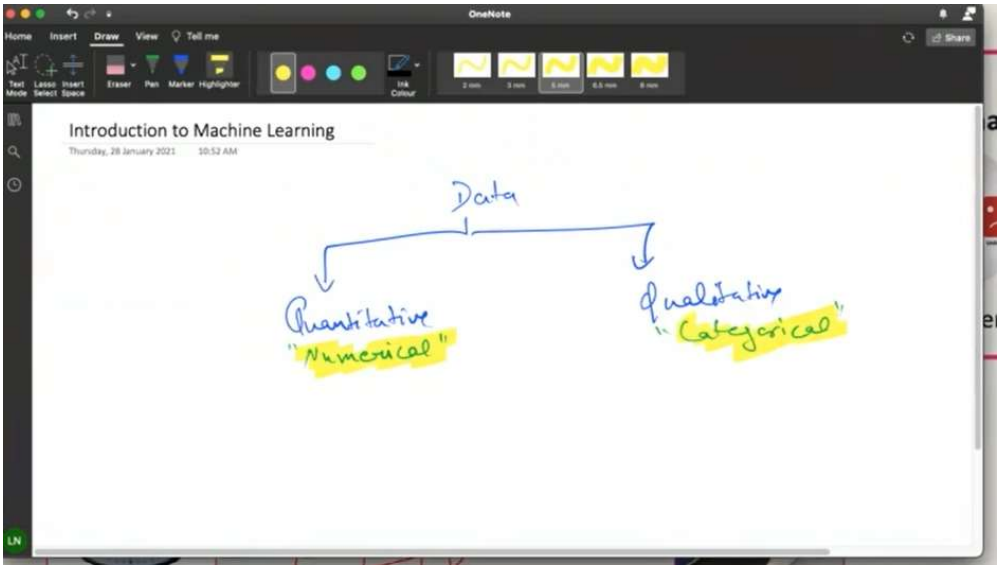
Approach 2

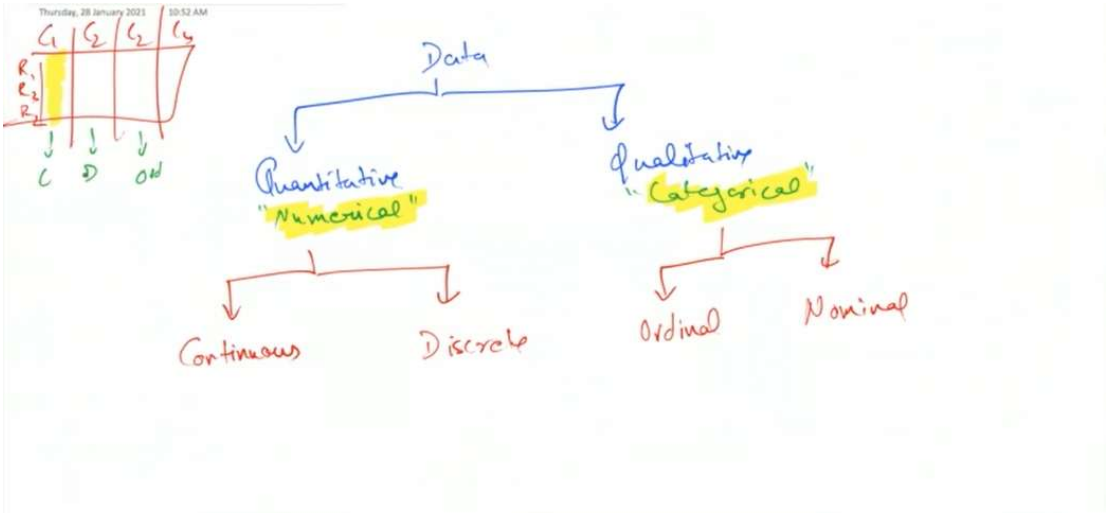
Unstructured data

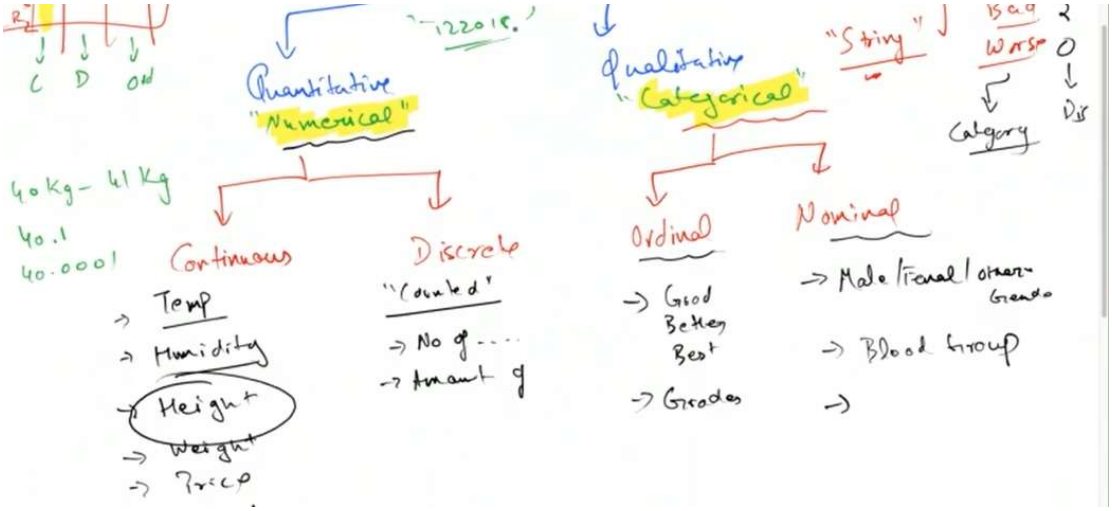
↳ Video, Audio, Images, Text

Graph data


TYPES OF DATA
AI
"Structured"
Quantitative
a. Discrete
b. Continuous
Two horses
Height
Qualitative
a. Nominal
b. Ordinal
Male
Female
Customer service
Interval
Ratio


OneNote
Introduction to Machine Learning
Data
Quantitative
"Numerical"
Qualitative
"Categorical"


Data
Quantitative
"Numerical"
Qualitative
"Categorical"
Continuous
Discrete
Ordinal
Nominal



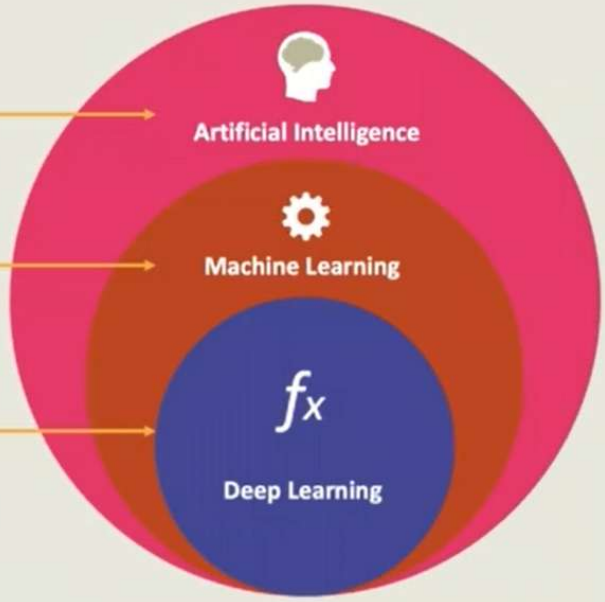
DATA SCIENCE, MACHINE LEARNING, DEEP LEARNING!

WHAT IS AI, ML, & DL?

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

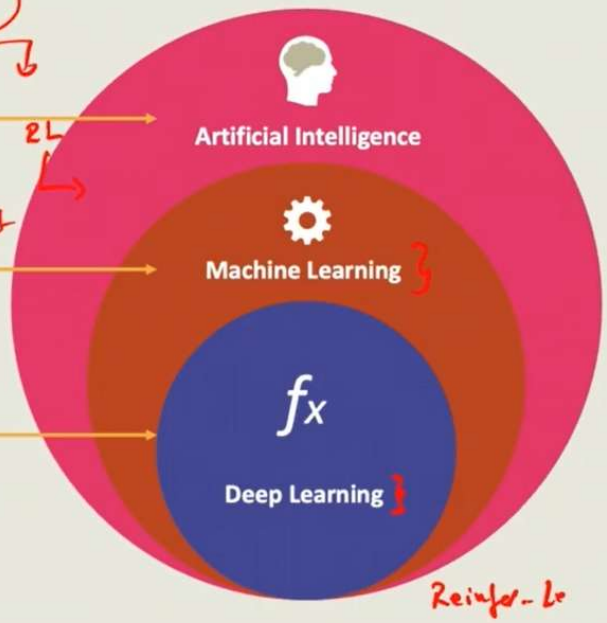


WHAT IS AI, ML, & DL?

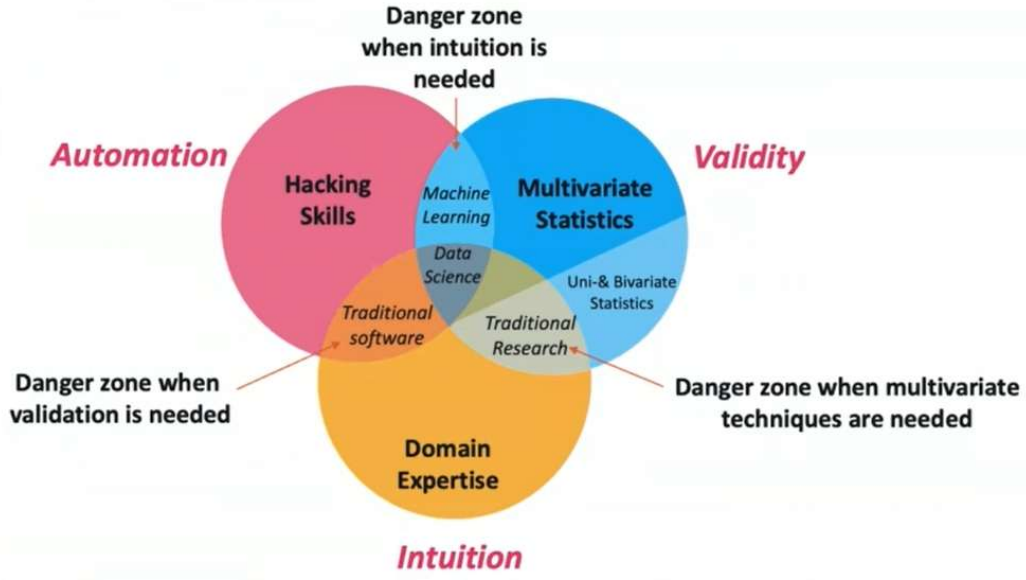
Artificial Intelligence
Any technique which enables computers to mimic human behavior.

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



WHAT IS DATA SCIENCE?



WHAT IS DATA SCIENCE?

Analytics/BI

Shows **WHAT** has happened

Data Science(EDA)

Shows **WHY** it happened

Machine Learning

Shows **HOW** we can predict & automate decision making systems

WHAT IS DATA SCIENCE?

Dashboards

Analytics/BI

{Shows **WHAT** has happened}

"Sherlock"

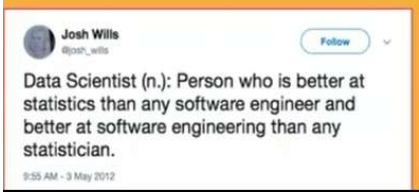
Data Science(EDA)

Shows **WHY** it happened

Machine Learning

Shows **HOW** we can predict & automate decision making systems

WHO IS DATA SCIENTIST?

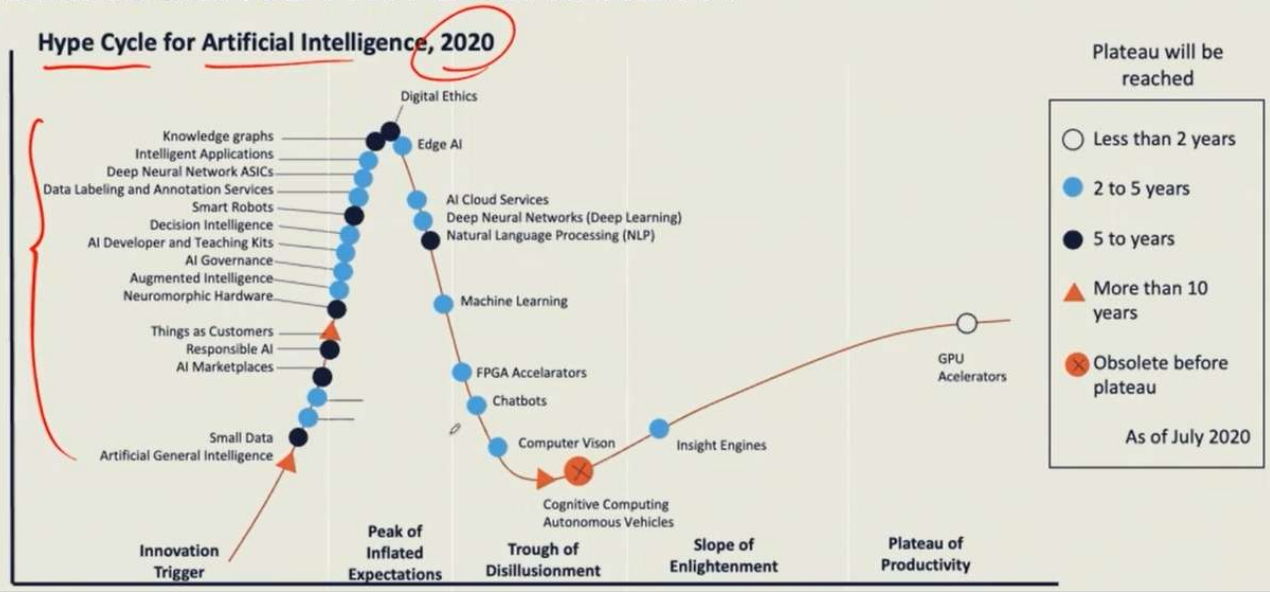


Responsibilities...

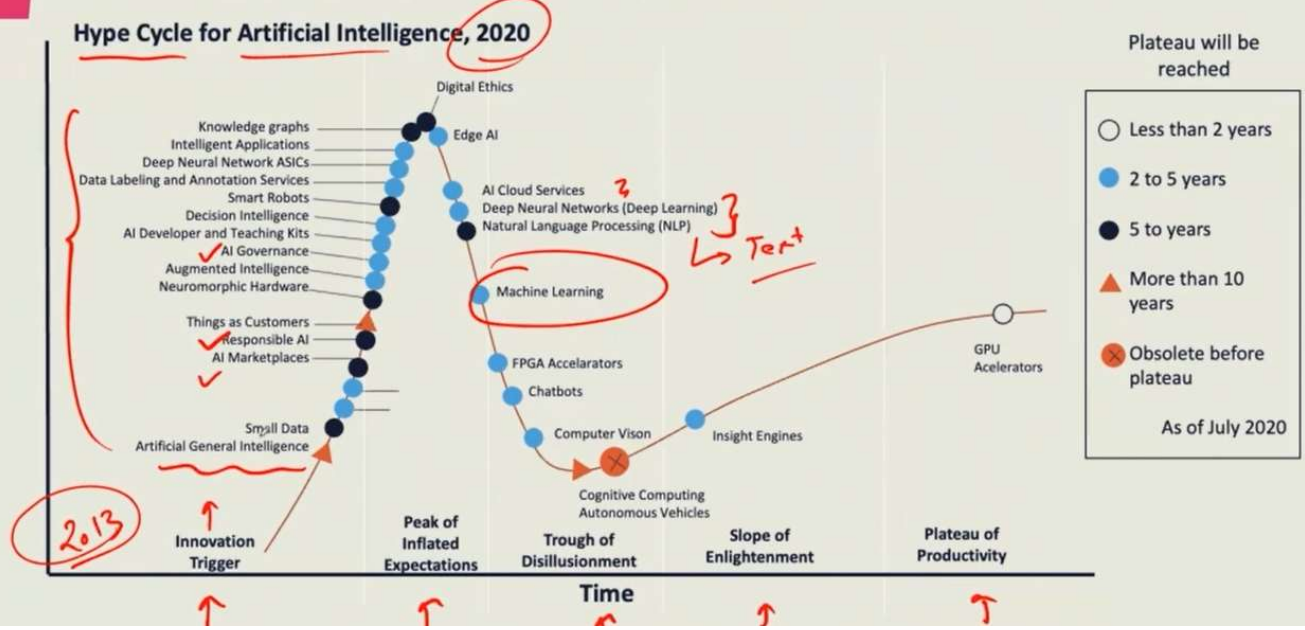
ON DAILY BASIS

- **Develop, construct, test, and maintain architectures** (such as databases and large-scale processing systems)
- **Ensure architecture** will support the requirements of the business
- Discover opportunities for **data acquisition**
- **Develop data set processes** for data modeling, mining and production
- Employ a variety of languages and tools (e.g. scripting languages) to **marry systems together**
- Conduct research to **answer industry and business questions**
- **Leverage large volumes of data** from internal and external sources to answer that business
- Employ sophisticated analytics programs, machine learning and statistical methods to **prepare data for use in predictive and prescriptive modeling**
- Explore and examine data to **find hidden patterns**
- **Automate work** through the use of predictive and prescriptive

DATA SCIENCE : HYPE OR REALITY?



DATA SCIENCE : HYPE OR REALITY?



VARIOUS PROBLEMS DATA SCIENCE CAN SOLVE!

PROBLEMS THAT DATA SCIENTISTS SOLVE

- Is this A or B?
- Is this weird?
- How much or how many?
- How is this organized?
- What should I do now?

PROBLEMS THAT DATA SCIENTISTS SOLVE

IS THIS A or B?

CLASSSIFICATION

Binary

Multi label

Classification

Data Type

Categ

WILL THIS APPLICANT BE ABLE TO REPAY THE LOAN?

IDENTIFY A HUMAN & A DOG?