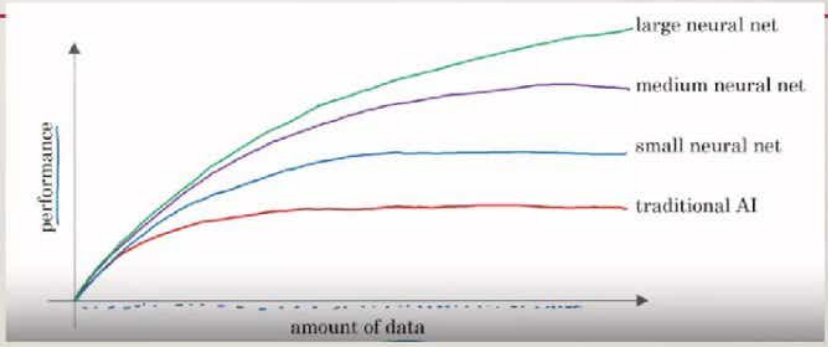
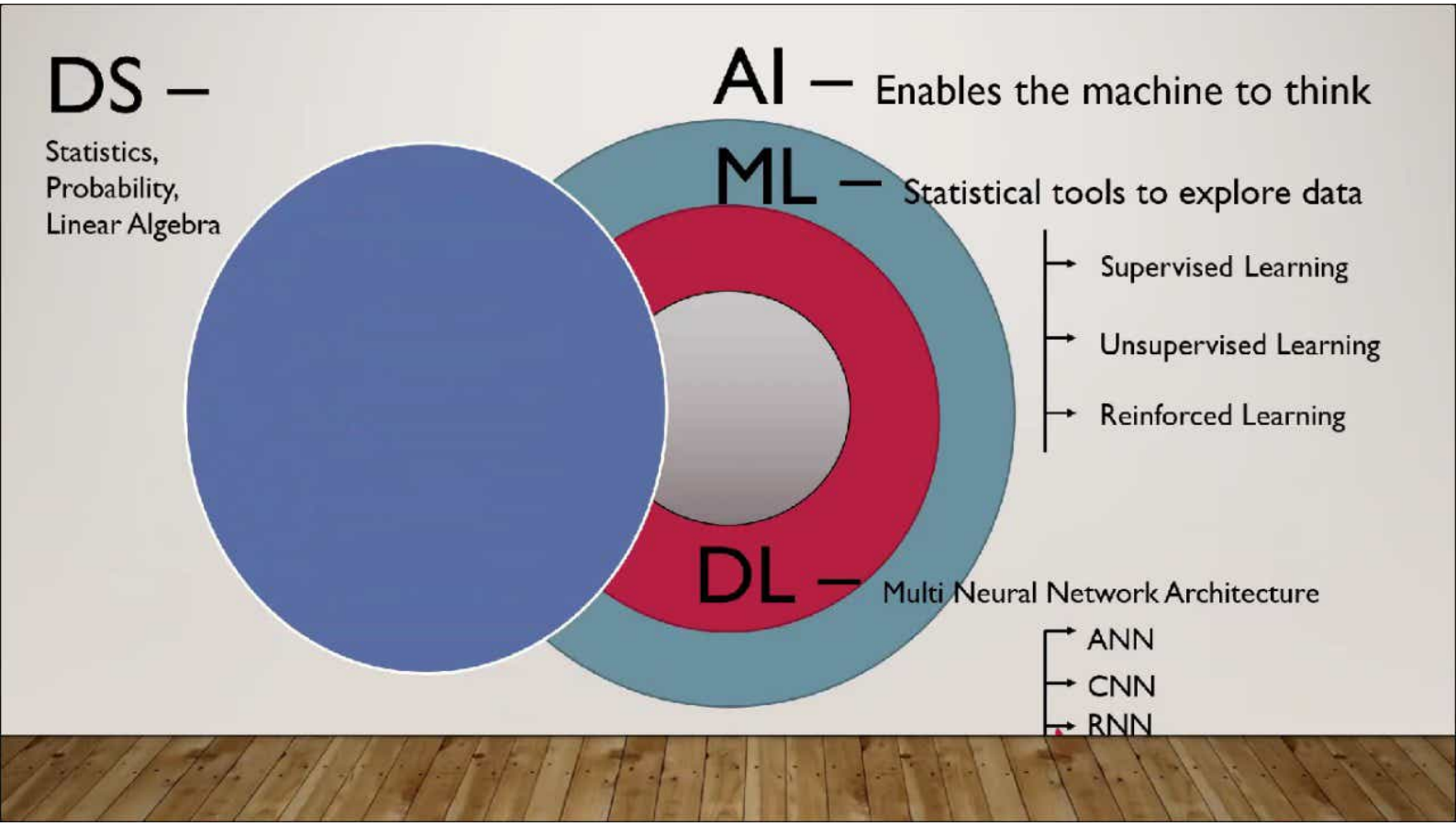
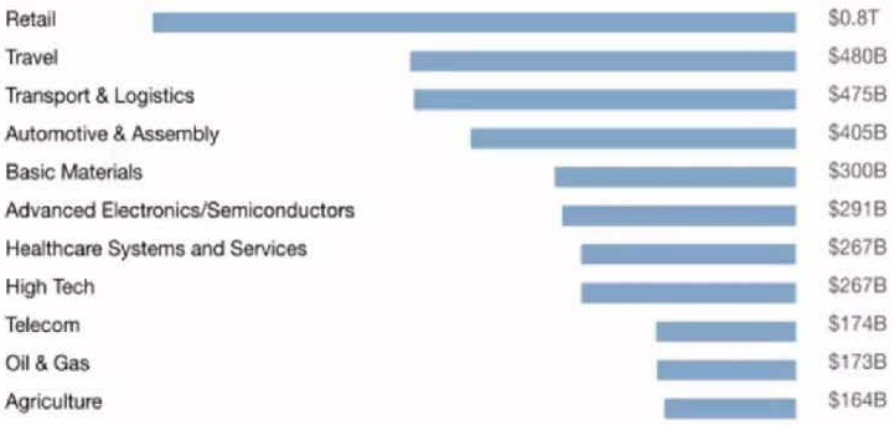
WHY NOW??? BIG DATA



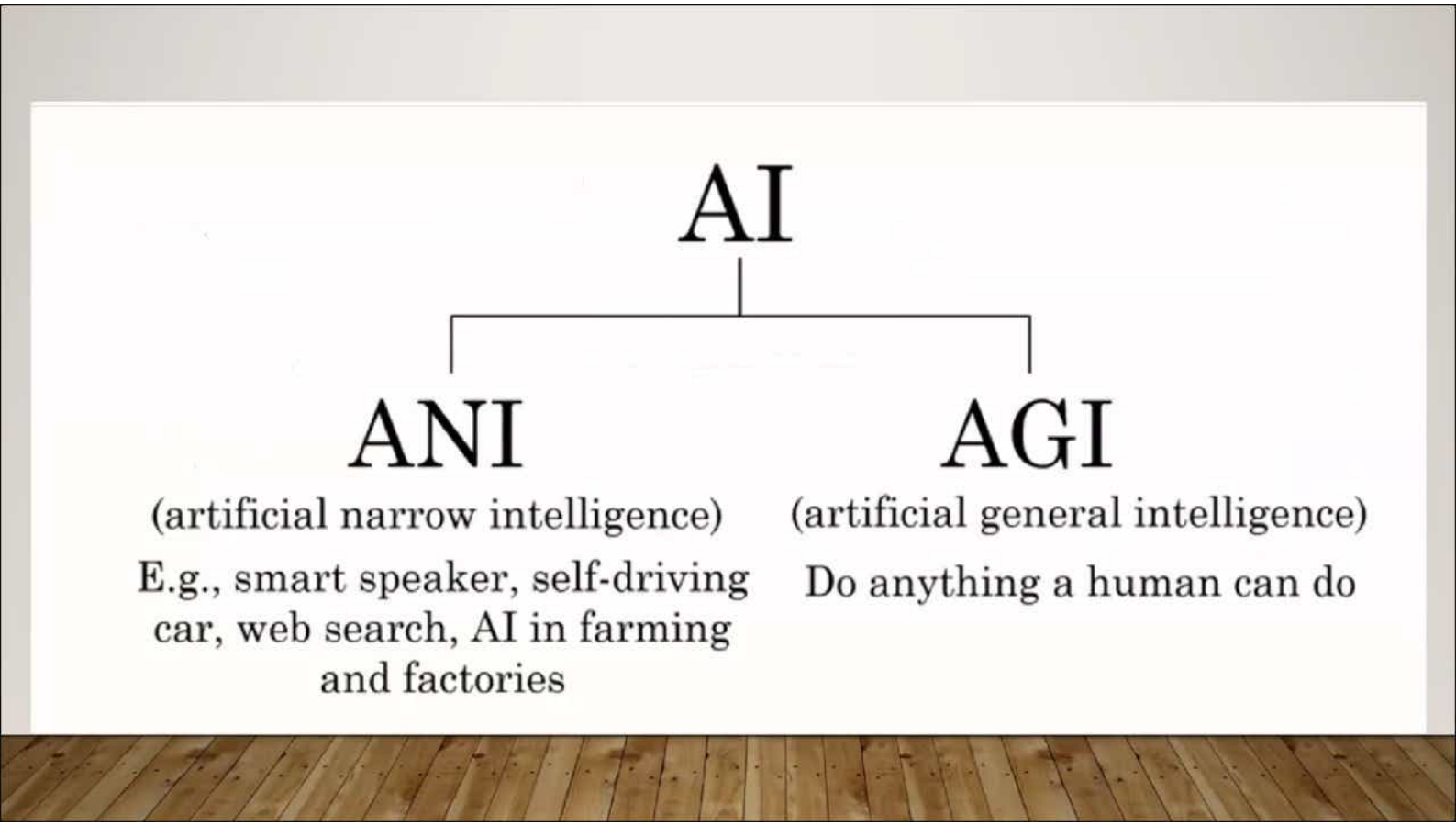

DS –
Statistics,
Probability,
Linear Algebra
AI – Enables the machine to think
ML – Statistical tools to explore data
Supervised Learning
Unsupervised Learning
Reinforced Learning
DL – Multi Neural Network Architecture
ANN
CNN
RNN

AI value creation by 2030

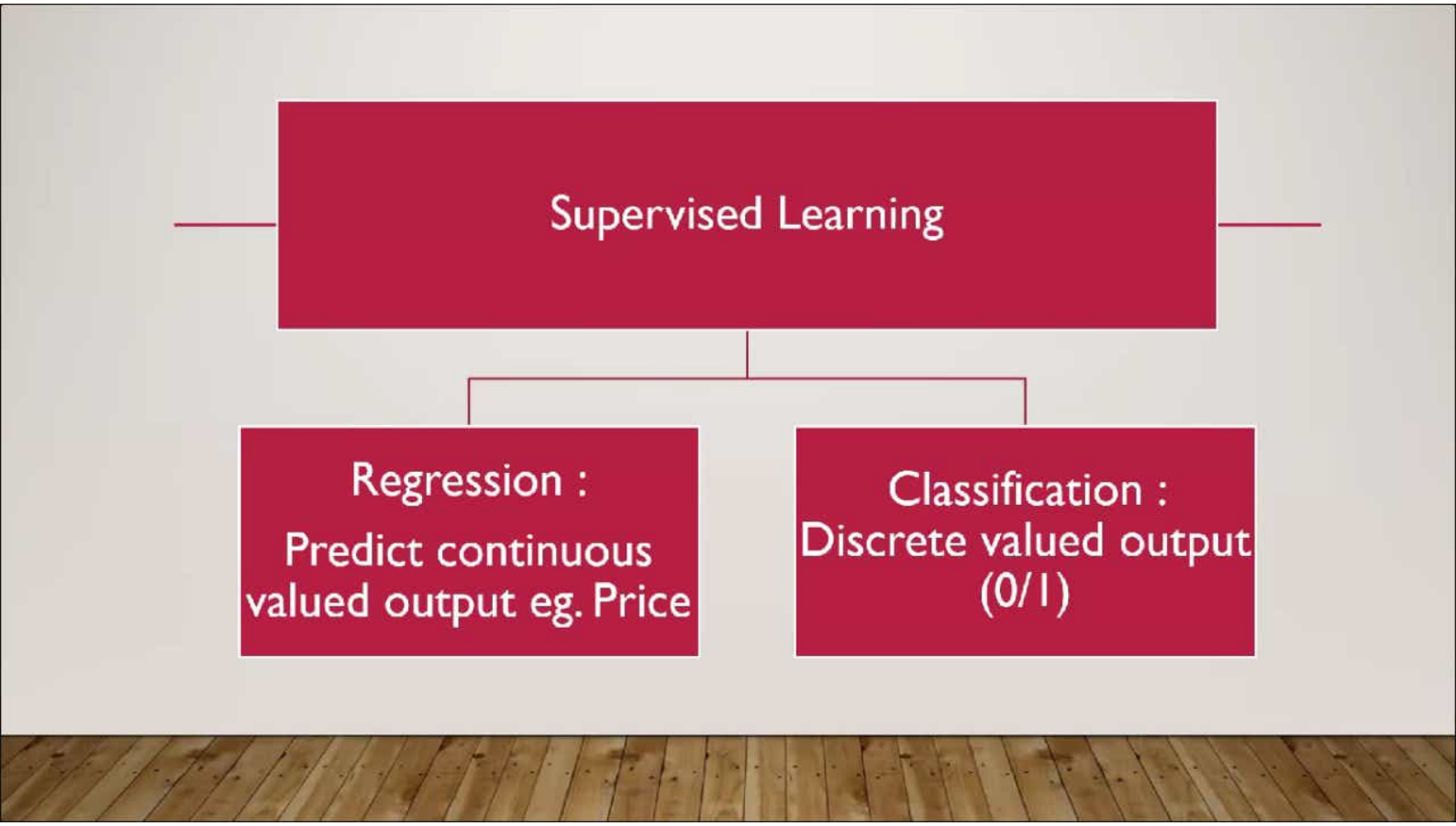
$13 trillion



[Source: McKinsey Global Institute.]

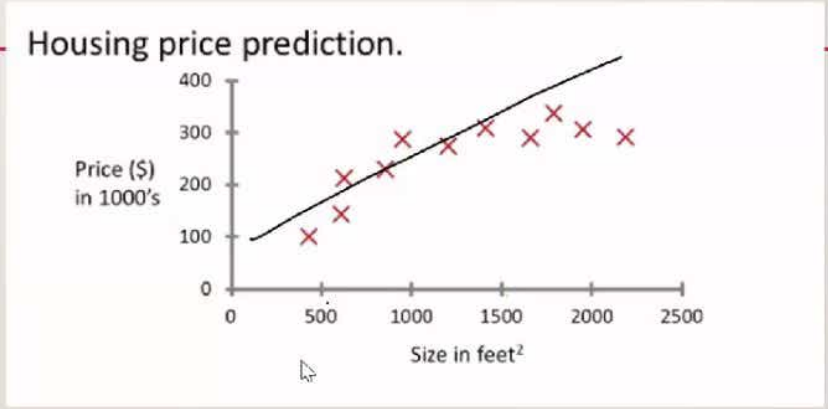

AI
ANI
(artificial narrow intelligence)
E.g., smart speaker, self-driving car, web search, AI in farming and factories
AGI
(artificial general intelligence)
Do anything a human can do

SUPERVISED LEARNING


Supervised Learning
Regression :
Predict continuous valued output eg. Price
Classification :
Discrete valued output (0/1)

REGRESSION

HOUSING PREDICTION



QUESTION : CLASSIFY THE PROBLEMS AS CLASSIFICATION OR REGRESSION?

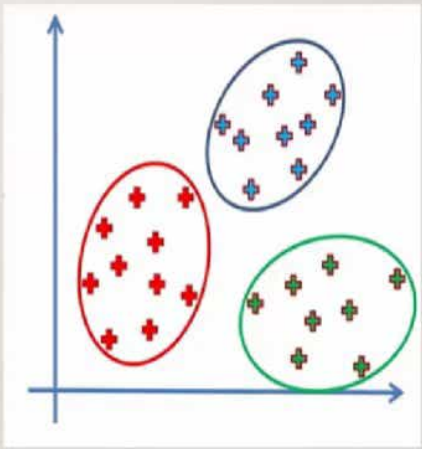
- 1. How many tshirts can be sold in next 3 months.
- 2. To decide if the email is spam or not
- 3. To detect if cup is broken or not
- 4. To determine the price of gold on certain date

STEPS

- 1. Gather input dataset
- 2. Train the model
- 3. Predict the result

CLUSTERING

- Clustering is similar to classification, but the basis is different.
- In Clustering we don't know what we are looking for, and we are trying to identify some segments or clusters in our data.
- When we use clustering algorithms on your dataset, unexpected things can suddenly pop up like structures, clusters and groupings we would have never thought of otherwise.



OBJECTIVE OF CLUSTERING

- Task ----→ Group similar points in one cluster

1. Points in one cluster are close together
2. Points in different clusters are far away

APPLICATIONS OF CLUSTERING

- https://en.wikipedia.org/wiki/Cluster_analysis#Applications
- Ecommerce -> To 'gp' similar customers
- Where manual labelling is time consuming, perform clustering as pre processing step and then do manual labelling

Aim : To predict the marks of student depending upon no of hours studied, IQ

Feature
Input
Independent
variable

Label
Output
Dependent
variable

# hrs	IQ	Marks
2	7	14
3	10	20
4	5	12