AI IN EVERYDAY LIFE

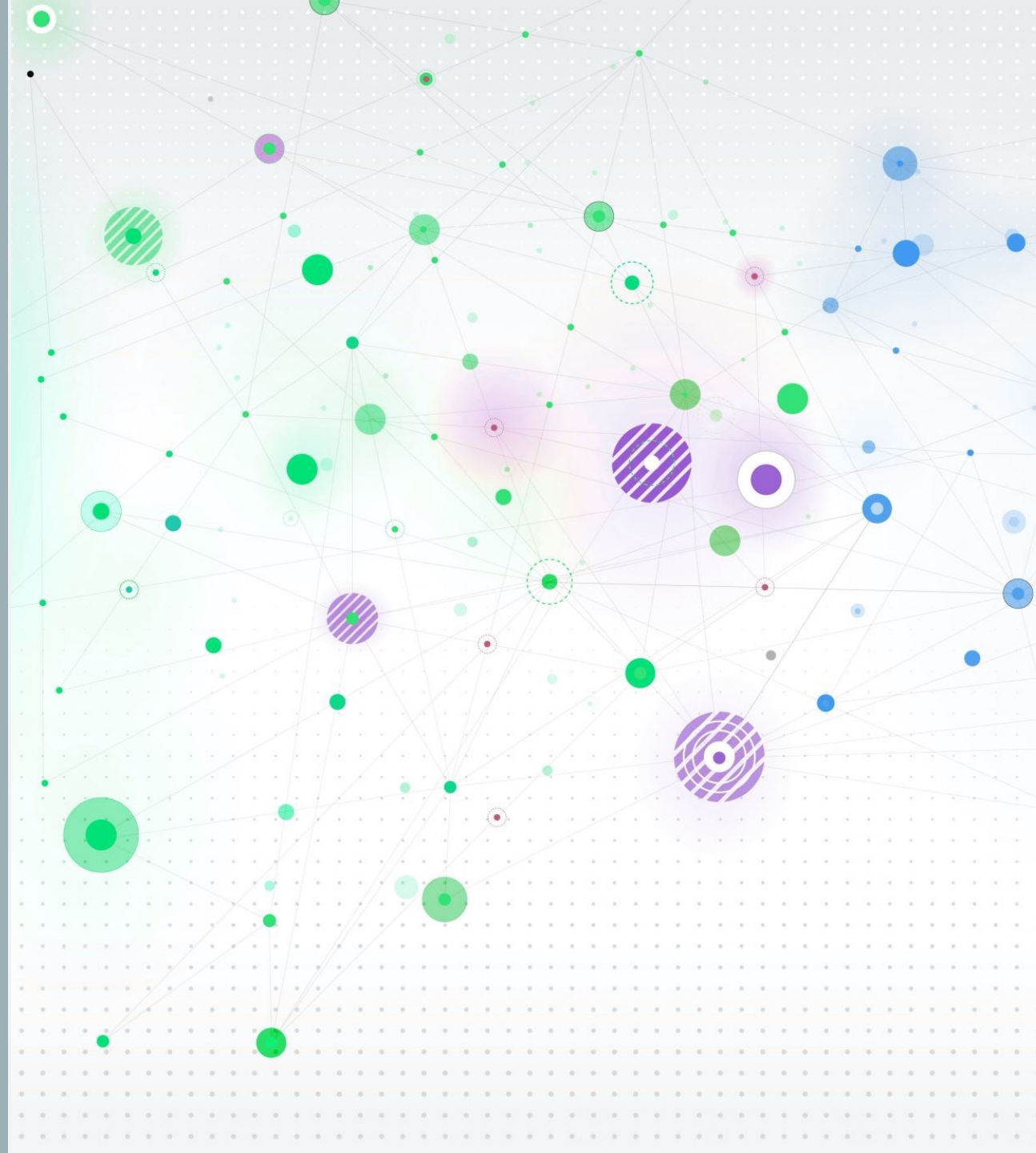Unit 3 – Machine Learning & Big Data

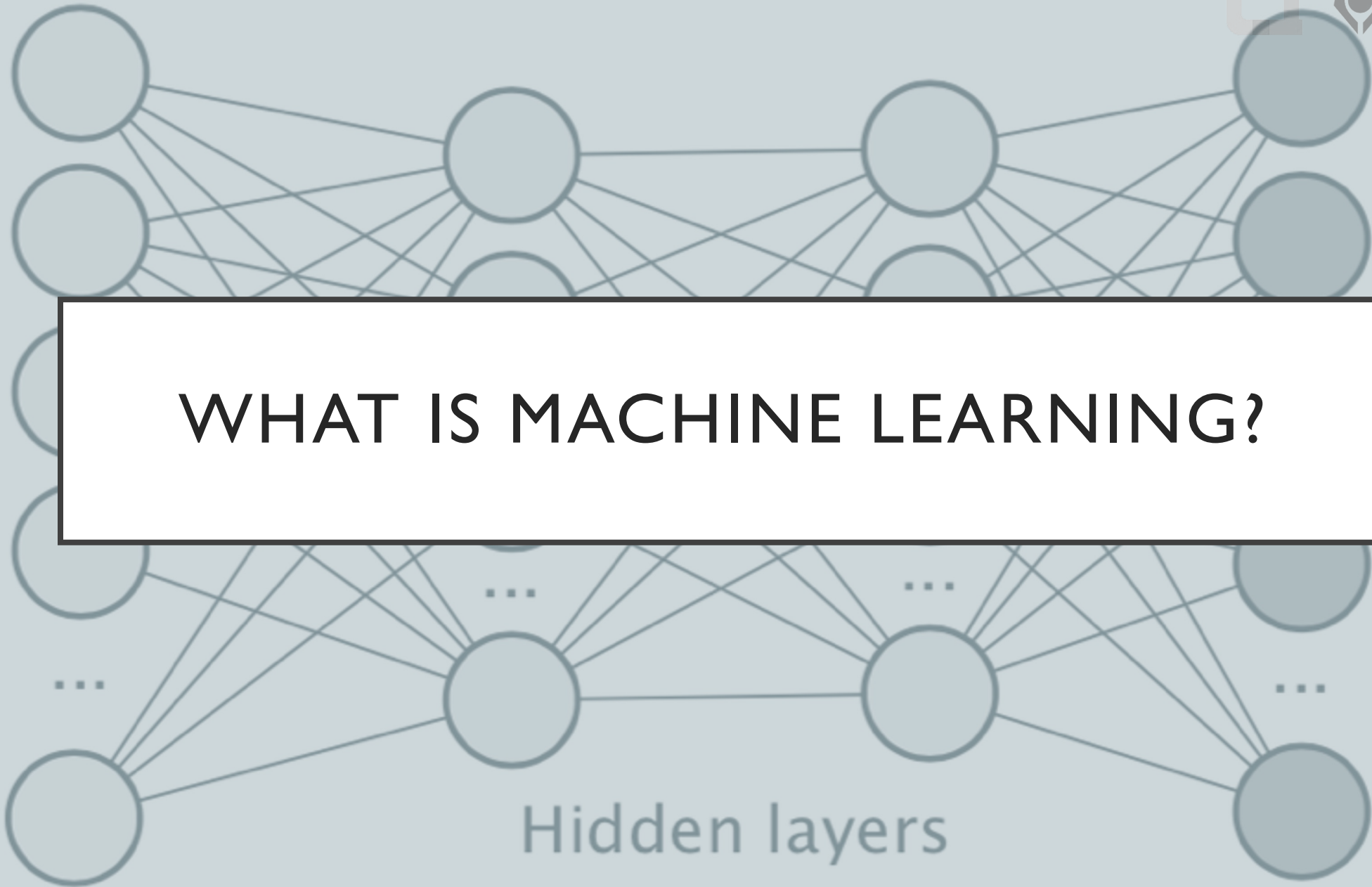# UNIT 3 – MACHINE LEARNING (& BIG DATA)

A brief history of Machine Learning

What is big about
Big Data?

How does machine learning use (big) data?

How does all of this relate to today's AI?

WHAT IS MACHINE LEARNING?

Products ∨   Solutions ∨   Consulting ∨   Support ∨   More ∨

What is machine learning?

This introduction to machine learning provides an overview of its history, important definitions, applications, and concerns within businesses today

Discover Watson Assistant →

What is machine learning?

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

IBM has a rich history with machine learning. One of its own, Arthur Samuel, is credited for coining the term, "machine learning" with his research (PDF, 481 KB) (link resides outside IBM) around the game of checkers. Robert Nealey, the self-proclaimed checkers master, played the game on an IBM 7094 computer in 1962, and he lost to the computer. Compared to what can be done today, this feat seems trivial, but it's considered a major milestone in the field of artificial intelligence.

Over the last couple of decades, the technological advances in storage and processing power have enabled some innovative products based on machine learning, such as Netflix's recommendation engine and self-driving cars.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights
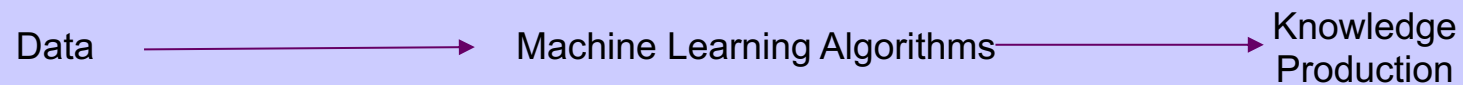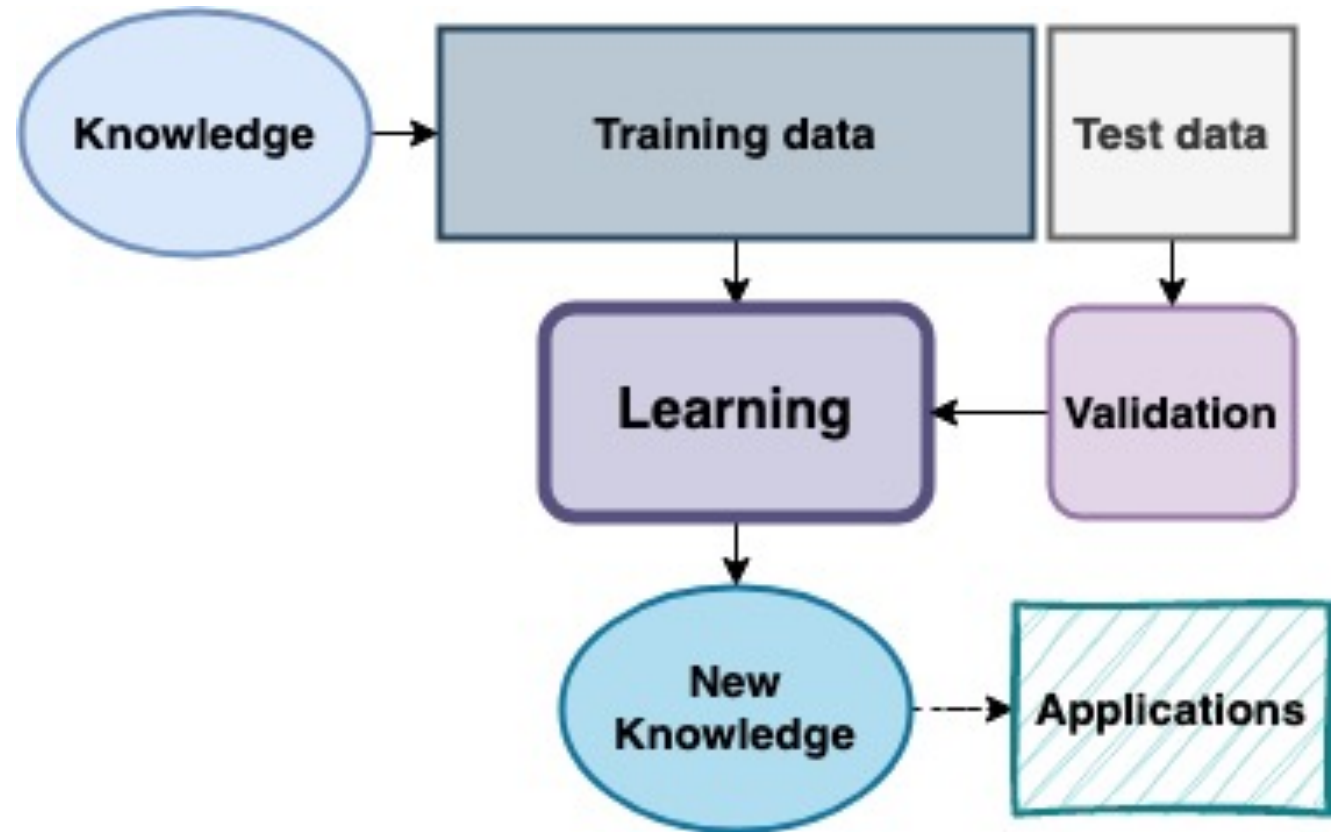
USEFUL
RESOURCE

WHAT IS MACHINE LEARNING?

Concerns the study and development of algorithms that can **learn from data and draw some conclusions.**

Types of conclusions include classification and prediction.

Data ⟶ Machine Learning Algorithms ⟶ Knowledge Production

PHASES OF MACHINE LEARNING

AI, MACHINE LEARNING, DEEP LEARNING
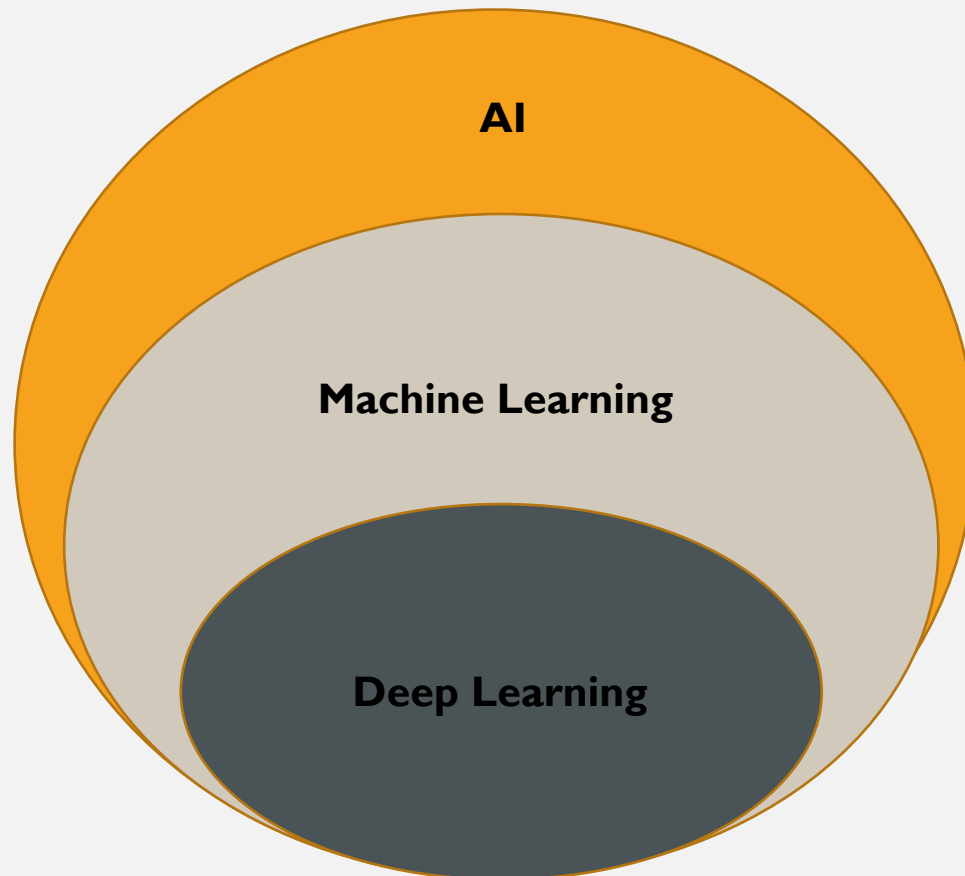
AI — Techniques to enable machines to have "intelligent" behaviors

Machine Learning — A subset of AI techniques that use data and statistics to help machines "learn" experience

Deep Learning — A subset of Machine Learning techniques that make computations using a multi-layer neural network

AI

Machine Learning

Deep Learning

A BRIEF HISTORY

Data → **Processing** → **Output**

Input layer — Hidden layers — Output layer
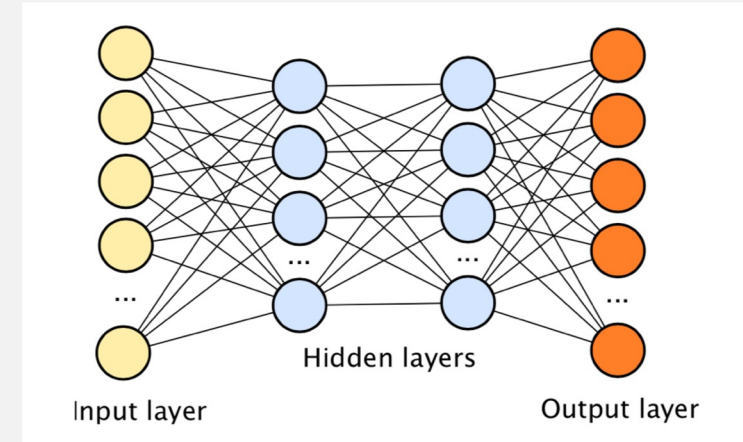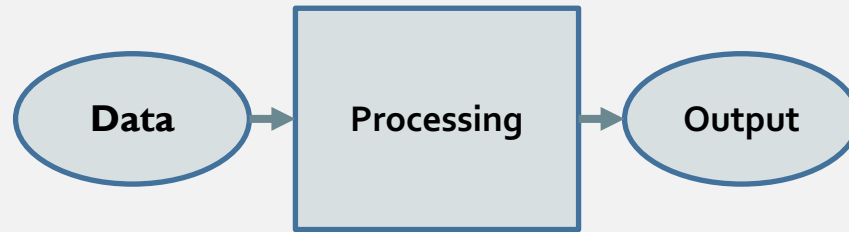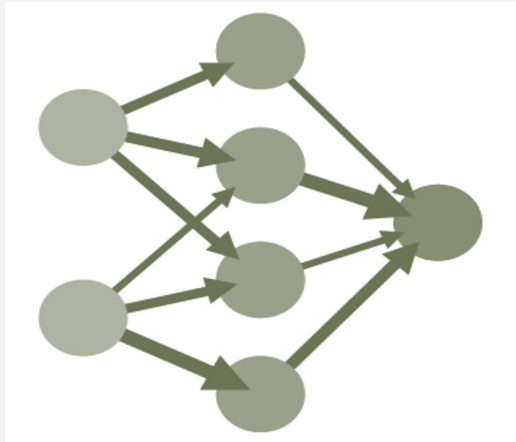
1950s–1970s	**1980s–2010s**	**Today**
Neural networks	**Machine Learning**	**Deep Learning**
First machines that can "think"	The rise of machine learning	Explosion of AI applications
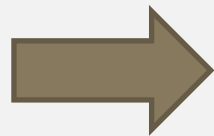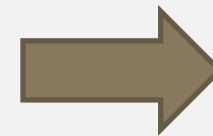
WHAT ABOUT BIG DATA?

What happens online in 60 seconds? (2012 - 2014)

BIG (OPEN) DATA

Source: https://www.anishsneh.com/2014/07/big-data-volume-velocity-and-variety.html
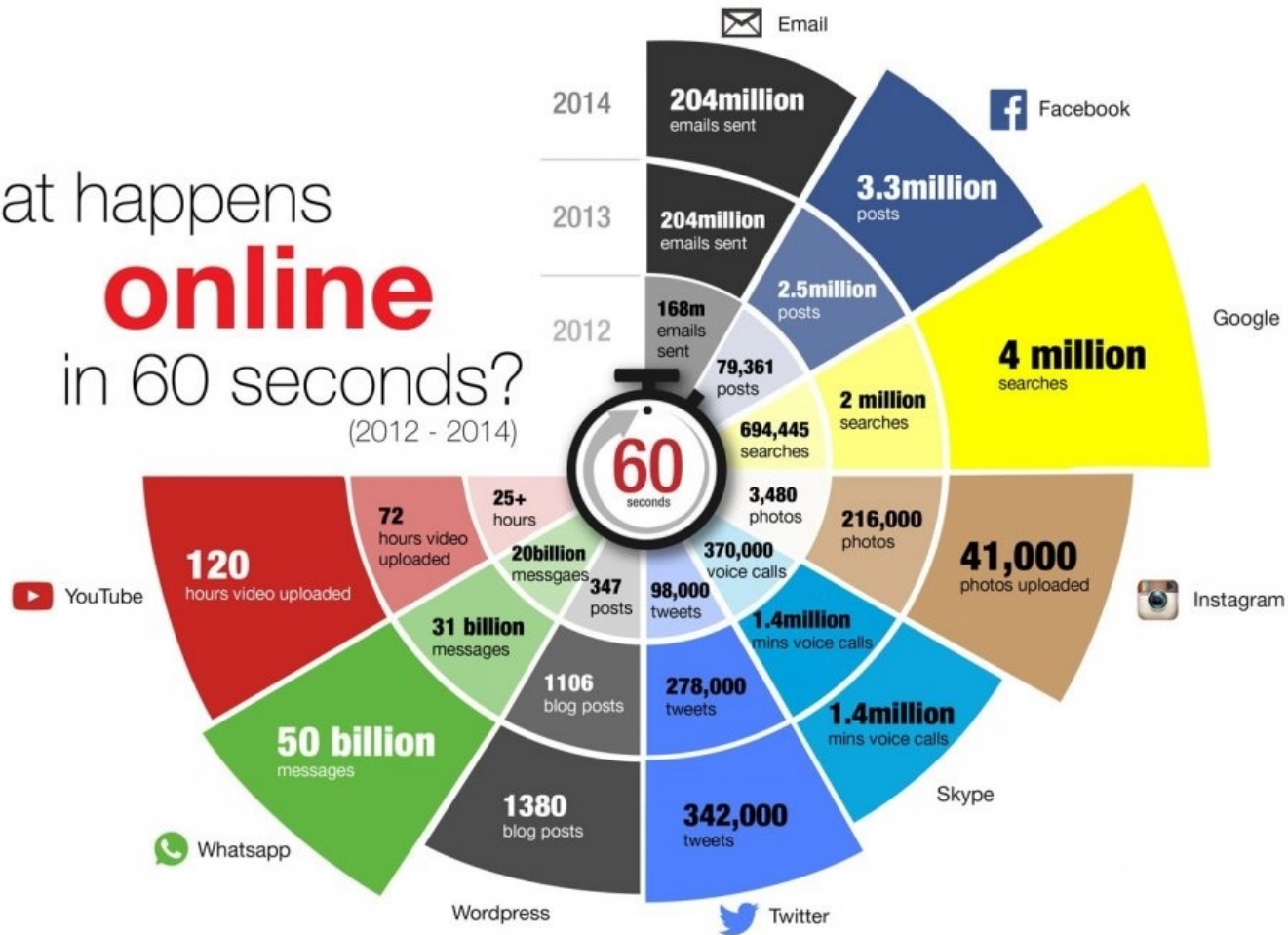
BIG DATA AND AI

- Science and technology have always been shaped by social and political factors
  - Big Data
  - New business models
- The new paradigm of AI – based on Big Data – is the first that is technically and economically feasible.

Source: Nello Cristianini. 2019. Shortcuts to artificial intelligence. In: Machines We Trust. Cambridge, MA: MIT Press.

HOW DOES MACHINE LEARNING USE DATA?

MACHINE LEARNING TECHNIQUES

- AI can adapt using *learning algorithms*, which allow the data to be its "guide"

- The algorithms find structure and regularities (i.e., patterns) in the data, in order to develop a "skill,"
e.g.,

  - <u>Prediction</u>: Predict the test score of a student

  - <u>Classification</u>: Distinguish between texts that have positive v. negative sentiment

MACHINE LEARNING TECHNIQUES

- Three main categories:
  - Supervised
  - Unsupervised
  - Semi-supervised

SUPERVISED MACHINE LEARNING

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Figure A: CLASSIFICATION

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Figure B: REGRESSION

TYPES OF SUPERVISED LEARNING

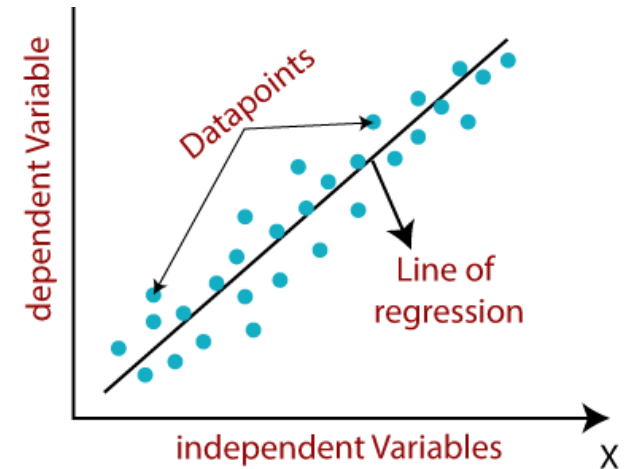| Algorithm / technique	Type of problem	Type of response variable
Linear Regression	Prediction	Continuous
Logistic Regression	Classification	Discrete
Decision Trees	Classification	Discrete
K-nearest Neighbor	Classification	Discrete
Support Vector Machines	Classification or Prediction	Discrete or Continuous
Neural Networks	Classification or Prediction	Discrete or Continuous

DECISION TREES

Figure A: CLASSIFICATION

LINEAR REGRESSION

- $Wind\ Speed = Temperature * b_1 + Pressure * b_2 + Humidity * b_3 + Wind\ Direction * b_4$

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Figure B: REGRESSION

UNSUPERVISED LEARNING

User ID	Gender	Age	Salary
15624510	Male	19	19000
15810944	Male	35	20000
15668575	Female	26	43000
15603246	Female	27	57000
15804002	Male	19	76000
15728773	Male	27	58000
15598044	Female	27	84000
15694829	Female	32	150000
15600575	Male	25	33000
15727311	Female	35	65000
15570769	Female	26	80000
15606274	Female	26	52000
15746139	Male	20	86000
15704987	Male	32	18000
15628972	Male	18	82000
15697686	Male	29	80000
15733883	Male	47	25000

Temperature	Pressure	Relative Humidity	Wind Direction
10.69261758	986.882019	54.19337313	195.7150879
13.59184184	987.8729248	48.0648859	189.2951202
17.70494885	988.1119385	39.11965597	192.9273834
20.95430404	987.8500366	30.66273218	202.0752869
22.9278274	987.2833862	26.06723423	210.6589203
24.04233986	986.2907104	23.46918024	221.1188507
24.41475295	985.2338867	22.25082295	233.7911987
23.93361956	984.8914795	22.35178837	244.3504333
22.68800023	984.8461304	23.7538641	253.0864716
20.56425726	984.8380737	27.07867944	264.5071106
17.76400389	985.4262085	33.54900114	280.7827454
11.25680746	988.9386597	53.74139903	68.15406036
14.37810685	989.6819458	40.70884681	72.62069702
18.45114201	990.2960205	30.85038484	71.70604706
22.54895853	989.9562988	22.81738811	44.66042709
24.23155922	988.796875	19.74790765	318.3214111

CUSTOMER SEGMENTATION

LINKS AND CONTACTS

https://datascientiafoundation.github.io/datascientia-education-eai-2023-24-unitn

http://knowdive.disi.unitn.it/

@knowdive

matteo.busso@unitn.it

THANK YOU!