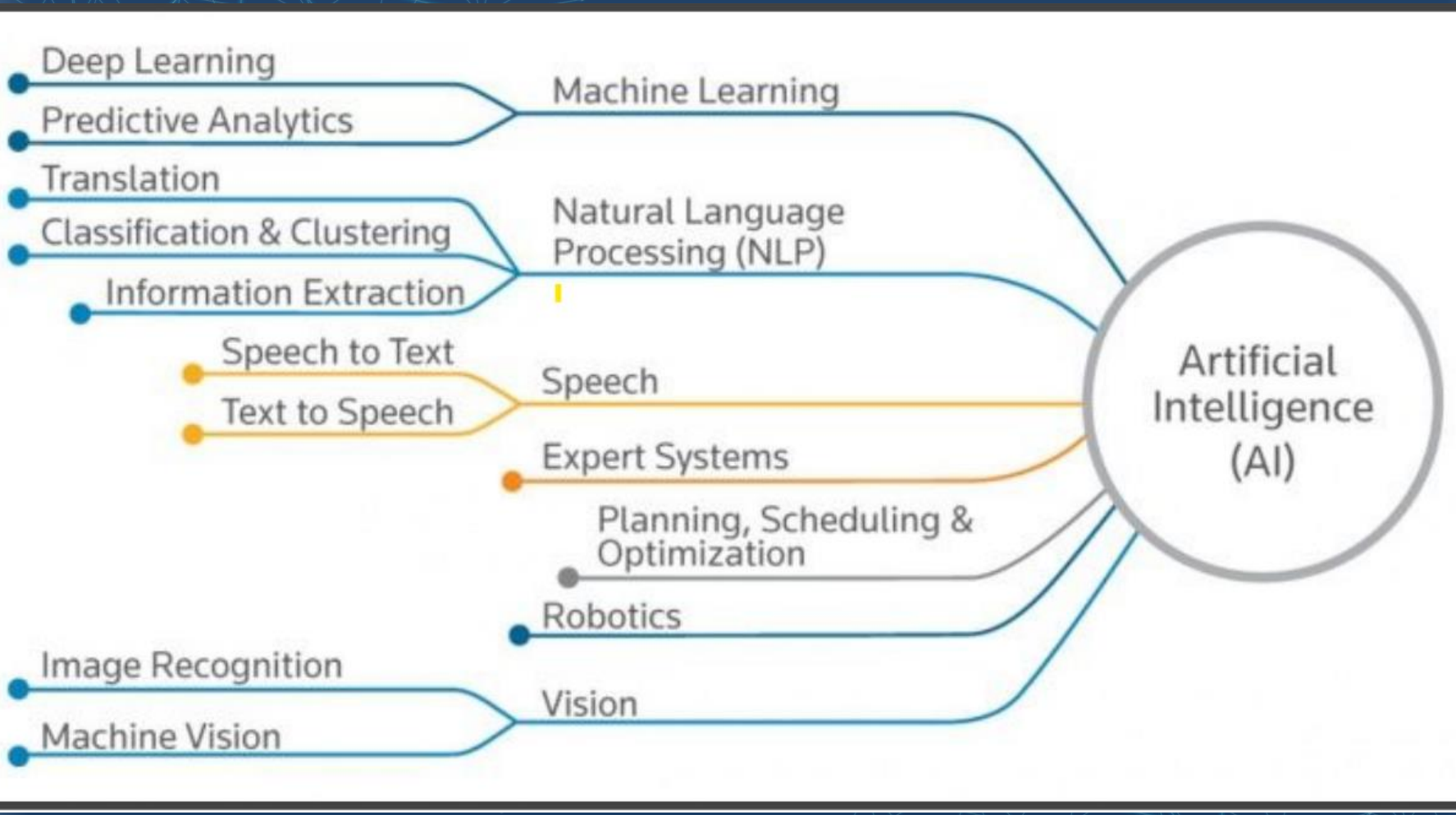
Fundamentals of AI & Machine Learning

Eng- Mohamed Khaled Idris
Eng- Mayar Swilam


Deep Learning
Predictive Analytics
Machine Learning
Translation
Classification & Clustering
Information Extraction
Natural Language Processing (NLP)
Speech to Text
Text to Speech
Speech
Expert Systems
Planning, Scheduling & Optimization
Robotics
Image Recognition
Machine Vision
Vision
Artificial Intelligence (AI)

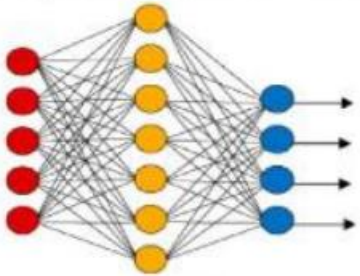
What is Machine Learning ?

Machine learning is a field of artificial intelligence (AI) that involves developing algorithms and statistical models that enable computers to learn and improve from experience, without being explicitly programmed.

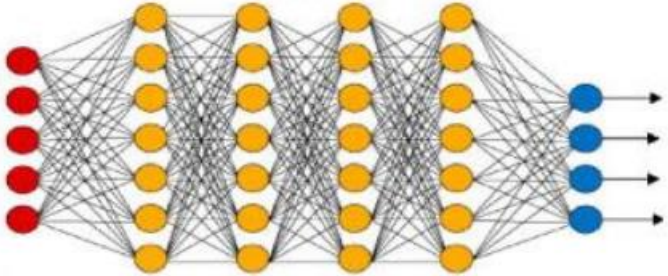
TRADITIONAL PROGRAMMING

MACHINE LEARNING

Simple Neural Network



Deep Learning Neural Network



Input Layer Hidden Layer Output Layer

Types of Machine Learning

Machine Learning

Supervised Learning

Task Driven
(Classification/Regression)

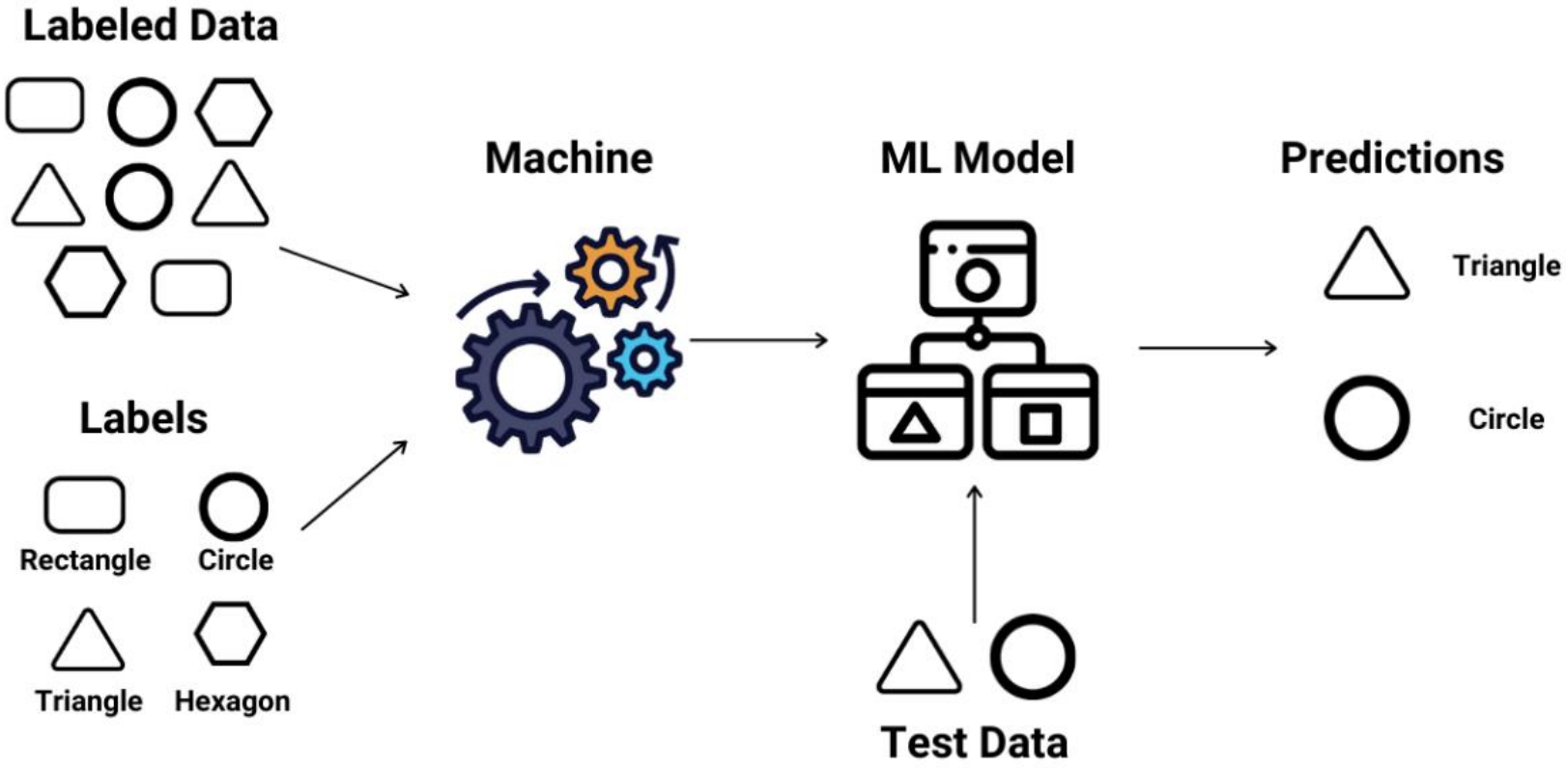
Unsupervised Learning

Data Driven
(Clustering)

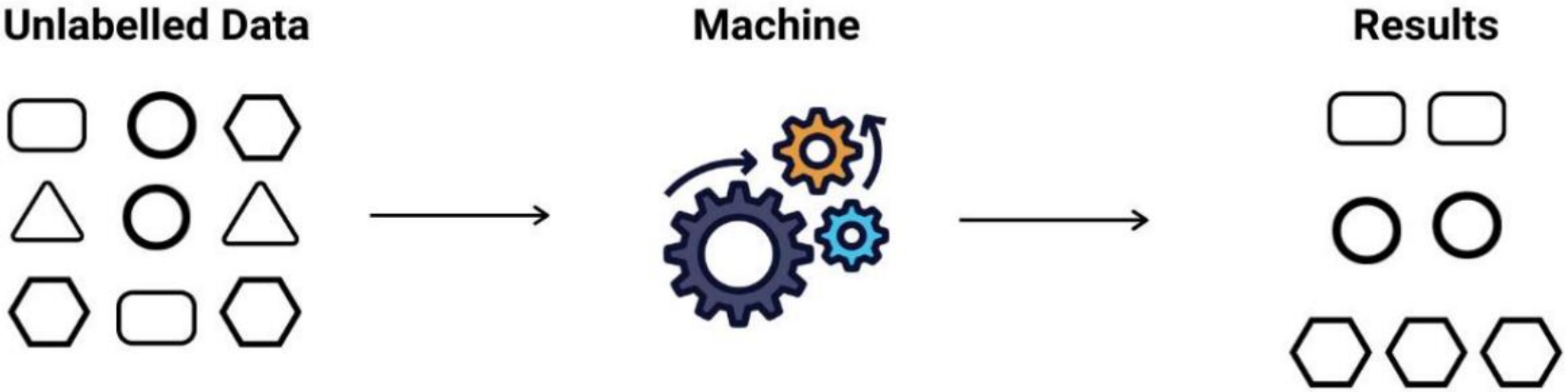
Reinforcement Learning

Learning from mistakes
(Playing Games)

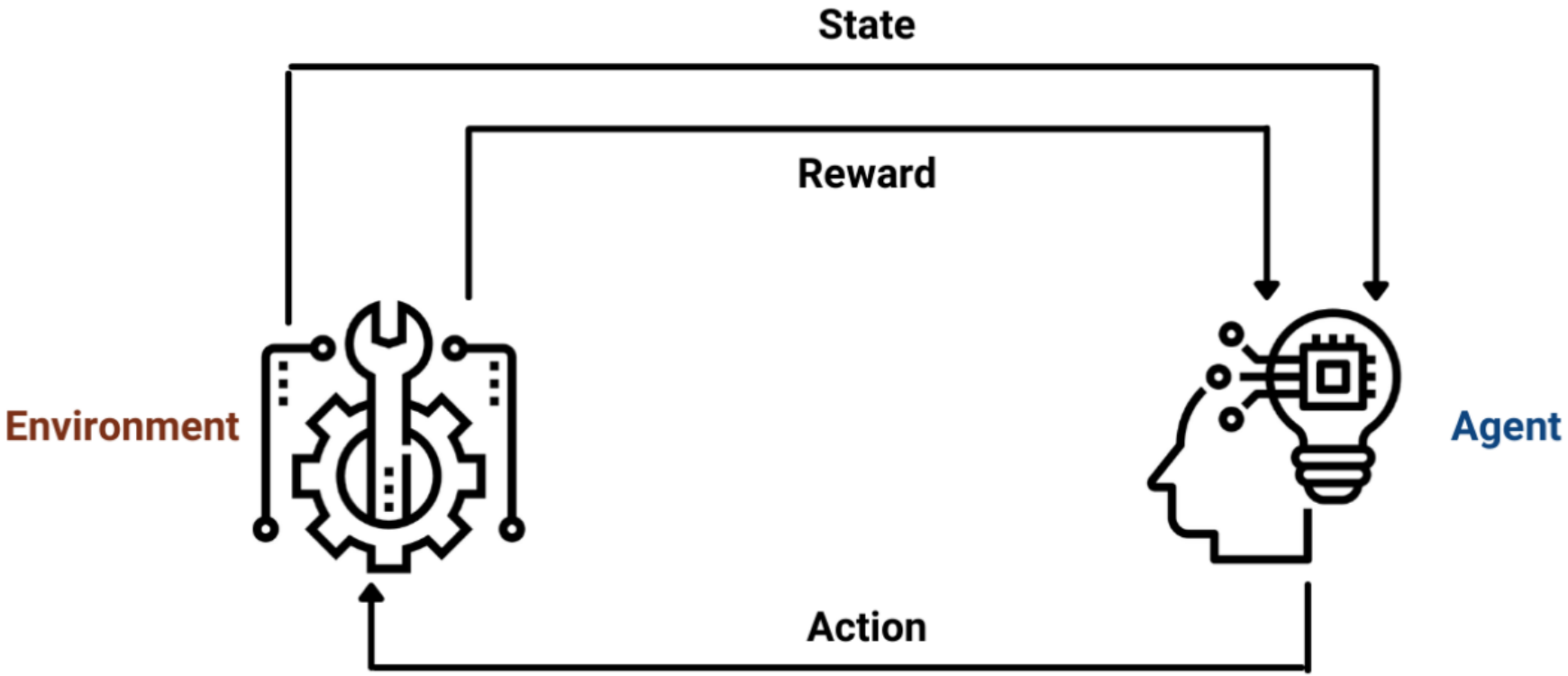
Supervised Learning



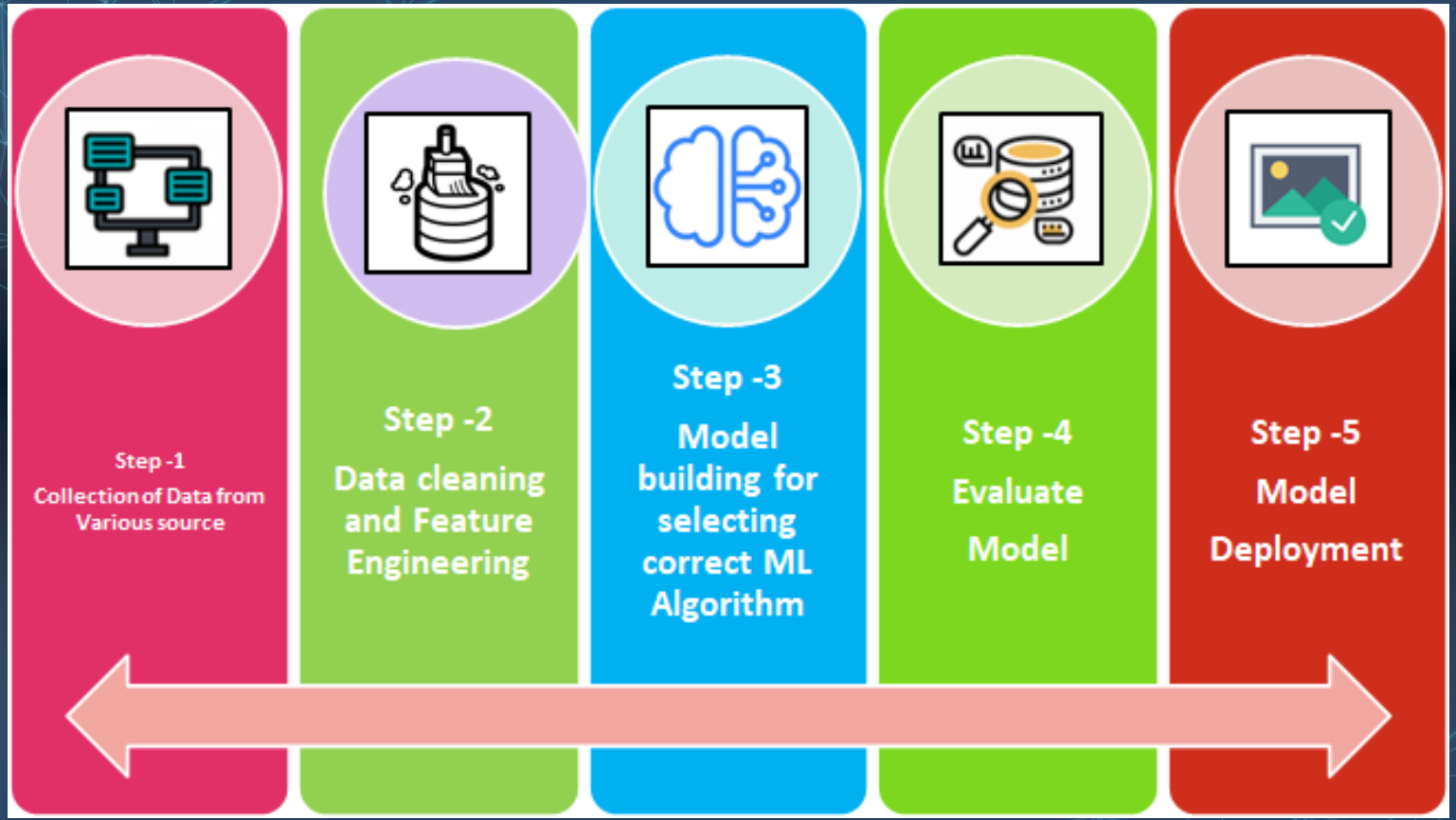
Unsupervised Learning



Reinforcement Learning



Machine Learning Process



THE DATA SCIENCE PROCESS

PIPELINE

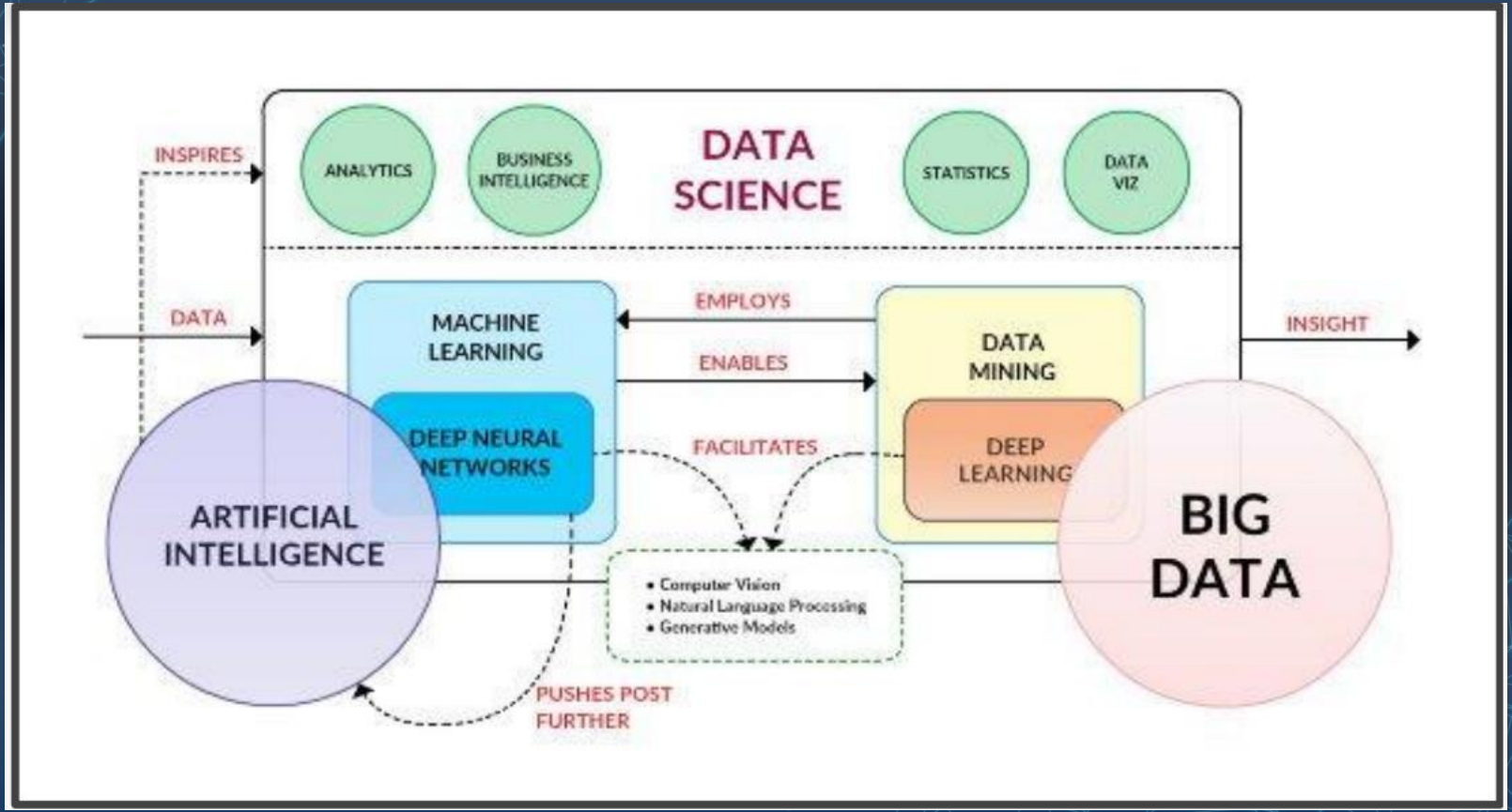
Data Engineers

Data Analysts

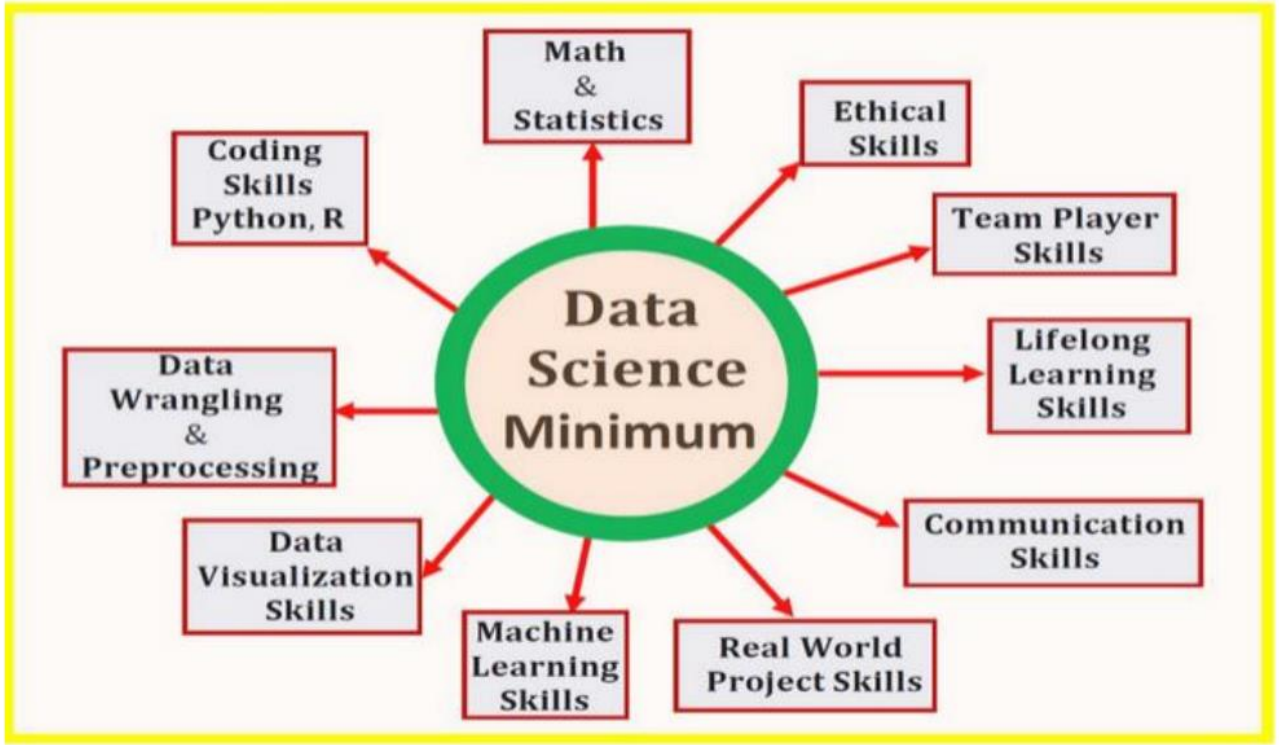
Machine Learning Engineers

Data Scientists

Data Science



Data Science

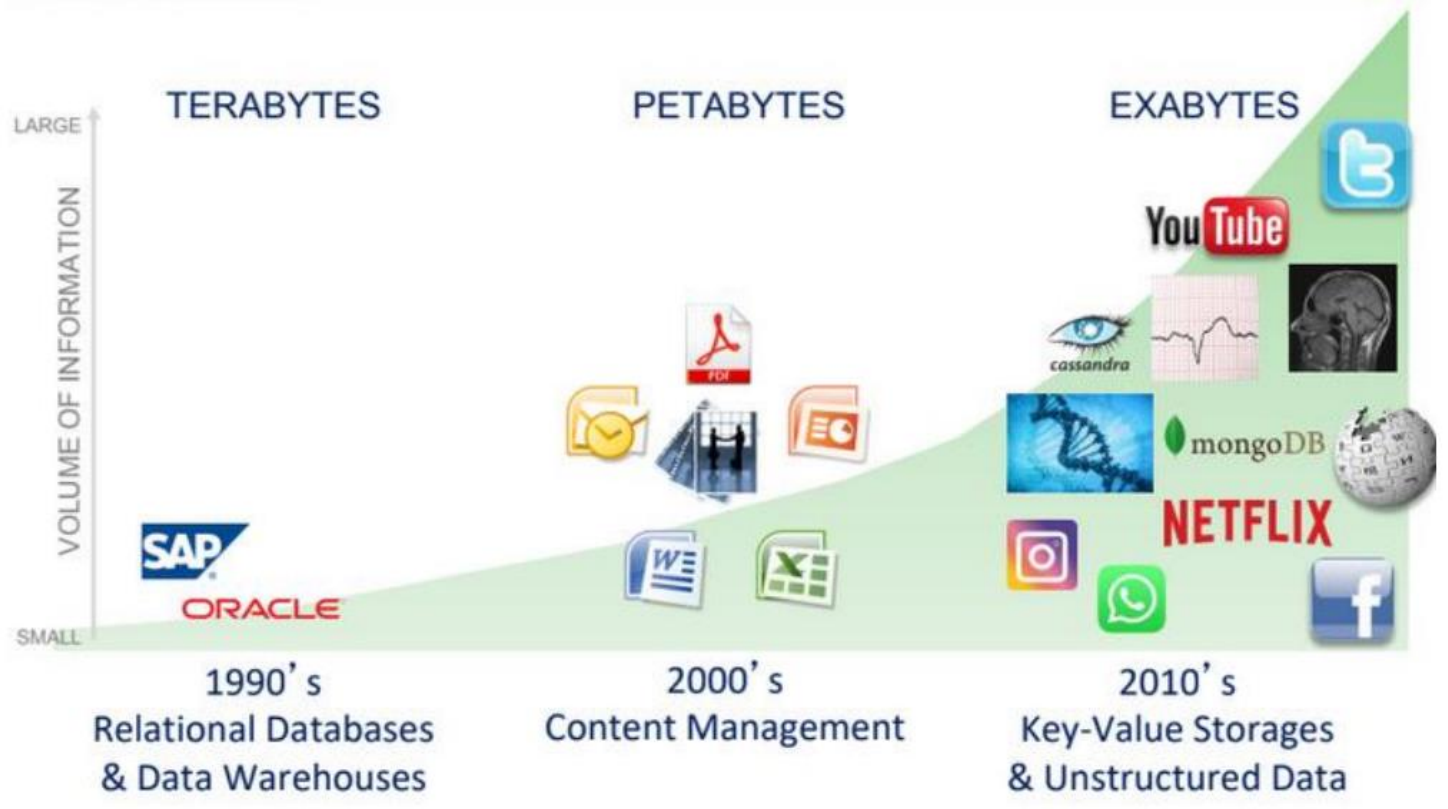


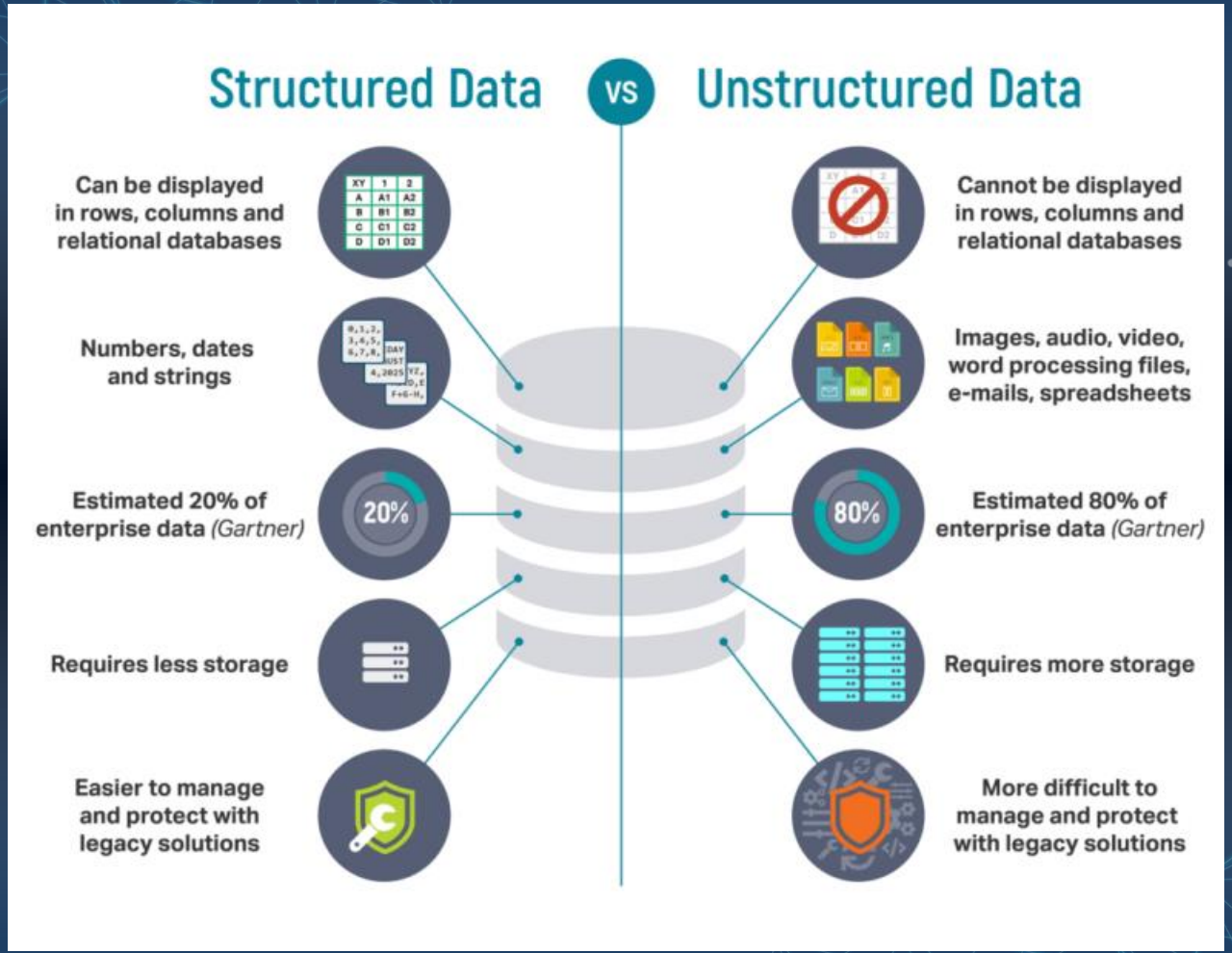
Big data:

refers to large and complex sets of data that are beyond the capacity of traditional data processing and management tools to handle effectively. Big data is characterized by the three Vs:

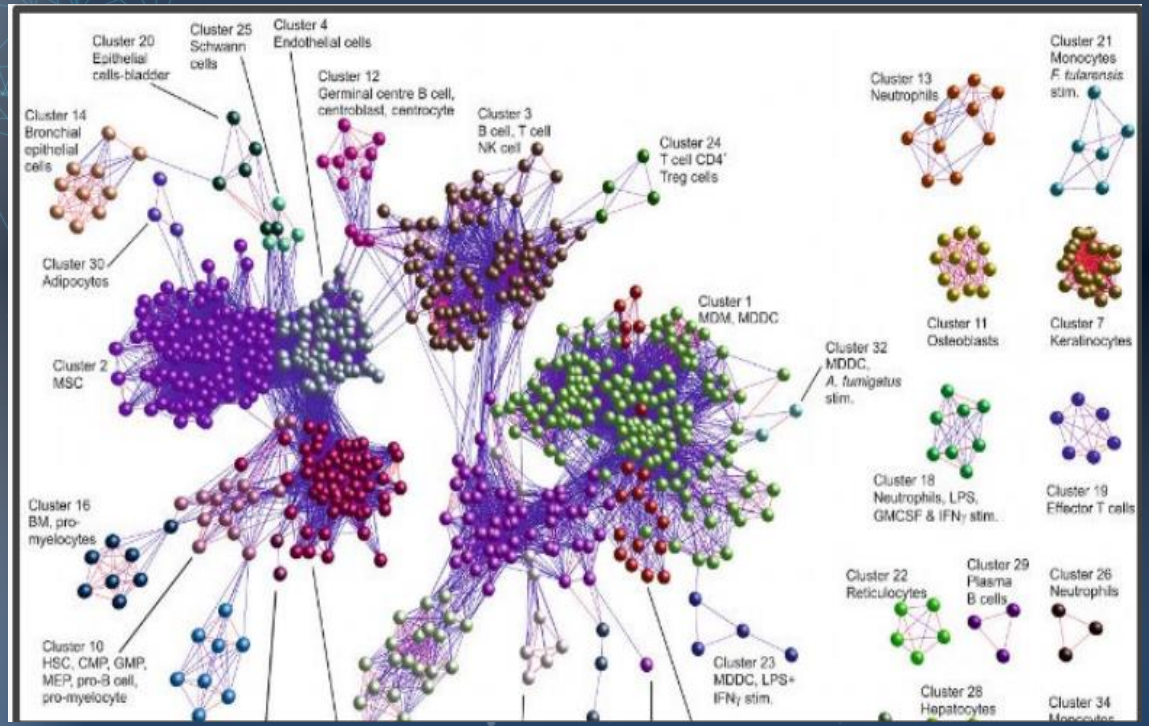
1. **Volume**: Big data involves a massive amount of data. This data can range from terabytes to petabytes or even exabytes in size. Traditional databases and data processing systems are not equipped to handle such enormous data volumes.
2. **Velocity**: Data in the big data context often comes in at high speeds. This can include data streaming in real-time from various sources like social media, sensors, and web applications. Managing and processing data at these high velocities requires specialized tools and techniques.
3. **Variety**: Big data encompasses diverse types of data, including structured data (e.g., databases and spreadsheets), semi-structured data (e.g., XML, JSON), and unstructured data (e.g., text documents, social media posts, images, and videos). This variety of data types adds complexity to data processing and analysis.


BIG
DATA
& ANALYTICS

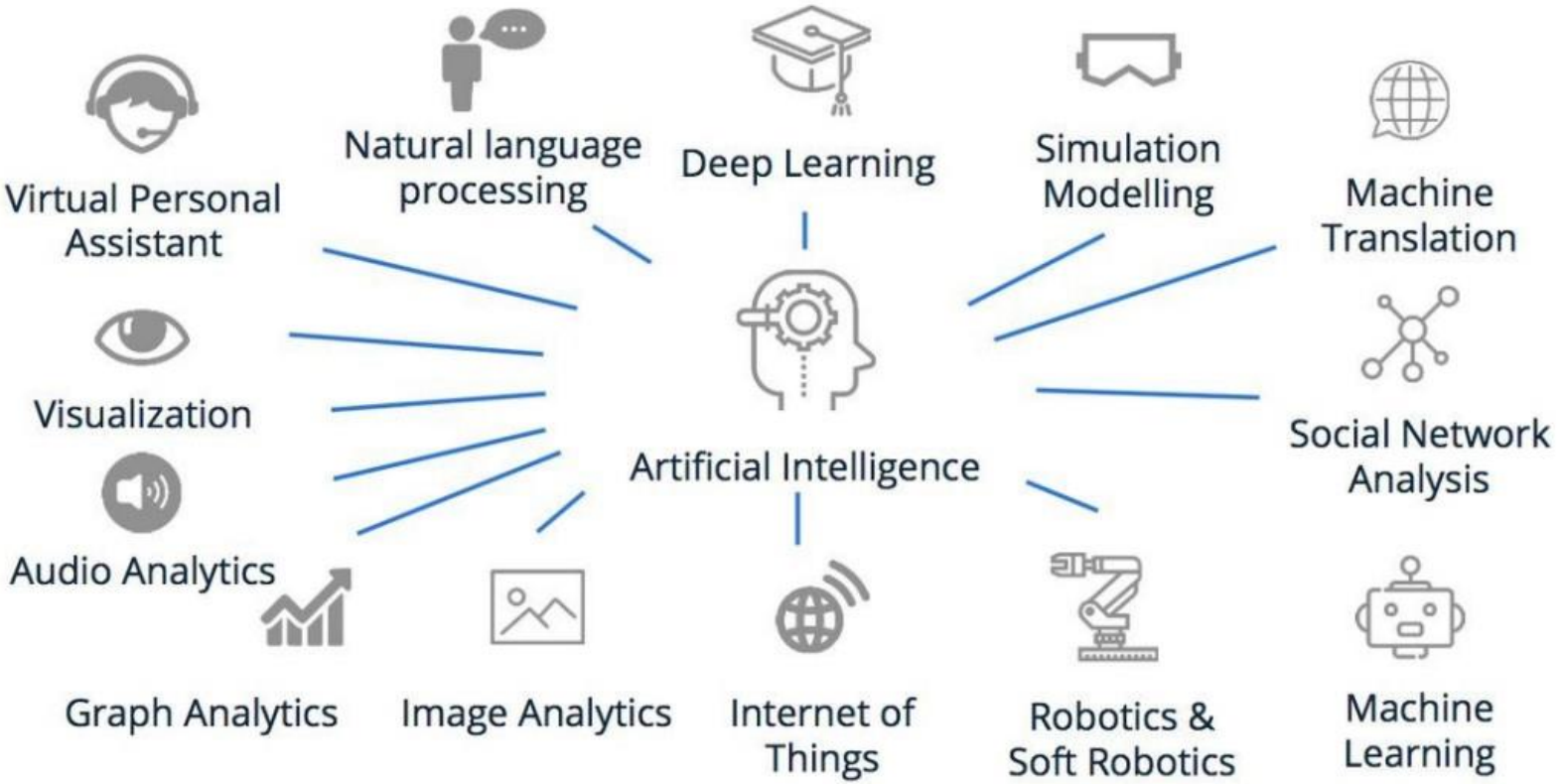

TERABYTES
PETABYTES
EXABYTES
LARGE
VOLUME OF INFORMATION
SMALL
SAP
ORACLE
YouTube
cassandra
mongoDB
NETFLIX
1990's
Relational Databases
& Data Warehouses
2000's
Content Management
2010's
Key-Value Storages
& Unstructured Data


Structured Data
VS
Unstructured Data
Can be displayed in rows, columns and relational databases
Cannot be displayed in rows, columns and relational databases
Numbers, dates and strings
Images, audio, video, word processing files, e-mails, spreadsheets
20%
Estimated 20% of enterprise data (Gartner)
80%
Estimated 80% of enterprise data (Gartner)
Requires less storage
Requires more storage
Easier to manage and protect with legacy solutions
More difficult to manage and protect with legacy solutions

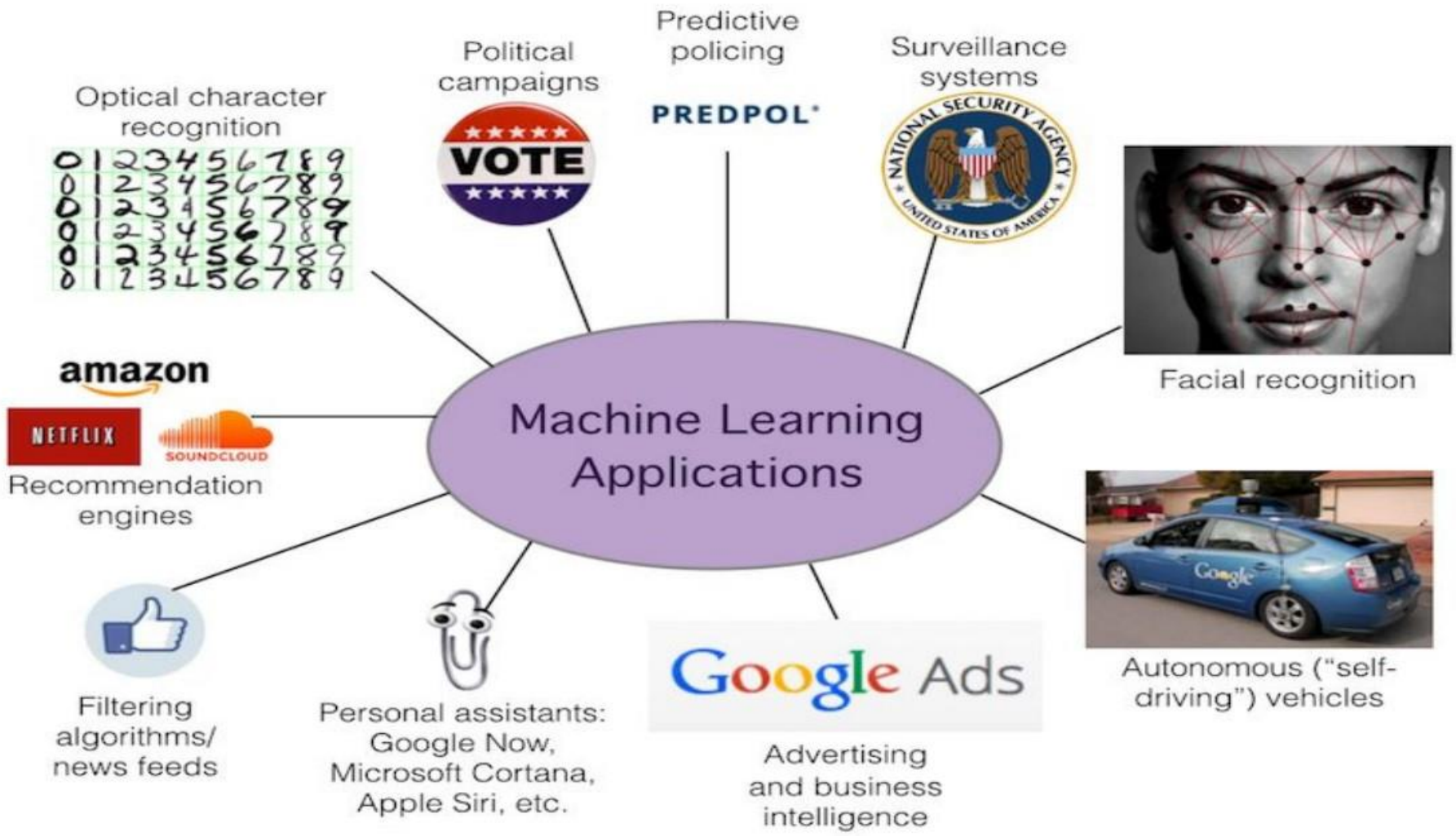
Big data Processing



Possible applications for Artificial Intelligence



source statista via @mikequindazzi


Machine Learning Applications
Optical character recognition
Political campaigns
VOTE
Predictive policing
PREDPOL
Surveillance systems
NATIONAL SECURITY AGENCY
UNITED STATES OF AMERICA
Facial recognition
amazon
NETFLIX
SOUNDCLOUD
Recommendation engines
Filtering algorithms/ news feeds
Personal assistants: Google Now, Microsoft Cortana, Apple Siri, etc.
Google Ads
Advertising and business intelligence
Google
Autonomous ("self-driving") vehicles

Course content:

- Getting Started with Python
- Mathematical Library with NumPy (Linear Algebra)
- Data Preprocessing and Visualization with Pandas & Matplotlib
- Supervised Machine Learning (Classification & Regression)
- Unsupervised Machine Learning (Clustering)
- Advanced Machine Learning
- Introduction to Deep Learning
- Introduction to Computer Vision
- Introduction to Natural Language Processing (NLP)

Why use Python in machine learning?

- Wide range of libraries
- Easy to use and learn
- Large community support
- Interoperability with other languages
- Scalability

PYTHON (JUPYTER NOTEBOOK):

- Most Preferred Programming Language in 2023 Recently used in
- most applications of Computer Science
- Jupyter Notebook is an open-source web app and the preferred IDE of Python since it organizes your code very well. Additionally, some similar platforms support it (Kaggle – Google Collab)

Interesting Fact:
Jupyter is a reference to
3 Programming Languages
Julia, Python, R

jupyter covid_19_dashboard Last Checkpoint: Last Friday at 11:45 PM (unsaved changes) ✔

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code Voila

In [13]:
```
# importing libraries

from __future__ import print_function
from ipywidgets import interact, interactive, fixed, interact_manual
from IPython.core.display import display, HTML

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
import folium
import plotly.graph_objects as go
import seaborn as sns
import ipywidgets as widgets
```

In [14]:
```
# loading data right from the source:
death_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_
confirmed_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_co
recovered_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_co
country_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/web-data/data/cases_country.csv')
```

In [15]:
```
confirmed_df.head()
```

In [16]:
```
recovered_df.head()
```

In [17]:
```
death_df.head()
```

In [18]:
```
country_df.head()
```

Python Libraries

1. **Scientifics Computing** Libraries

Pandas
(Data structures & tools)

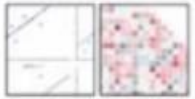
NumPy
(Arrays & matrices)

SciPy
(Integrals, solving differential equations, optimization)

Python Libraries

2. **Visualization** Libraries

Matplotlib
(plots & graphs, most popular)

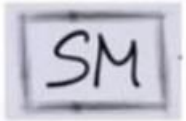


Seaborn
(plots : heat maps, time series, violin plots)

Python Libraries

3. **Algorithmic** libraries

Scikit-learn
(Machine Learning : regression, classification,...)



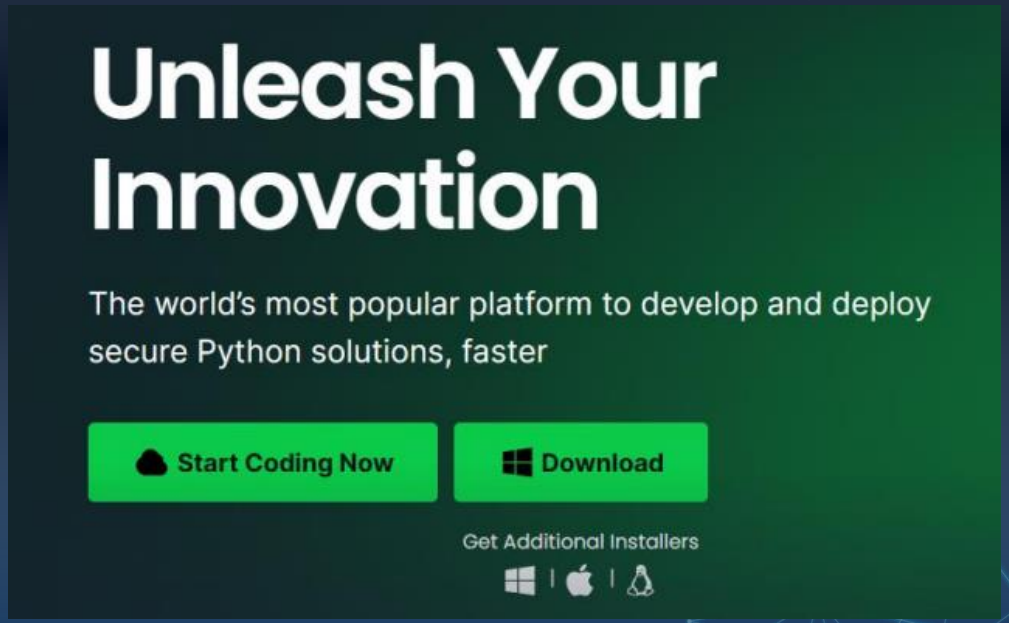
Statsmodels
(Explore data, estimate statistical models, and perform statistical tests.)

Python Libraries

TENSORFLOW & PYTORCH

USING ANACONDA PLATFORM

- **Download Anaconda From The Given Link (anaconda.com)**



SETUP YOUR KAGGLE ACCOUNT :

- The Biggest Data Science and Machine Learning Platform in the world.
- It provides a HUGE number of Datasets freely to be used in the projects.
- There are some free courses for short hours of some topics such as (Data Visualization, Data Analysis, Machine Learning, Advanced Machine Learning … etc).

Git & GitHub:

GitHub is a web-based platform for version control and collaboration that allows developers and teams to work on software projects together. It provides a wide range of tools and features for managing and tracking changes to code, collaborating with other developers, and hosting and sharing software projects.

Important Links:

- Home - UCI Machine Learning Repository
- Numpy_Cheat_Sheet.pdf (datacamp.com)
- Pandas_Cheat_Sheet.pdf (datacamp.com)
- Matplotlib_Cheat_Sheet.pdf (datacamp.com)

Tasks:

- Install Anaconda
- Prepare Kaggle Account
- Prepare GitHub Account
- Write First Jupyter Notebook

Code For testing

In mathematics, the **symmetric difference** of two sets is the set of elements which are in either of the sets, but not in their intersection. Two following sets are given:

1.A = {2, 4, 6, 8}
2.B = {4, 10}

Using the appropriate method, extract the symmetrical difference of sets A and B and print the result to the console as shown below.

Expected result:

1.Symmetric difference: {2, 6, 8, 10}

The variable *x*:

```
x = -1.5
```

and the following expression are given:

```
expression = 'x**2 + x'
```

Using the appropriate function, calculate the value of this expression and print the result to the console.

Tip: Use the `eval()` function.

Expected result:

```
0.75
```

The following variables are given:

```
1.var1 = 'Python'
2.var2 = ('Python')
3.var3 = ('Python',)
4.var4 = ['Python']
5.var5 = {'Python'}
```

Using the appropriate function, check if the variables are instances of *tuple* class. Print the result to the console as shown below.

Tip: Use the `isinstance()` built-in function.

Expected result:

```
1.False
2.False
3.True
4.False
5.False
```

The following list is given:

```
1.characters = ['k', 'b', 'c', 'j', 'z', 'w']
```

Using the built-in functions, return the first and the last letter in alphabetical order from this list and print the result to the console as shown below.

Tip: Use the `min()` and `max()` functions.

Expected result:

```
1.First: b
2.Last: z
```


Thank You