



KESHAV MEMORIAL INSTITUTE OF TECHNOLOGY

AN AUTONOMOUS INSTITUTION - ACCREDITED BY NAAC WITH 'A' GRADE

Narayanaguda, Hyderabad.

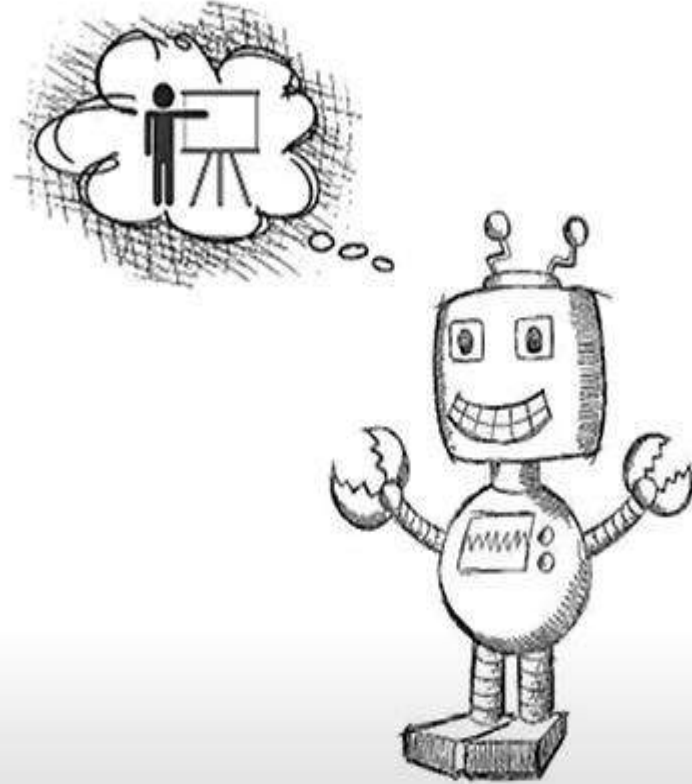
Deep Learning

SESSION-1

**BY
ASHA**

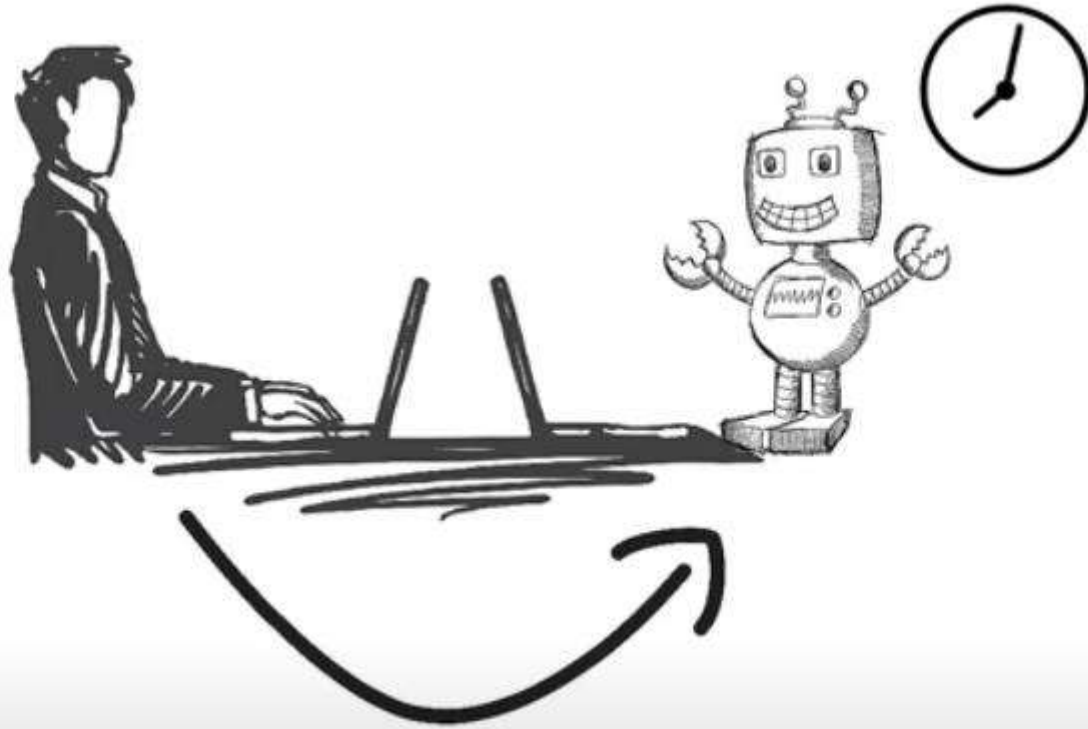


**HUMANS LEARN FROM
PAST EXPERIENCES**



**MACHINES FOLLOW INSTRUCTIONS
GIVEN BY HUMANS**

WHAT IF HUMANS CAN TRAIN THE MACHINES...



Execute Instructions

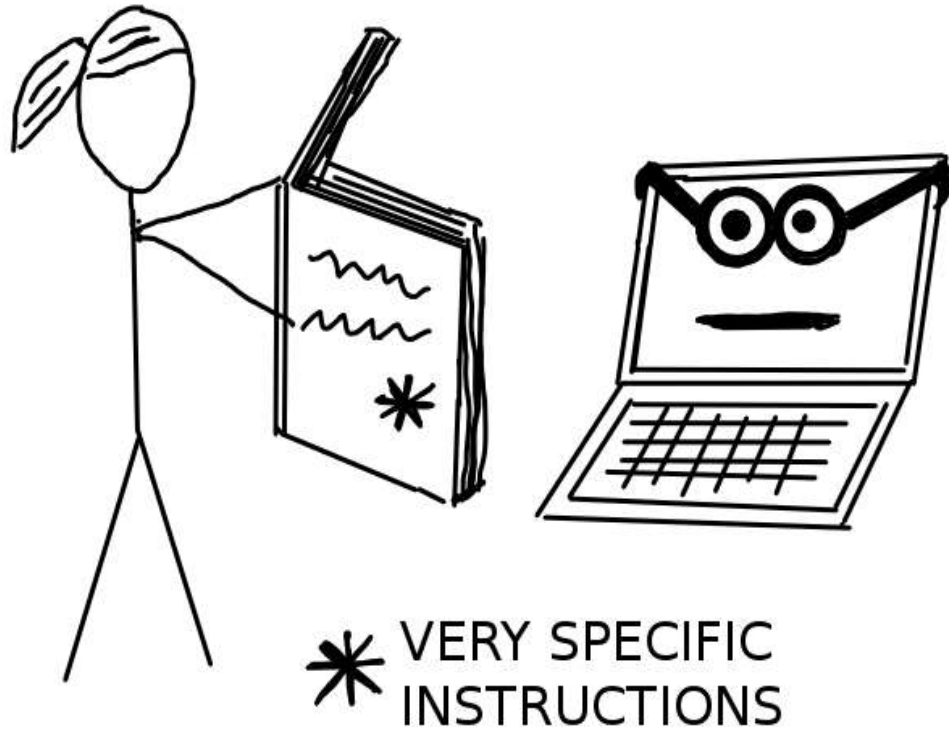
```
function withdraw(amount) {  
  if (amount > balance) {  
    fail("Hey you ain't got the cash!")  
  } else {  
    balance = balance - amount  
  }  
}
```



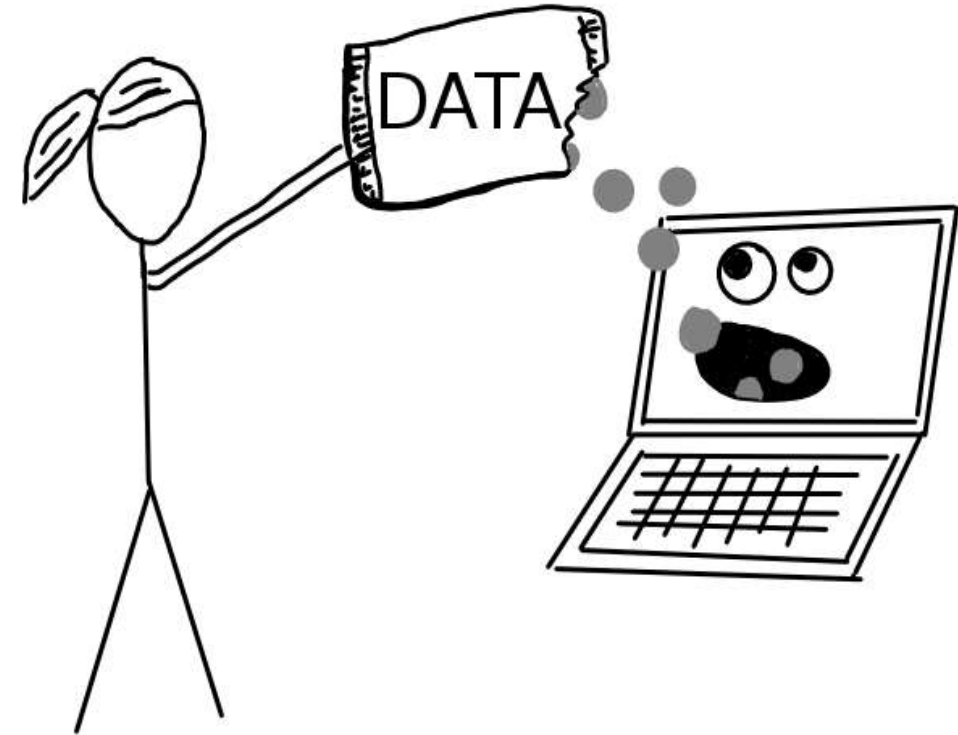
Learn + Think

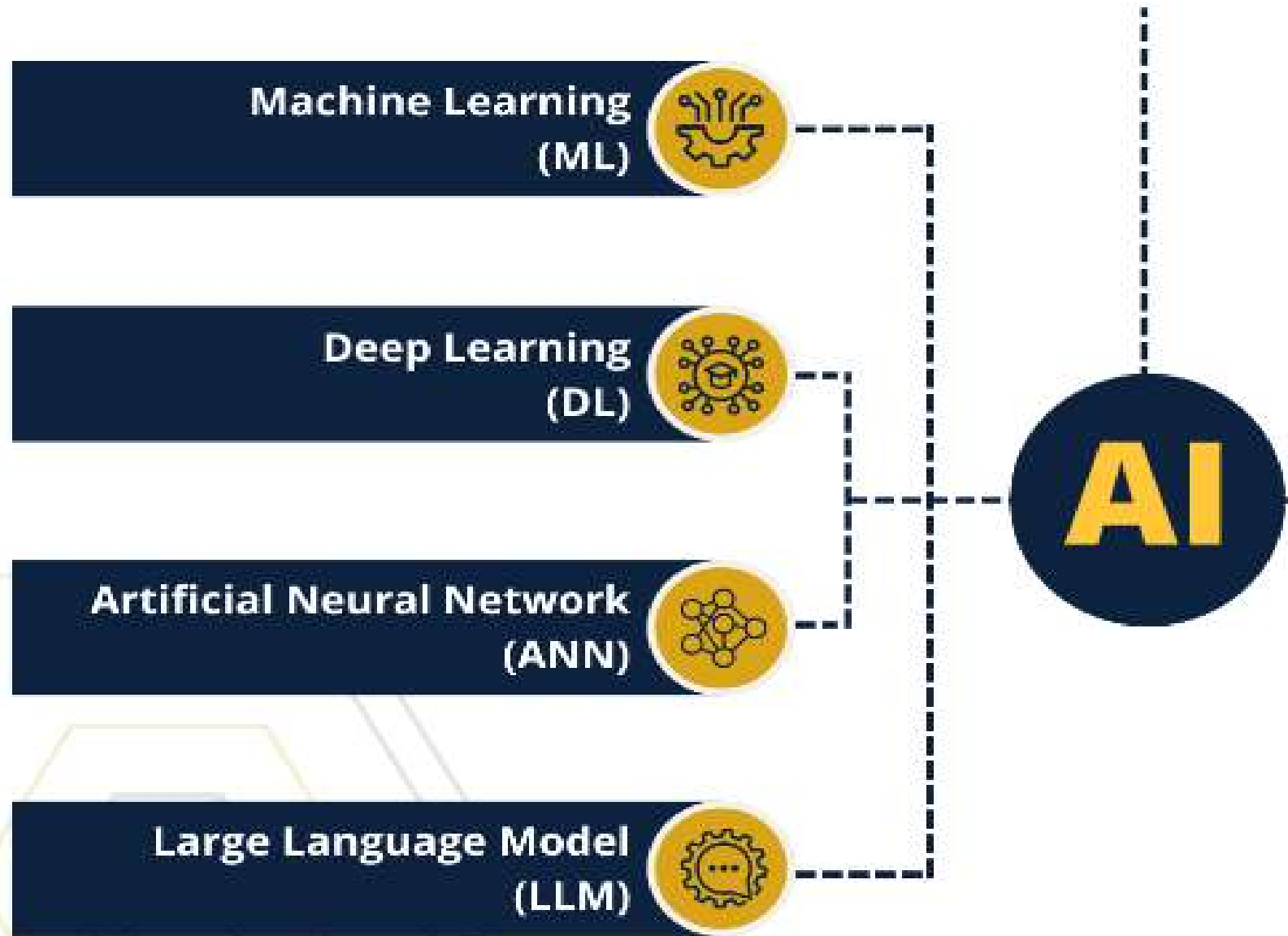


Without Machine Learning



With Machine Learning





ARTIFICIAL INTELLIGENCE

- ARTIFICIAL INTELLIGENCE REFERS TO THE DEVELOPMENT OF COMPUTER SYSTEMS THAT CAN PERFORM TASKS THAT TYPICALLY REQUIRE HUMAN INTELLIGENCE.



Tesla Autopilot Car



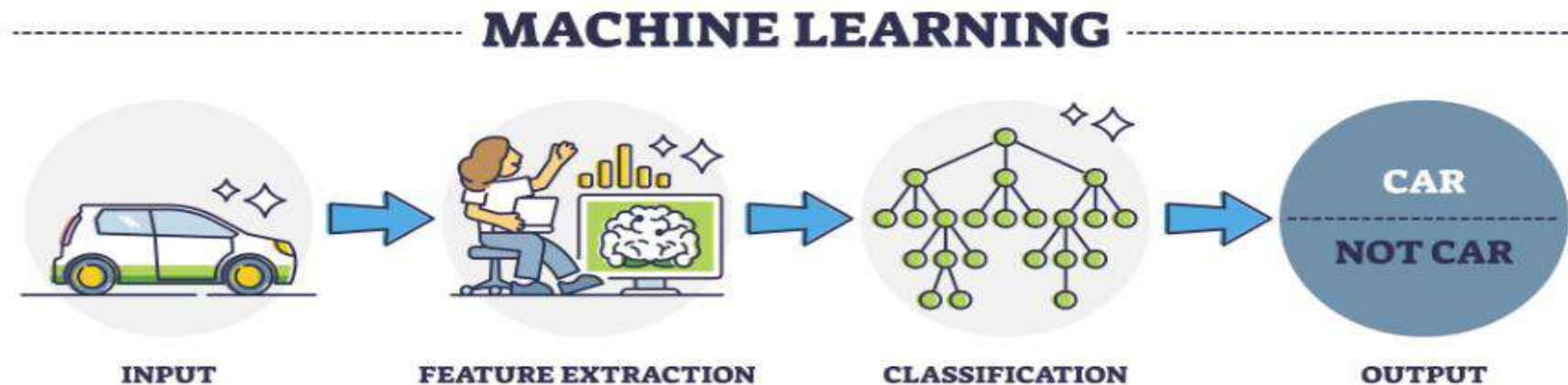
Sophia Humanoid Robot



Amazon Alexa

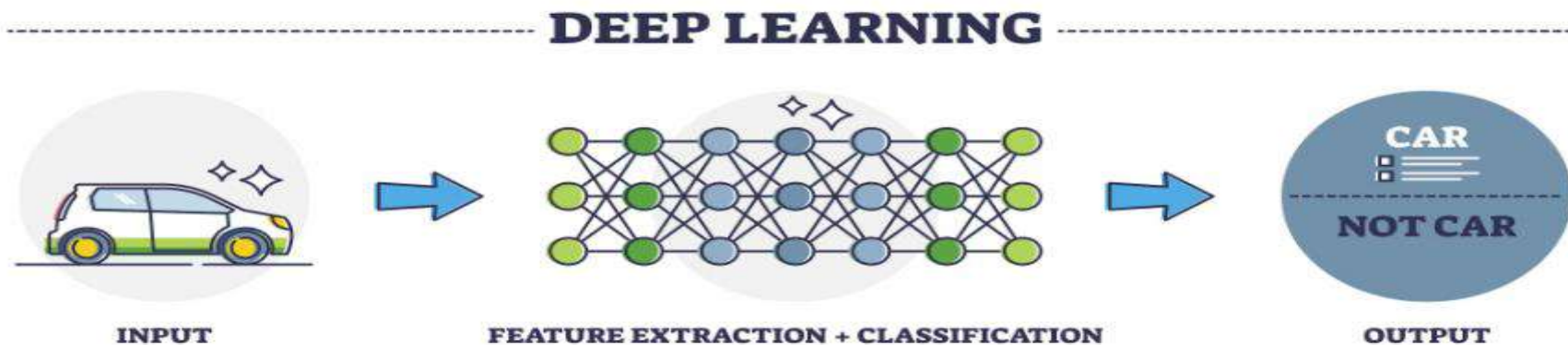
MACHINE LEARNING

- MACHINE LEARNING INVOLVES ALGORITHMS AND STATISTICAL MODELS THAT ENABLE COMPUTERS TO IMPROVE THEIR PERFORMANCE ON A SPECIFIC TASK WITHOUT EXPLICIT PROGRAMMING.
- IT FOCUSES ON PATTERN RECOGNITION AND LEARNING FROM DATA.
- MACHINE LEARNING IS A SUBSET OF ARTIFICIAL INTELLIGENCE.



DEEP LEARNING

- DEEP LEARNING IS A SUBSET OF MACHINE LEARNING THAT INVOLVES NEURAL NETWORKS WITH MULTIPLE LAYERS (DEEP NEURAL NETWORKS).
- THESE NETWORKS CAN AUTOMATICALLY LEARN TO EXTRACT FEATURES FROM DATA AND MAKE COMPLEX DECISIONS BASED ON LARGE AMOUNTS OF DATA.



Machine learning vs. deep learning



Machine learning

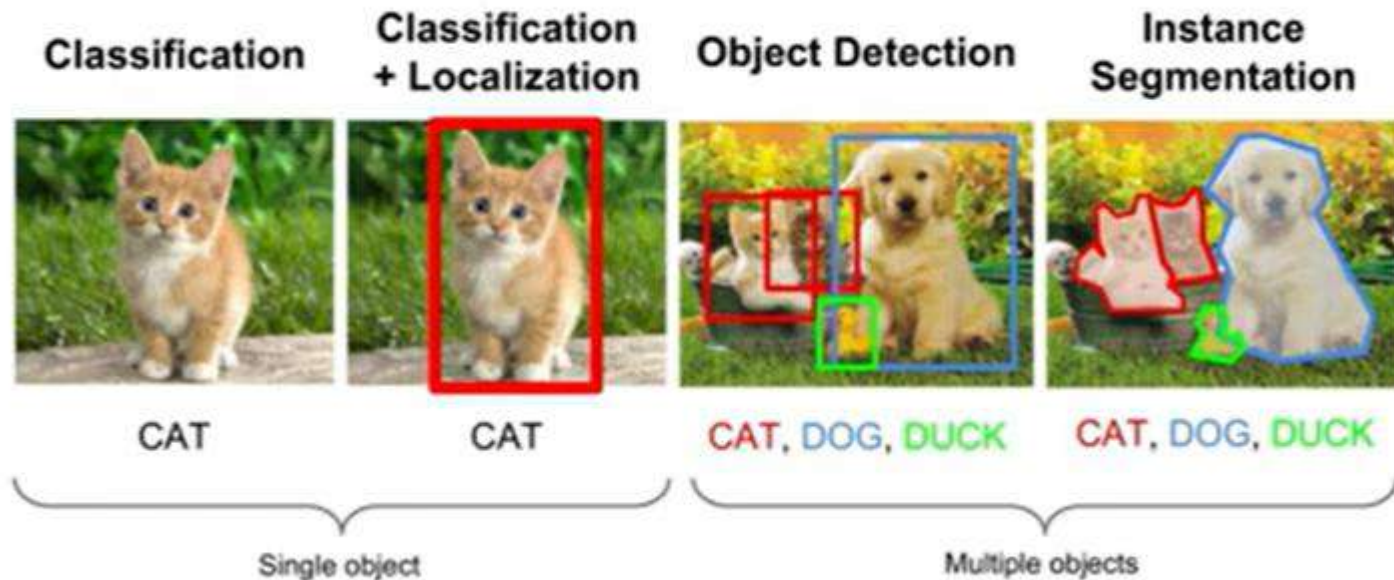
Uses algorithms and learns on its own but may need human intervention to correct errors



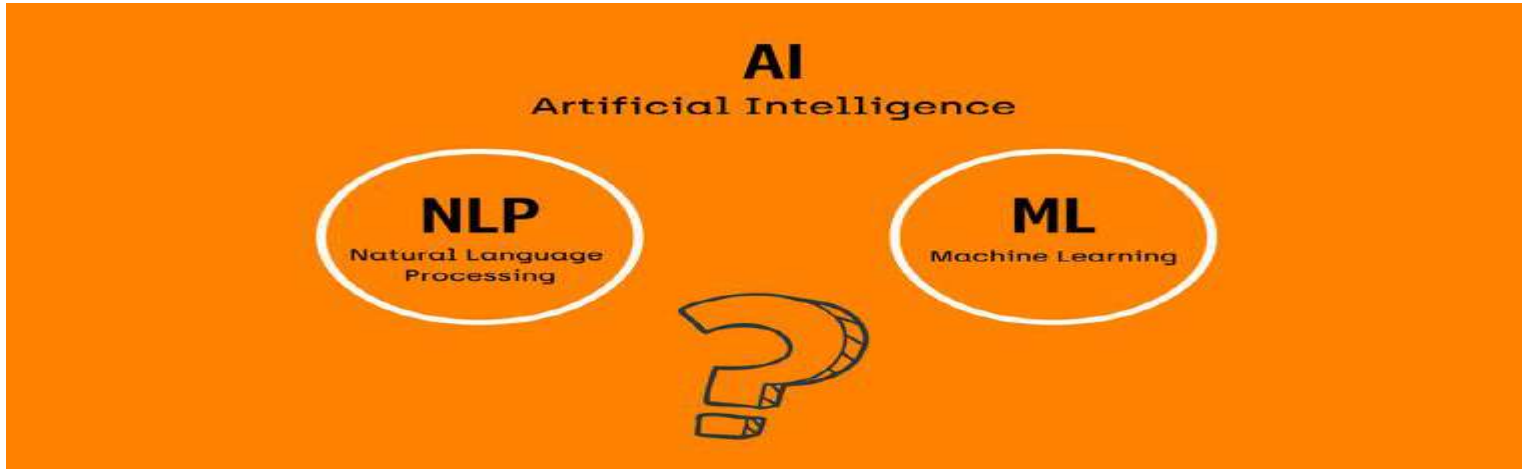
Deep learning

Uses advanced computing, its own neural network, to adapt with little to no human intervention

COMPUTER VISION (CV) IS A FIELD OF ARTIFICIAL INTELLIGENCE THAT ENABLES COMPUTERS TO INTERPRET AND UNDERSTAND THE VISUAL WORLD. USING DIGITAL IMAGES FROM CAMERAS AND VIDEOS AND DEEP LEARNING MODELS, MACHINES CAN ACCURATELY IDENTIFY AND CLASSIFY OBJECTS, AND THEN REACT TO WHAT THEY "SEE."



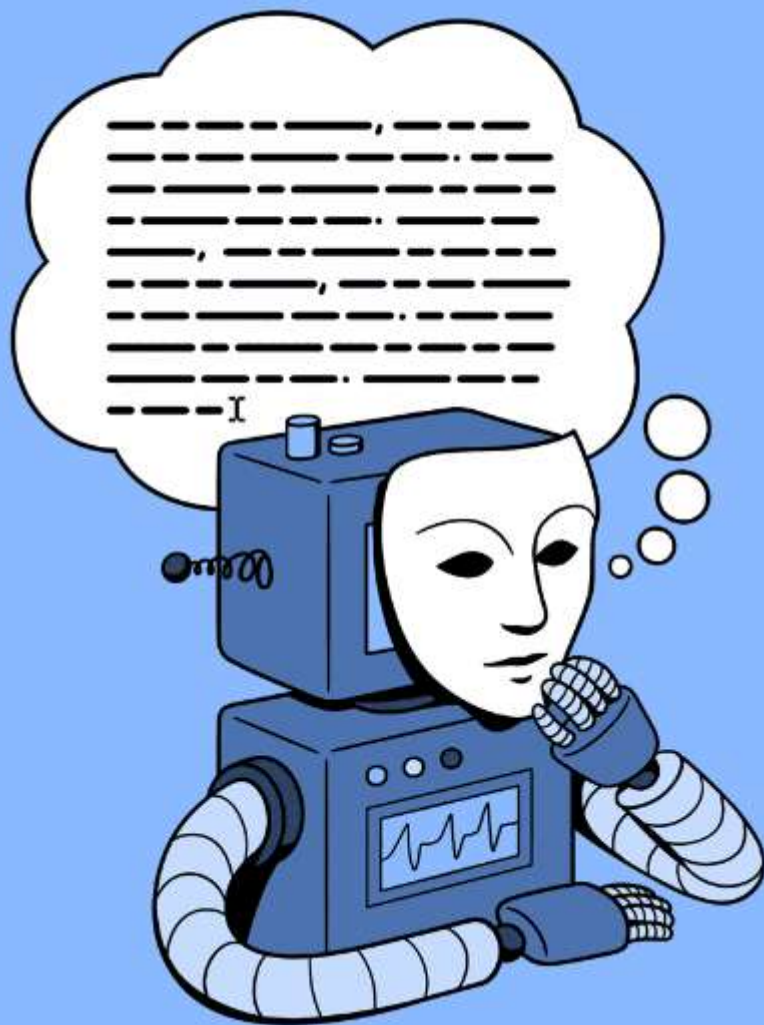
NATURAL LANGUAGE PROCESSING



- NATURAL LANGUAGE PROCESSING (NLP) IS A SUBFIELD OF ARTIFICIAL INTELLIGENCE THAT FOCUSES ON THE INTERACTION BETWEEN COMPUTERS AND HUMANS THROUGH NATURAL LANGUAGE.
- THE ULTIMATE GOAL OF NLP IS TO ENABLE COMPUTERS TO UNDERSTAND, INTERPRET, AND GENERATE HUMAN LANGUAGES IN A VALUABLE WAY.

LARGE LANGUAGE MODEL

- LARGE LANGUAGE MODELS ARE ADVANCED AI MODELS TRAINED ON VAST AMOUNTS OF TEXT DATA, ENABLING THEM TO UNDERSTAND AND GENERATE HUMAN-LIKE LANGUAGE.
- VIRTUAL ASSISTANTS LIKE SIRI OR ALEXA UTILIZE LARGE LANGUAGE MODELS TO UNDERSTAND AND RESPOND TO NATURAL LANGUAGE QUERIES.



Large Language Model (LLM)

['lärj 'laŋ-gwij 'mä-dəl]

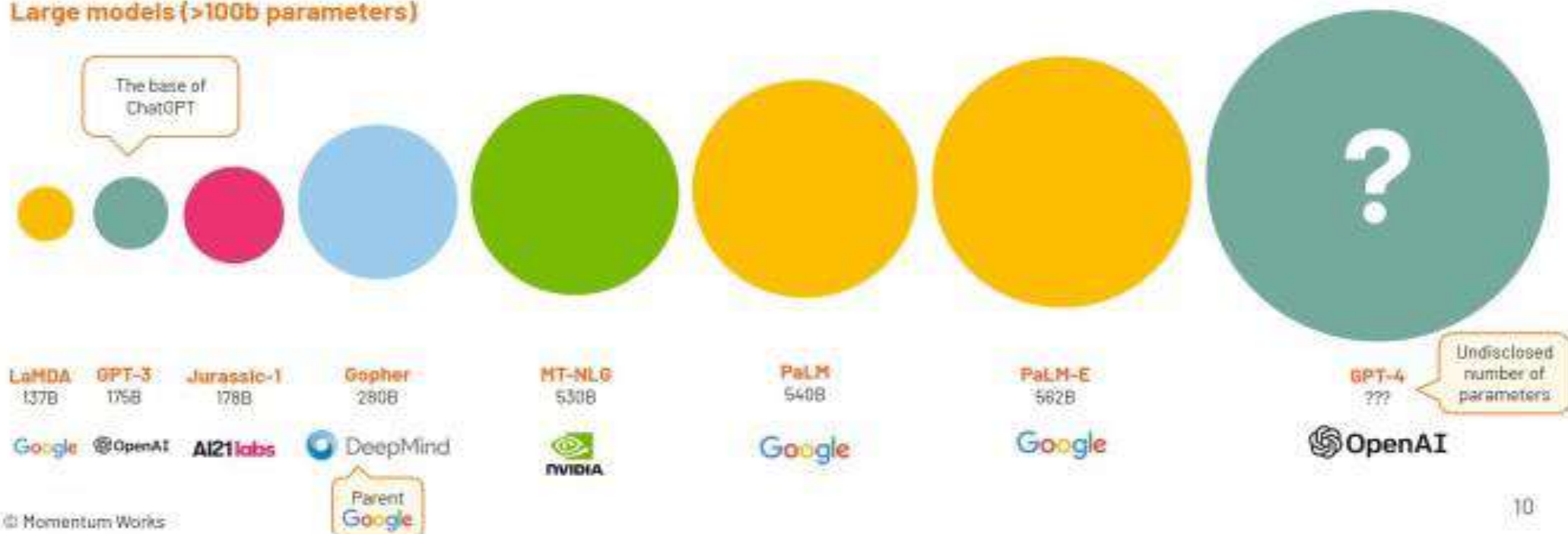
A deep learning algorithm that's equipped to summarize, translate, predict, and generate human-sounding text to convey ideas and concepts.

Large Language Models are becoming very large indeed

Small models ($\leq 100b$ parameters)



Large models ($>100b$ parameters)



GENERATIVE AI

- GENERATIVE AI REFERS TO ARTIFICIAL INTELLIGENCE SYSTEMS THAT ARE CAPABLE OF CREATING NEW CONTENT SUCH AS TEXT, IMAGES, OR MUSIC.
- THESE SYSTEMS LEARN FROM EXISTING DATA PATTERNS AND GENERATE FRESH, ORIGINAL CONTENT.

GENERATIVE AI

who are you? explain it to a 10 year old kid



Hi there! I'm ChatGPT, and I'm like a super-smart robot that can talk to you and help answer your questions. Imagine I'm a really big library that knows a lot about almost everything, but I can talk and understand what you say. Here's a simple way to understand me:

MACHINE LEARNING



In 1959, Arthur Samuel, a computer scientist who pioneered the study of artificial intelligence, described machine learning as “The study that gives computers the ability to learn.”

Machine learning is a subset of AI, which enables the machine to automatically learn from data, improve performance from past experiences, and make predictions.



Artificial Intelligence

AI involves techniques that equip computers to emulate human behavior, enabling them to learn, make decisions, recognize patterns, and solve complex problems in a manner akin to human intelligence.

Machine Learning

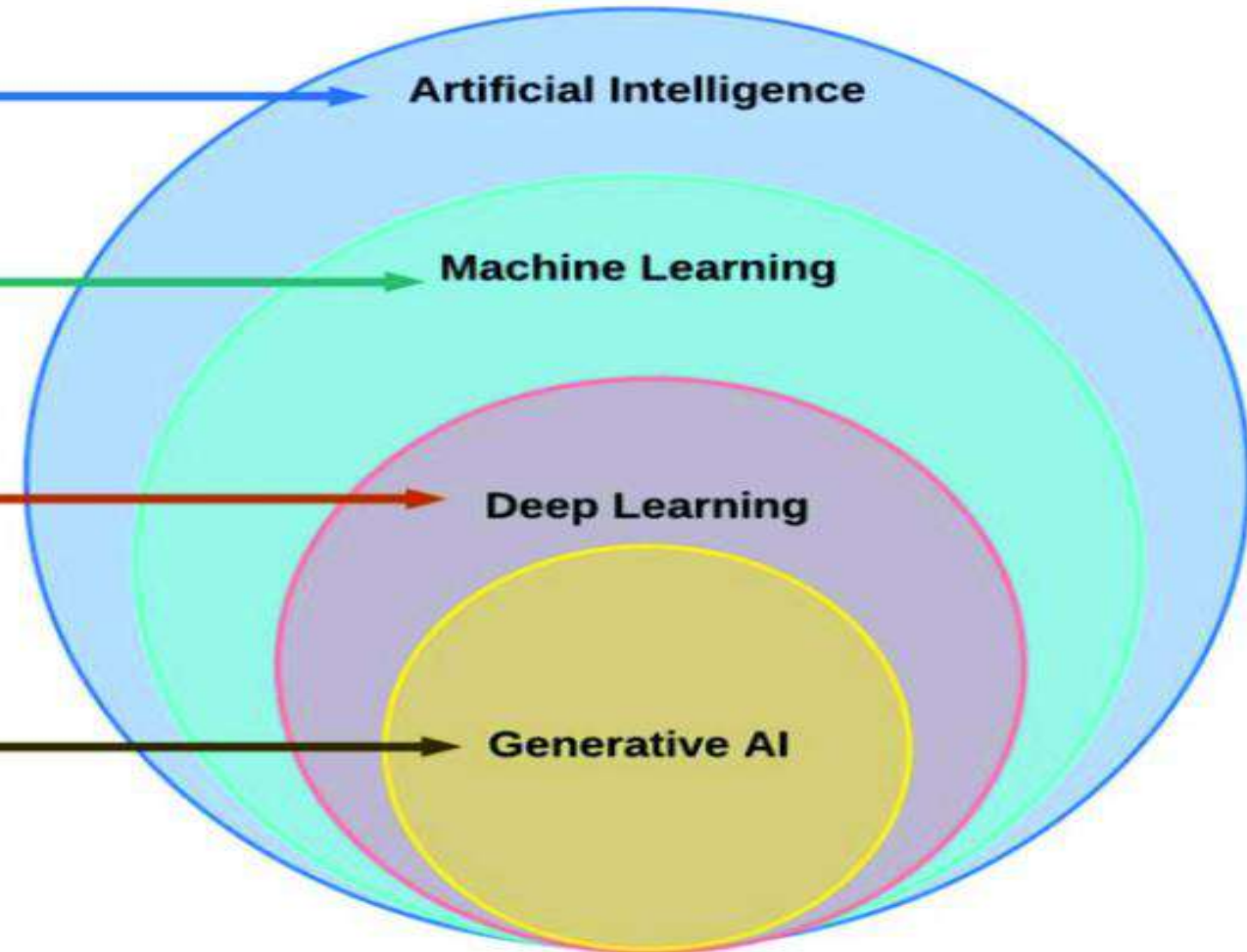
ML is a subset of AI, uses advanced algorithms to detect patterns in large data sets, allowing machines to learn and adapt. ML algorithms use supervised or unsupervised learning methods.

Deep Learning

DL is a subset of ML which uses neural networks for in-depth data processing and analytical tasks. DL leverages multiple layers of artificial neural networks to extract high-level features from raw input data, simulating the way human brains perceive and understand the world.

Generative AI

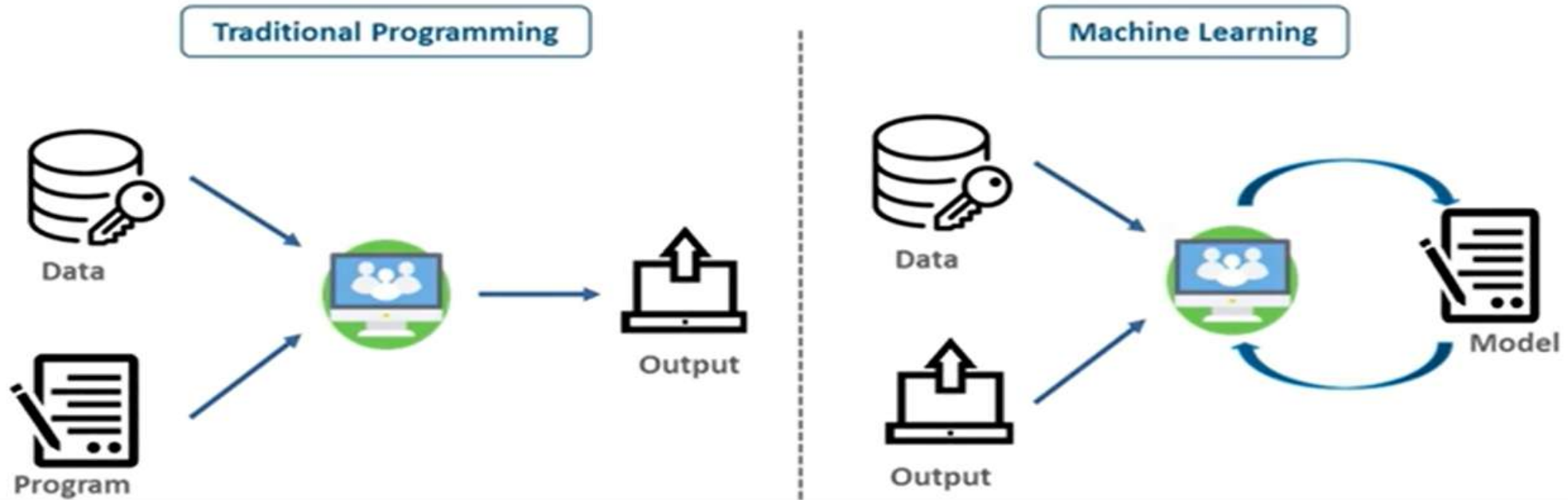
Generative AI is a subset of DL models that generates content like text, images, or code based on provided input. Trained on vast data sets, these models detect patterns and create outputs without explicit instruction, using a mix of supervised and unsupervised learning.



MACHINE LEARNING IS A SUBSET OF ARTIFICIAL INTELLIGENCE THAT AIMS TO MIMIC HOW HUMAN BEINGS LEARN BY USING DATA.

A more technical definition given by Tom M. Mitchell's (1997) : “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.”

HOW MACHINE LEARNING WORKS?

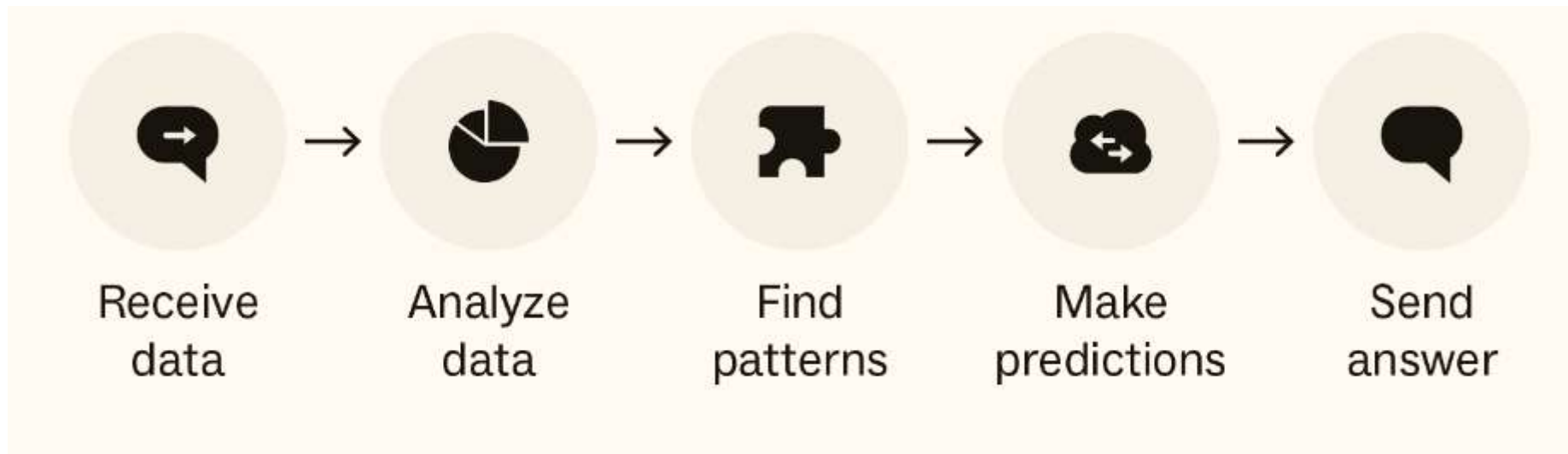


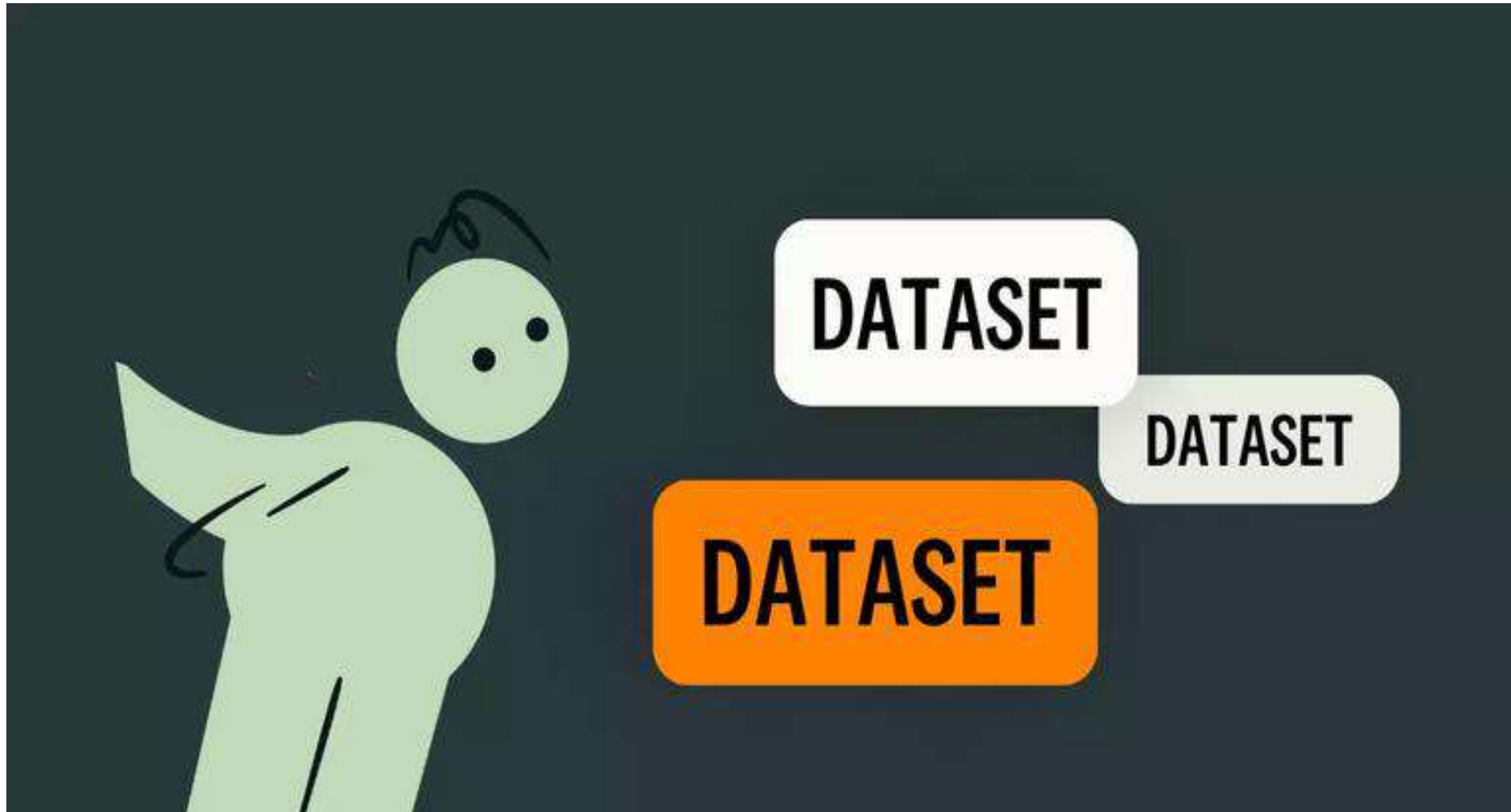
Learn from Data

Find Hidden Insights

Train and Grow

HOW MACHINE LEARNING WORKS?





DATASET

- A dataset is a collection of data in which data is arranged in some order. A dataset can contain any data from a series of an array to a database table.
- A tabular dataset can be understood as a database table or matrix, where each column corresponds to a particular variable, and each row corresponds to the fields of the dataset. The most supported file type for a tabular dataset is "Comma Separated File," or CSV.

NEED OF DATASET

To work with machine learning projects, we need a huge amount of data. Collecting and preparing the dataset is one of the most crucial parts while creating an ML/AI project.

POPULAR SOURCES FOR MACHINE LEARNING DATASETS

- Kaggle Datasets
- UCI Machine Learning Repository
- Datasets via AWS
- Google's Dataset Search Engine
- Microsoft Datasets

DATA PREPROCESSING

DATA PRE-PROCESSING IS A PROCESS OF CLEANING THE RAW DATA I.E. THE DATA IS COLLECTED IN THE REAL WORLD AS MOST OF THE REAL-WORLD DATA IS MESSY, SOME OF THESE TYPES OF DATA ARE:

1. MISSING DATA
2. NOISY DATA
3. INCONSISTENT DATA

WHY IS DATA PREPROCESSING IMPORTANT?

THE MAJORITY OF THE REAL-WORLD DATASETS FOR MACHINE LEARNING ARE HIGHLY SUSCEPTIBLE TO BE MISSING, INCONSISTENT, AND NOISY.

- DATA PROCESSING IS, THEREFORE, IMPORTANT TO IMPROVE THE OVERALL DATA QUALITY.
- DUPLICATE OR MISSING VALUES MAY GIVE AN INCORRECT VIEW OF THE OVERALL STATISTICS OF DATA
- OUTLIERS AND INCONSISTENT DATA POINTS OFTEN TEND TO DISTURB THE MODEL'S OVERALL LEARNING, LEADING TO FALSE PREDICTIONS.

DATA REDUCTION

- THE SIZE OF THE DATASET IN A DATA WAREHOUSE CAN BE TOO LARGE TO BE HANDLED BY DATA ANALYSIS AND DATA MINING ALGORITHMS.
- ONE POSSIBLE SOLUTION IS TO OBTAIN A REDUCED REPRESENTATION OF THE DATASET THAT IS MUCH SMALLER IN VOLUME BUT PRODUCES THE SAME QUALITY OF ANALYTICAL RESULTS.

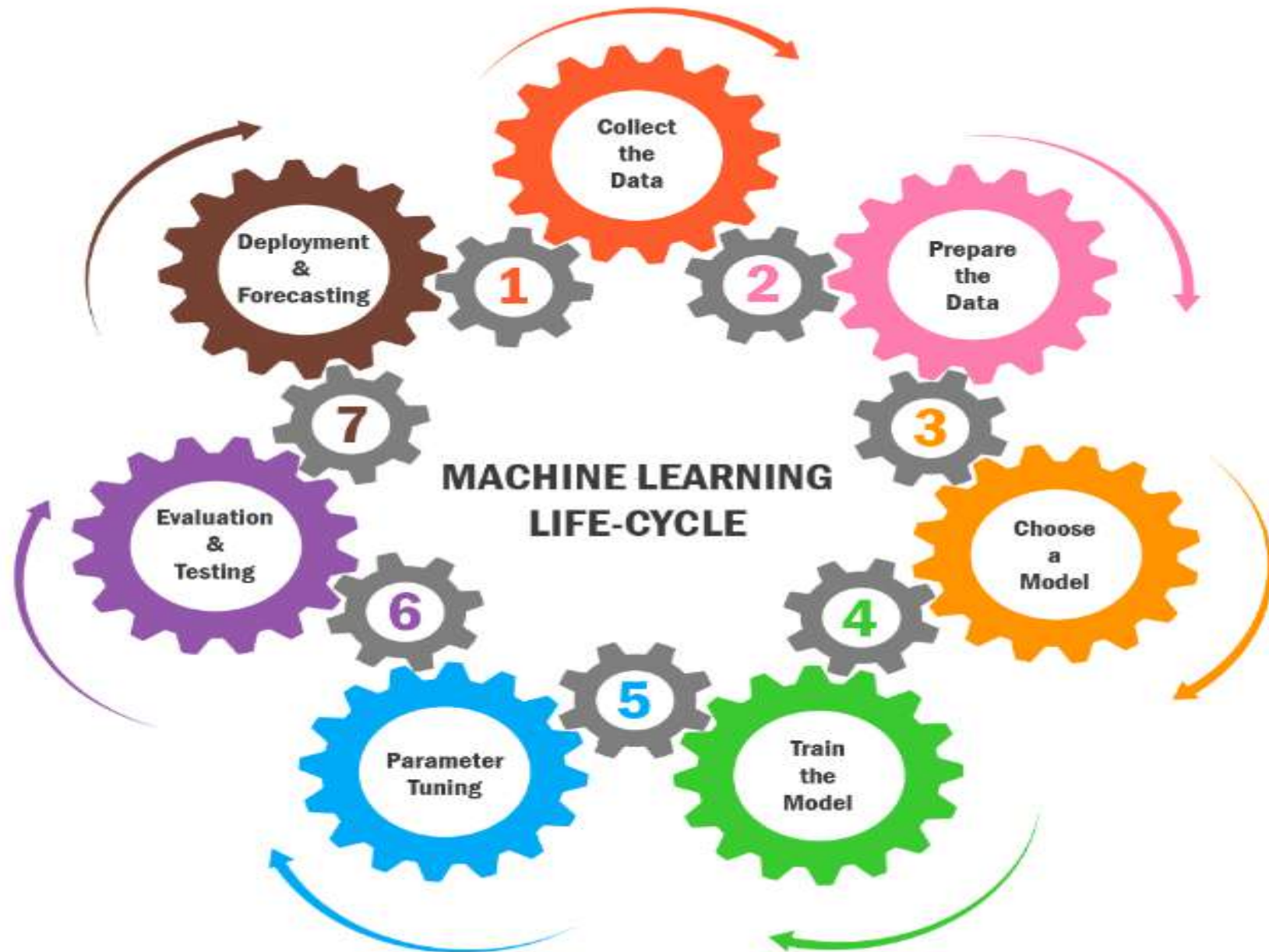
- **HANDLING MISSING VALUES:** TECHNIQUES INCLUDE REMOVING INSTANCES WITH MISSING VALUES, IMPUTING MISSING VALUES WITH THE MEAN, MEDIAN, OR MODE, OR USING ADVANCED TECHNIQUES LIKE KNN IMPUTATION.
- **REMOVING DUPLICATES:** IDENTIFYING AND REMOVING DUPLICATE INSTANCES TO ENSURE THE DATASET IS CLEAN.
- **FEATURE SCALING:**
- **NORMALIZATION:** RESCALING THE FEATURES TO A RANGE OF $[0, 1]$.
- **STANDARDIZATION:** RESCALING THE FEATURES TO HAVE A MEAN OF 0 AND A STANDARD DEVIATION OF 1.

ENCODING CATEGORICAL DATA:

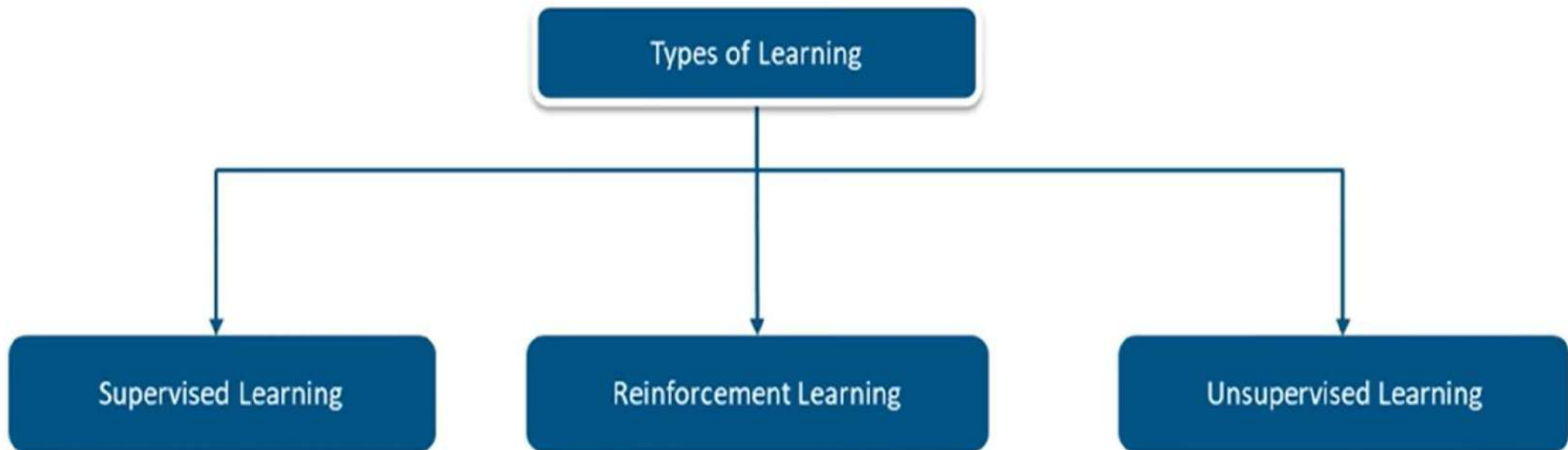
- **ONE-HOT ENCODING:** CONVERTING CATEGORICAL VARIABLES INTO BINARY VECTORS.
- **LABEL ENCODING:** CONVERTING CATEGORICAL VARIABLES INTO INTEGER VALUES.

SPLITTING DATA:

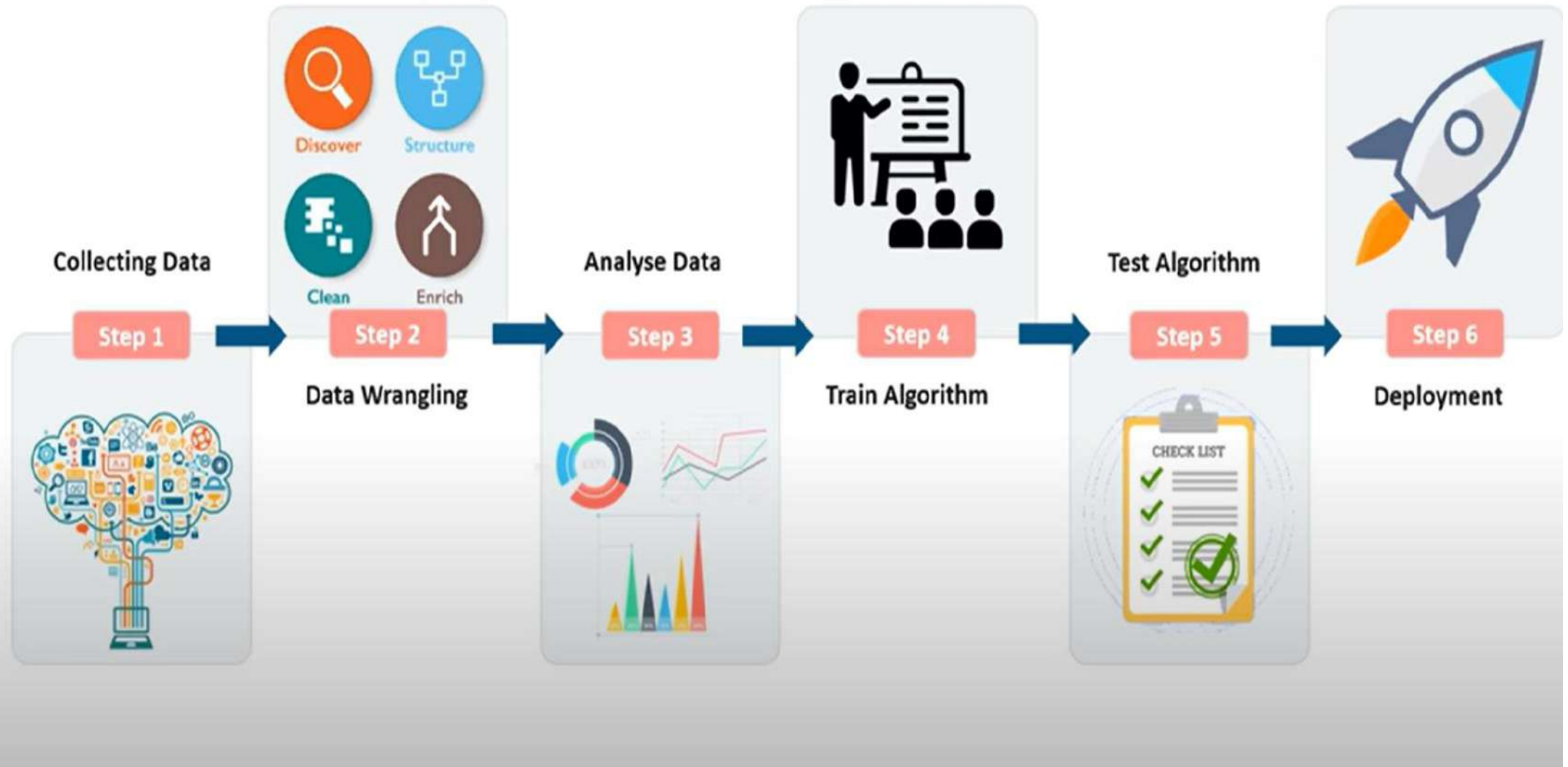
- DIVIDING THE DATASET INTO TRAINING AND TESTING SETS TO EVALUATE THE MODEL'S PERFORMANCE.



MACHINE LEARNING TYPES ?



ALGORITHM DEVELOPMENT STEPS



BASIC TERMINOLOGY

FEATURES AND LABELS:

- **FEATURES:** THE INPUT VARIABLES (INDEPENDENT VARIABLES) USED BY THE MODEL TO MAKE PREDICTIONS.
- **LABELS:** THE OUTPUT VARIABLE (DEPENDENT VARIABLE) THAT THE MODEL IS TRYING TO PREDICT.

← Features →					Label
Position	Experience	Skill	Country	City	Salary (\$)
Developer	0	1	USA	New York	103100
Developer	1	1	USA	New York	104900
Developer	2	1	USA	New York	106800
Developer	3	1	USA	New York	108700
Developer	4	1	USA	New York	110400
Developer	5	1	USA	New York	112300
Developer	6	1	USA	New York	114200
Developer	7	1	USA	New York	116100
Developer	8	1	USA	New York	117800
Developer	9	1	USA	New York	119700
Developer	10	1	USA	New York	121600

TRAINING AND TESTING:

- **TRAINING SET:** A SUBSET OF THE DATASET USED TO TRAIN THE MODEL.
- **TESTING SET:** A SUBSET OF THE DATASET USED TO EVALUATE THE MODEL'S PERFORMANCE.

TRAINING SET

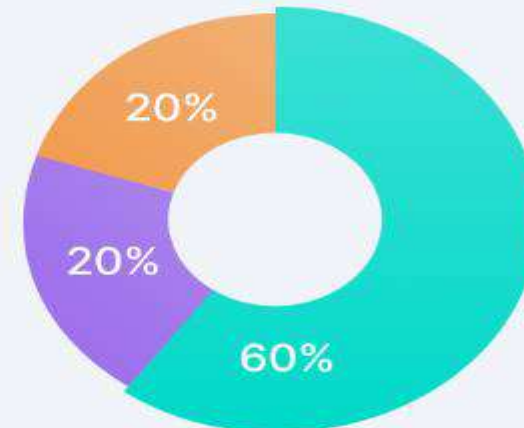
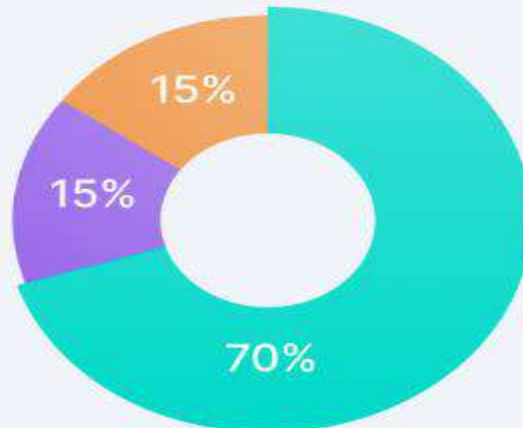
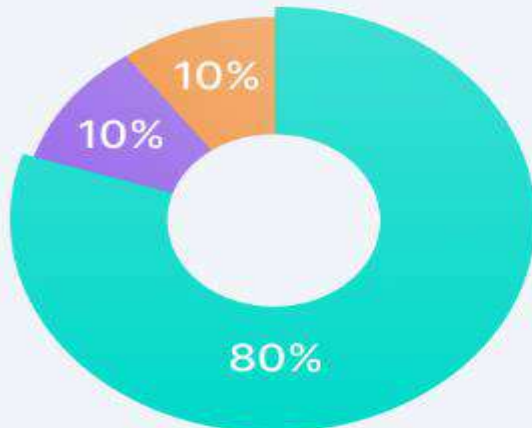
The subset of data used to train a machine learning model

TEST SET

The subset of data used to evaluate the performance of a trained machine learning model on unseen examples, simulating real-world data

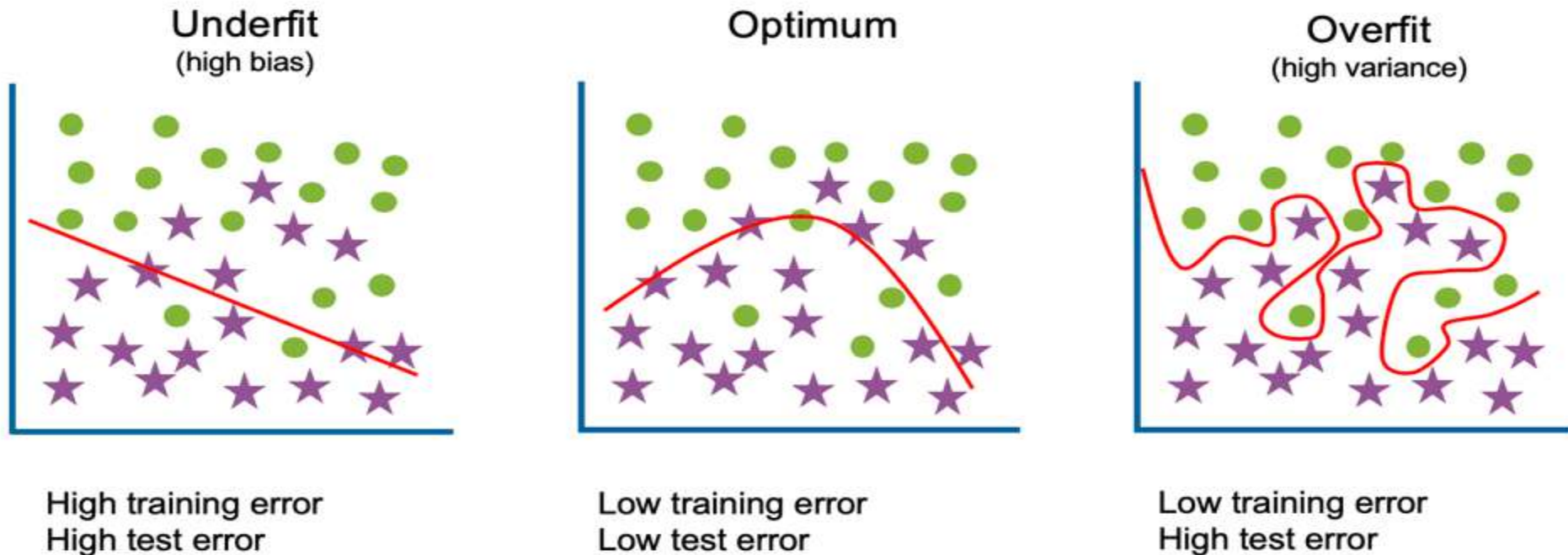
VALIDATION SET

The intermediary subset of data used during the model development process to fine-tune hyperparameters



OVERFITTING AND UNDERFITTING:

- **OVERFITTING:** WHEN THE MODEL PERFORMS WELL ON THE TRAINING DATA BUT POORLY ON THE TESTING DATA BECAUSE IT HAS LEARNED NOISE AND DETAILS FROM THE TRAINING DATA.
- **UNDERFITTING:** WHEN THE MODEL PERFORMS POORLY ON BOTH THE TRAINING AND TESTING DATA BECAUSE IT IS TOO SIMPLE TO CAPTURE THE UNDERLYING PATTERNS IN THE DATA.



WHAT LIBRARIES DO WE USE FOR MACHINE LEARNING?

Seaborn

Focused on the visual of statistical models which include heat maps and depict the overall distributions

Matplotlib

It enables you to make- Bar charts, Scatter plots, Line Charts, Histograms, Pie charts, Contour plots, Quiver plots

Scikit-Learn

Simple and efficient or data mining and data analysis, Built on NumPy and matplotlib, Open source

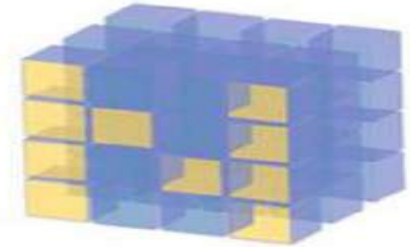
Pandas

Perfect tool for data wrangling, designed for quick and easy data manipulation, aggregation, and visualization

Numpy

Stands for Numerical Python, provides an abundance of useful features for operations on n-arrays and matrices in Python

NumPy



- NumPy – Numerical python is a very popular python library for array and matrix processing, with the help of a large collection of high-level mathematical functions.
- It is very useful for fundamental scientific computations in Machine Learning.

Pandas



- Pandas-Panel data is a popular Python library for data analysis.
- It is not directly related to Machine Learning but the dataset must be prepared before training for which Pandas are useful as it is developed specifically for data extraction and preparation.
- It provides data structures and wide variety tools for data analysis. It provides many inbuilt methods for filtering, combining and grouping data.

Matplotlib

- Matplotlib is a Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It is needed when a programmer wants to visualize the patterns in the data.
- A module named pyplot makes it easy for programmers for plotting .

Scikit-learn



- Scikit-learn is one of the most popular ML libraries for classical ML algorithms.
- Scikit-learn supports most of the supervised and unsupervised learning algorithms

TensorFlow



- TensorFlow is a popular open-source library for high performance numerical computation.
- It can train and run deep neural networks that can be used to develop several AI applications. TensorFlow is widely used in the field of deep learning research and application.

ANY
QUESTIONS



THANK YOU