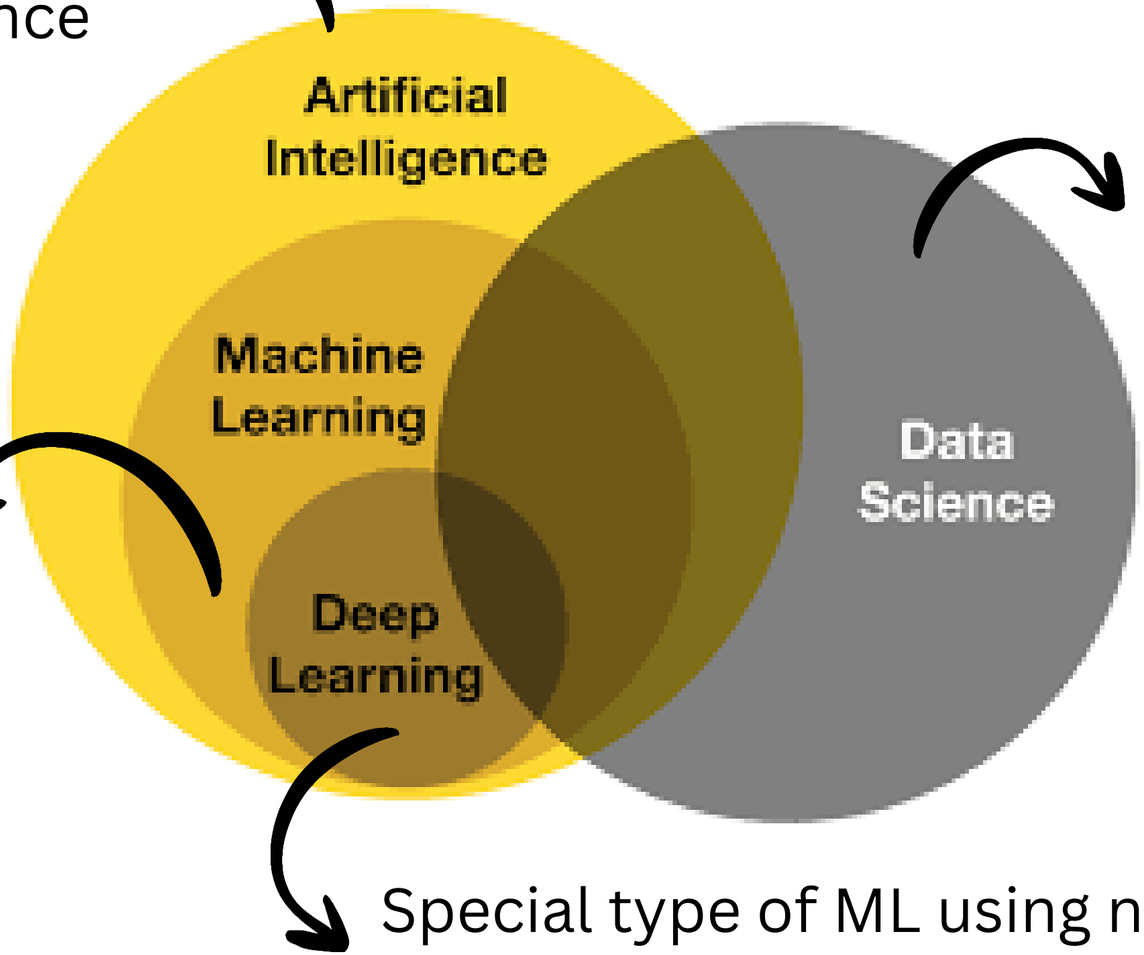
Will AI take our jobs?

Machine mimicking human intelligence



Working with data (cleaning, analyzing, visualizing).

Training machines on data so they learn patterns.

Special type of ML using neural networks (like our brain).

ML Pipeline

1. Ingredients

1. Data Collection

ML Pipeline

1. Ingredients
2. Wash & chop vegetables

1. Data Collection
2. Data Processing

ML Pipeline

1. Ingredients
2. Wash & chop vegetables
3. Adding spice

1. Data Collection
2. Data Processing
3. Feature Engineering

ML Pipeline

1. Ingredients
2. Wash & chop vegetables
3. Adding spice
4. Cooking

1. Data Collection
2. Data Processing
3. Feature Engineering
4. Model Training.

ML Pipeline

1. Ingredients
2. Wash & chop vegetables
3. Adding spice
4. Cooking
5. Tasting

1. Data Collection
2. Data Processing
3. Feature Engineering
4. Model Training.
5. Evaluate

ML Pipeline

1. Ingredients
2. Wash & chop vegetables
3. Adding spice
4. Cooking
5. Tasting
6. Adding salt if needed

1. Data Collection
2. Data Processing
3. Feature Engineering
4. Model Training.
5. Evaluate
6. Tuning

ML Pipeline

1. Ingredients
2. Wash & chop vegetables
3. Adding spice
4. Cooking
5. Tasting
6. Adding salt if needed
7. Serving food to friends

1. Data Collection
2. Data Processing
3. Feature Engineering
4. Model Training.
5. Evaluate
6. Tuning
7. Deploy

What happens if you don’t clean your vegetables properly?

Key Concepts

Overfitting

Overfitting happens when a model learns too much from the training data, including details that don't matter (like noise or outliers).

Underfitting

Underfitting is the opposite of overfitting. It happens when a model is too simple to capture what's going on in the data.

Note: The underfitting model has High bias and low variance.

Is it a good practice to always increase model complexity?

Key Concepts

Confusion Matrix summarizes classifier results with 4 values

Imagine you are a doctor diagnosing whether someone has COVID or not.
You can be right or wrong in 4 ways:
True Positive, **True Negative**, **False Positive**, **False Negative**.

Scenario: COVID Testing of 100 People

Out of 100 people:
40 people actually have COVID (**actual positives).**
60 people are healthy **(actual negatives)**.

Now suppose the test works like this:
It correctly detects 30 sick people (**TP = 30).**
It misses 10 sick people **(FN = 10).**
It wrongly marks 5 healthy people as sick **(FP = 5)**.
It correctly says 55 people are healthy **(TN = 55).**

Key Concepts

Precision & Recall

explains efficiency of the model

- **Precision**- Out of all positive predictions(True Positives + False Positives), how many were correct?

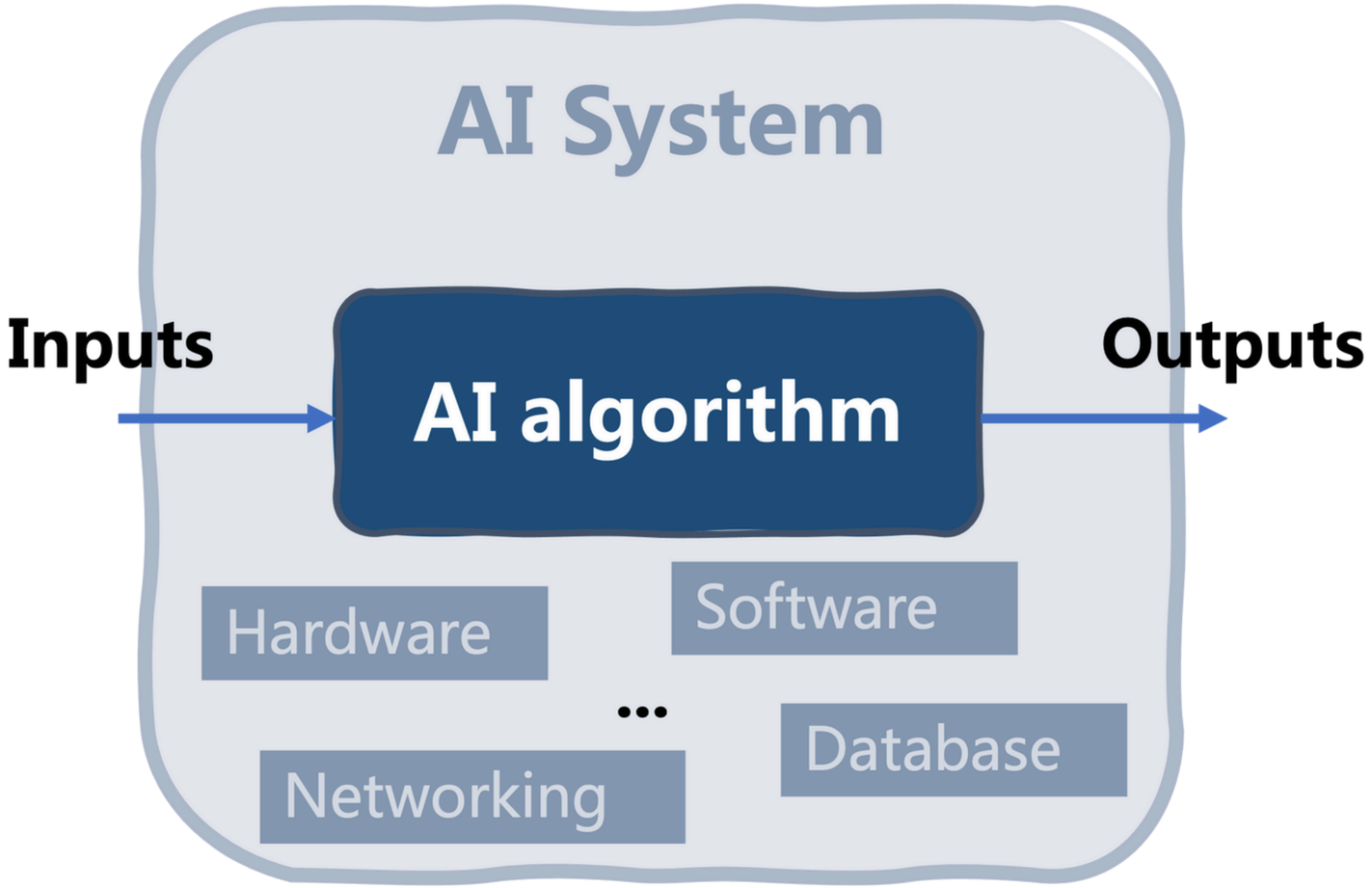
Focuses on minimizing false positives.
Calculation: TP/TP+FP

- **Recall**- Out of all actual positives(only True Positives), how many did you catch?

Focuses on minimizing false negatives.
Calculaion: TP/TP+FN

Algorithm and AI Systems

Are AI Algorithms and AI Systems the same?


AI System
Inputs
AI algorithm
Outputs
Hardware
Software
...
Database
Networking

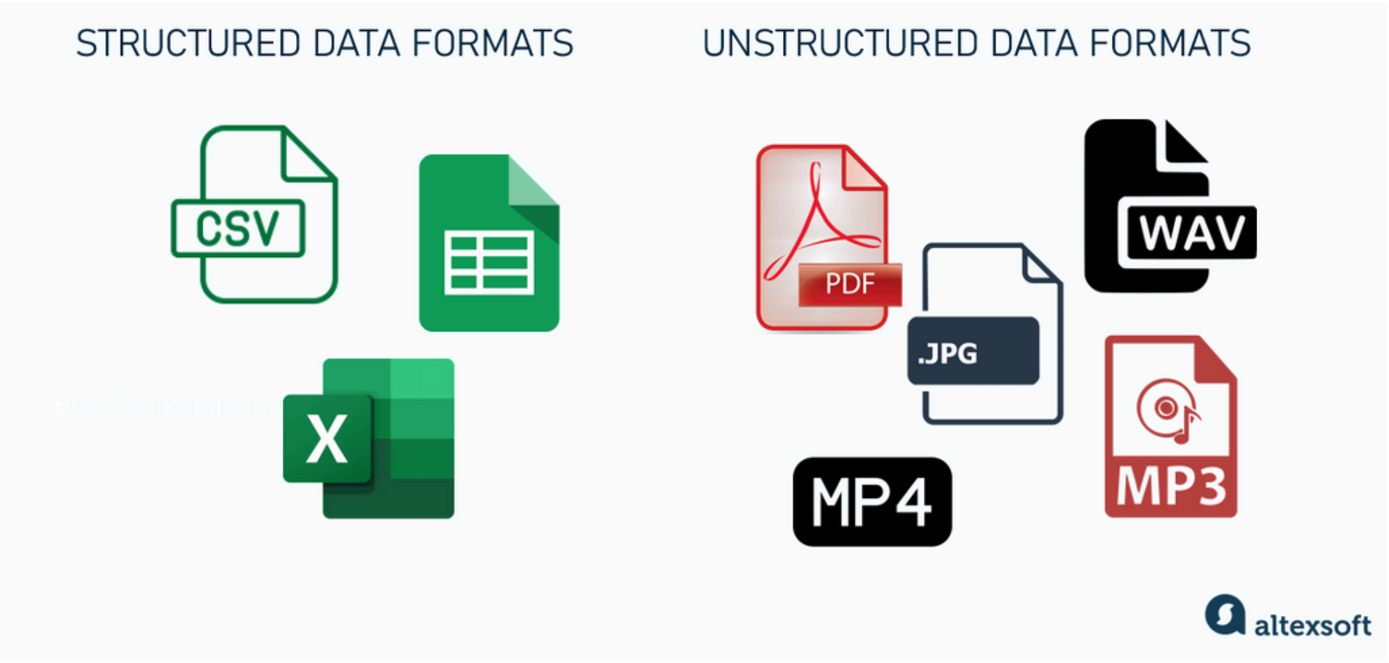
Everything you need to know about DATA

- **Data acquisition**
- **Data inspection**
- **Summary statistics**
- **Data visualisation**

Splitting the Dataset:

- Training Dataset
- Testing Dataset

Types of Data: Structured and Unstructured



Types of ML Algorithms

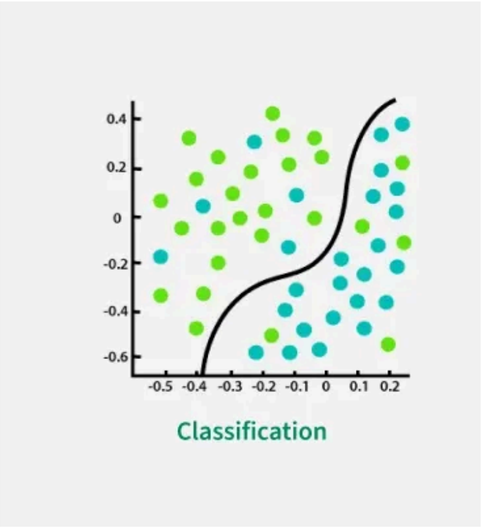
Supervised Learning

- Trained on Labelled Dataset
- Input-Output pairs provided during training

Classification Model

- predicts discrete labels
- O/P variable is a category : such as spam or not spam.

eg: Spam detection



Types of ML Algorithms

Supervised Learning

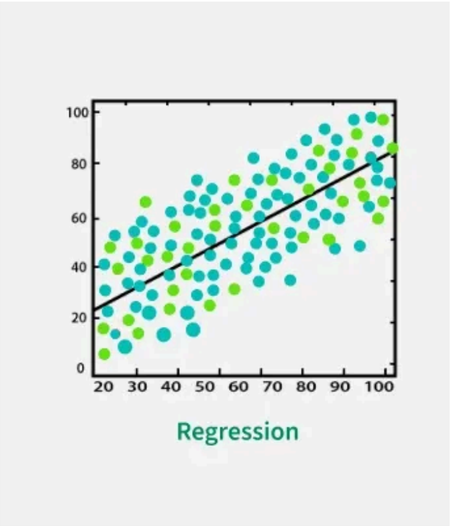
- Trained on Labelled Dataset
- Input-Output pairs provided during training

Some examples of Regression algorithms are: Linear Regression, Random-Forest, Decision Tree

Regression Model

- predicts continuous O/P
- O/P variable is a real value.

eg: Predicting house prices



Any Limitations?

Types of ML Algorithms

Unsupervised Learning

Training models on **unlabeled data** to discover patterns or structures within the data.

- Data exploration
- Visualization
- Dimensionality reduction
- Anomaly detection

Clustering

Aims to group similar data points together into clusters
eg: K-means clustering algorithm

Types of ML Algorithms

Unsupervised Learning

Training models on **unlabeled data** to discover patterns or structures within the data.

- Data exploration
- Visualization
- Dimensionality reduction

Association

Association rule learning is a technique for discovering relationships between items in a dataset.

Types of ML Algorithms

Reinforcement Learning

Here the algorithm learns from interaction with an environment.

- It corrects itself through trial and error techniques.
- Keeps on increasing its performance using Reward Feedback to learn behavior or pattern

Examples:
Self driving cars, AlphaGo

Quick run through model families and when to use them.

Linear models: linear regression, logistic regression — use for simple relationships.
Tree-based: Decision Trees, Random Forests — good for mixed numerical/categorical data and interpretability.
Example: loan approval decisions (if income > X and credit score > Y → approve).
KNN: predict label based on nearest neighbors — good for small datasets, explainable.
Ensembles: combine multiple models for better performance — Random Forests average many trees.
Deep Learning: images, speech — big data & compute.