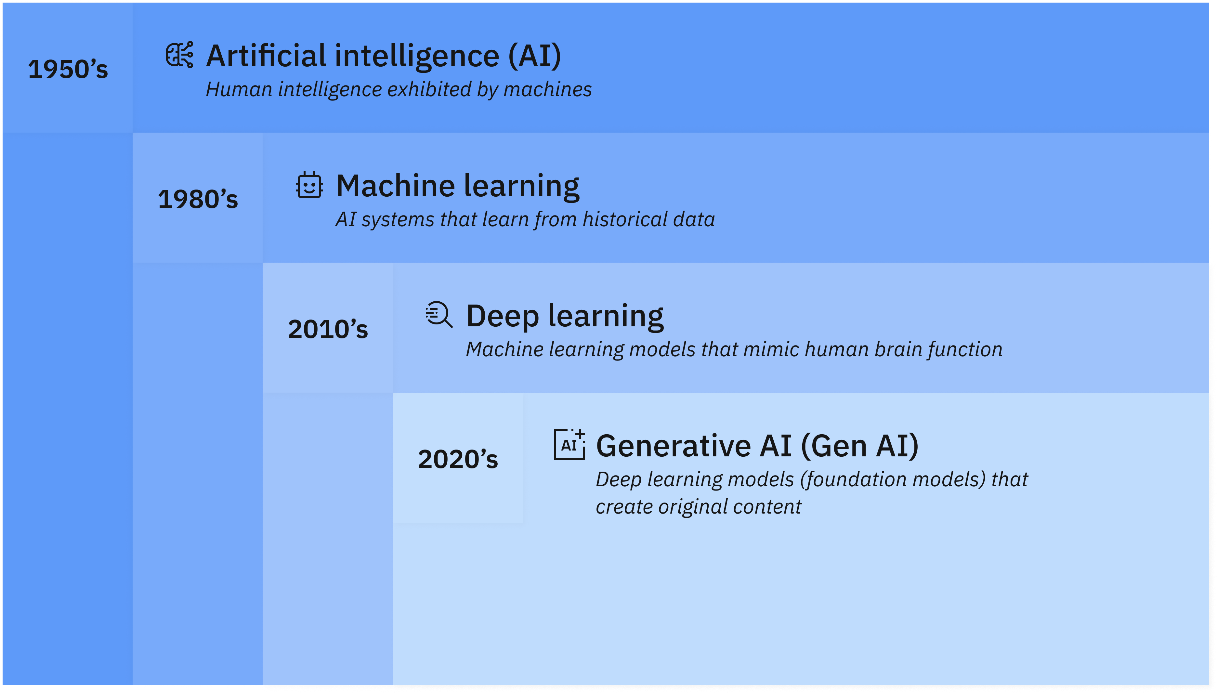
**ARTIFICIAL INTELLIGENCE A**

Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.



**Goals of AI**

* Replicate human intelligence
* Solve Knowledge-intensive tasks
* An intelligent connection of perception and action
* Building a machine which can perform tasks that requires human intelligence such as:
  + Proving a theorem
  + Playing chess
  + Plan some surgical operation
  + Driving a car in traffic
* Creating some system which can exhibit intelligent behaviour, learn new things by itself, demonstrate, explain, and can advise to its user.

**Advantages of AI**

* Accuracy with less errors: AI machines or systems are prone to less errors and high accuracy as it takes decisions as per pre-experience or information.
* Speed: AI systems can be of very high-speed and fast-decision making, because of that AI systems can beat a chess champion in the Chess game.
* Reliability: AI machines are highly reliable and can perform the same action multiple times with high accuracy.
* Useful for risky areas: AI machines can be helpful in situations such as defusing a bomb, exploring the ocean floor, where employing a human can be risky.
* Digital Assistant: AI can be very useful to provide digital assistance to the users Eg. E-commerce websites to show the products as per customer requirement.
* Useful as a public utility: AI can be very useful for public utilities such as a self-driving car which can make our journey safer and hassle-free, facial recognition for security purpose, Natural language processing to communicate with the human in human-language, etc.

**AI challenges and risks**

* Data risks
  + AI systems rely on data sets that might be vulnerable to data poisoning, data tampering, data bias or cyberattacks that can lead to data breaches. Organizations can mitigate these risks by protecting data integrity and implementing security and availability throughout the entire AI lifecycle, from development to training and deployment and post-deployment.
* Model risks
  + Threat actors can target AI models for theft, reverse engineering or unauthorized manipulation. Attackers might compromise a model’s integrity by tampering with its architecture, weights or parameters; the core components that determine a model’s behaviour, accuracy and performance.
* Operational risks
  + Like all technologies, models are susceptible to operational risks such as model drift, bias and breakdowns in the governance structure. Left unaddressed, these risks can lead to system failures and cybersecurity vulnerabilities that threat actors can use.

**AI ethics and governance**

* Explainability and interpretability
  + As AI becomes more advanced, humans are challenged to comprehend and retrace how the algorithm came to a result. Explainable AI is a set of processes and methods that enables human users to interpret, comprehend and trust the results and output created by algorithms.
* Fairness and inclusion
  + Although machine learning, by its very nature, is a form of statistical discrimination, the discrimination becomes objectionable when it places privileged groups at systematic advantage and certain unprivileged groups at systematic disadvantage, potentially causing varied harms. To encourage fairness, practitioners can try to minimize algorithmic bias across data collection and model design, and to build more diverse and inclusive teams.
* Robustness and security
  + Robust AI effectively handles exceptional conditions, such as abnormalities in input or malicious attacks, without causing unintentional harm. It is also built to withstand intentional and unintentional interference by protecting against exposed vulnerabilities.
* Accountability and transparency
  + Organizations should implement clear responsibilities and governance structures for the development, deployment and outcomes of AI systems. In addition, users should be able to see how an AI service works, evaluate its functionality, and comprehend its strengths and limitations. Increased transparency provides information for AI consumers to better understand how the AI model or service was created.
* Privacy and compliance
  + Many regulatory frameworks, including GDPR, mandate that organizations abide by certain privacy principles when processing personal information. It is crucial to be able to protect AI models that might contain personal information, control what data goes into the model in the first place, and to build adaptable systems that can adjust to changes in regulation and attitudes around AI ethics.

**K-MEANS ALGORITHM a**

* [WCSS](https://medium.com/@sreeku.ralla/wcss-how-many-clusters-are-good-enough-74f91c06dc75)
* [Choosing K](https://medium.com/@nirmalsankalana/k-means-clustering-choosing-optimal-k-process-and-evaluation-methods-2c69377a7ee4#:~:text=Elbow%20Method,like%20shape%20in%20the%20plot.)
* [Manhattan Distance](https://www.datacamp.com/tutorial/manhattan-distance)
* [K-Means clustering](https://neptune.ai/blog/k-means-clustering)
* [Euclidean distance](https://en.wikipedia.org/wiki/Euclidean_distance)

**NORMALIZATION a**

Normalization is a data preprocessing technique used to transform the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values or losing information. It’s about adjusting the scale of your data to level the playing field for all the features in your dataset.

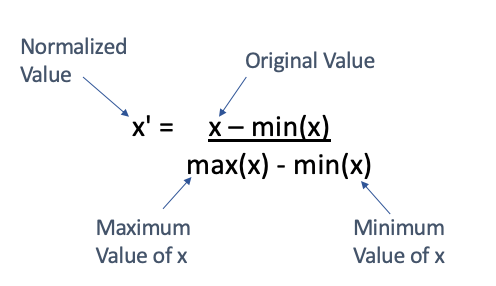
**WHY NORMALIZE DATA?**

* **Improving Model Accuracy:** Many machine learning algorithms, like gradient descent, converge faster with normalized data. Without normalization, features with higher magnitude can dominate the learning process, leading to less accurate models.
* **Facilitating Model Training:** Normalized data helps ensure that each feature contributes equally to the model training process, making it easier for the model to learn the patterns.
* **Enhancing Compatibility:** Some algorithms, especially those involving distance calculations like k-nearest neighbors (KNN) and k-means clustering, require normalized data to function correctly because they are sensitive to the magnitude of the data.

**MIN-MAX NORMALIZATION**

It is one of the simplest methods where the values are scaled to a fixed range — usually 0 to 1.

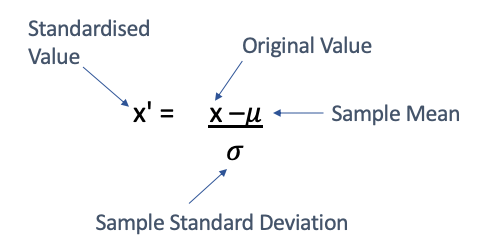
The formula for calculating the Min-Max Scaling is:



are the minimum and maximum values in the feature, respectively. This method is best used when the distribution is not Gaussian or when the standard deviation is very small. However, it’s sensitive to outliers.

**STANDARD NORMALIZATION**

Standardization (or Z-Score Normalization) transforms the features so they have the properties of a standard normal distribution with a mean of 0 and a standard deviation of 1:



This method is less affected by outliers and is suitable for algorithms that assume the input data is normally distributed.

**IMPORTANT CONCEPTS [KNOW THE MATH BEHIND IT – BASIC CONCEPTS] a**

* Min-max normalization
* Standard normalization [z-score]
* Linear correlation coefficient (r)
  + Covariance
* Probability distribution
  + Gaussian distribution
  + Poisson distribution
* Correlation matrix

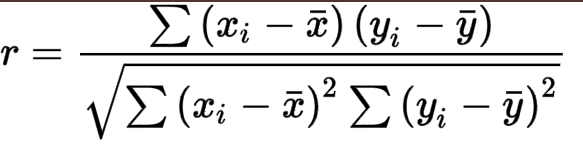
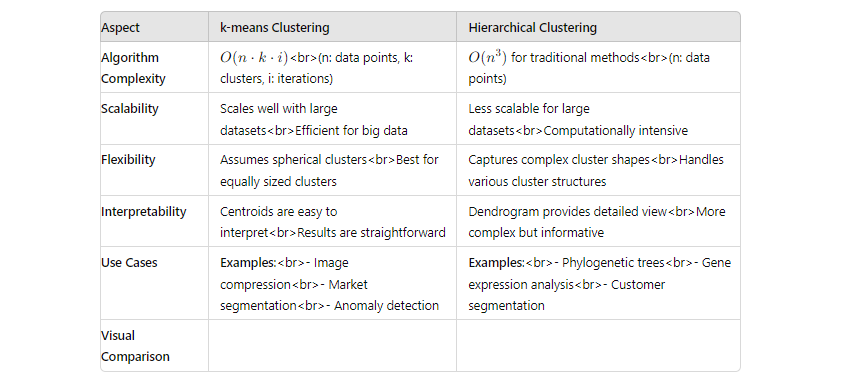


Figure LINEAR CORRELATION COEFFICIENT

* NEGATIVE correlation doesn’t mean bad. It just means it has a negative slope. -1 is strongly correlated.

**QUESTIONS:**

1. is normalization important for k means and HAC?
   * Yes
2. Difference between HAC and K-means clustering
   * [](https://medium.com/@amit25173/k-means-clustering-vs-hierarchical-clustering-171c217b2968)

Click for more

1. Are HAC and K-means supervised or unsupervised?
   * Unsupervised.