



Data Science & Machine Learning with Python

Batch no: 19



Visit Our Website
www.aiquest.org



What, Why, How?

Data Science



What is Data Science?

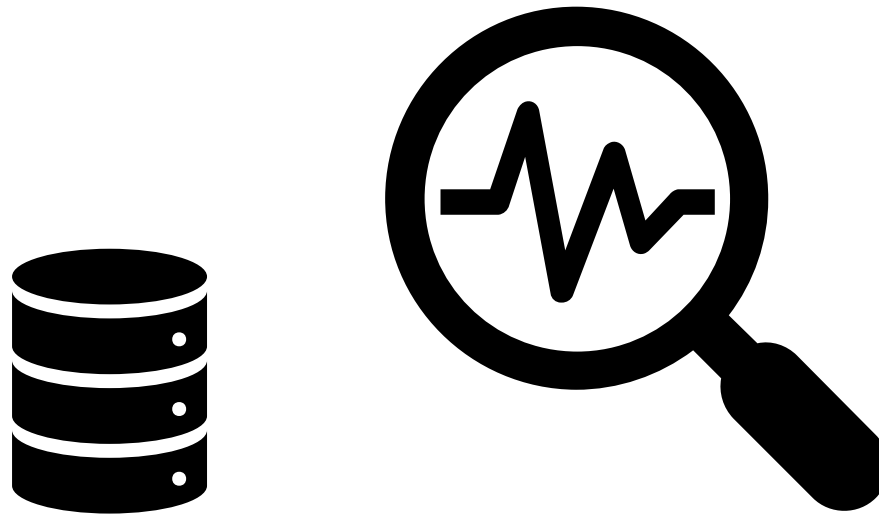
Why it is So Popular?

Does Data Science Have Future?

How Much Money Do Data Scientists Make?

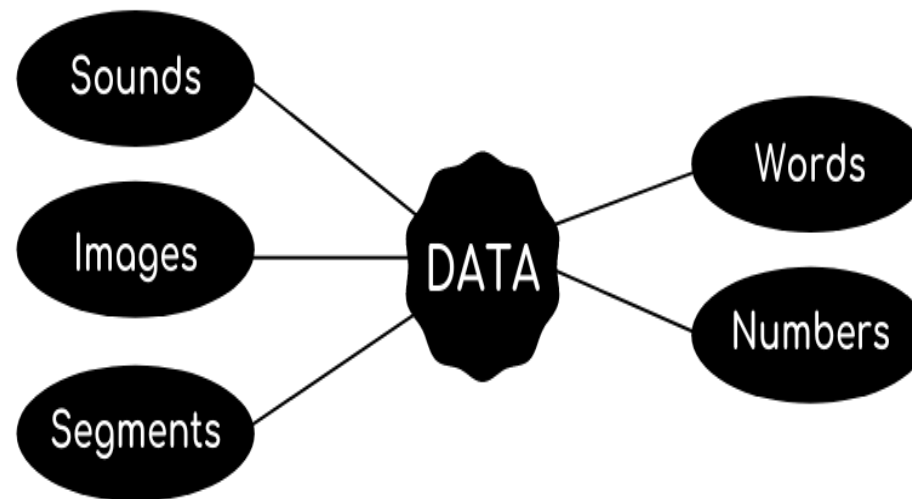


Let's talk about Data and Information



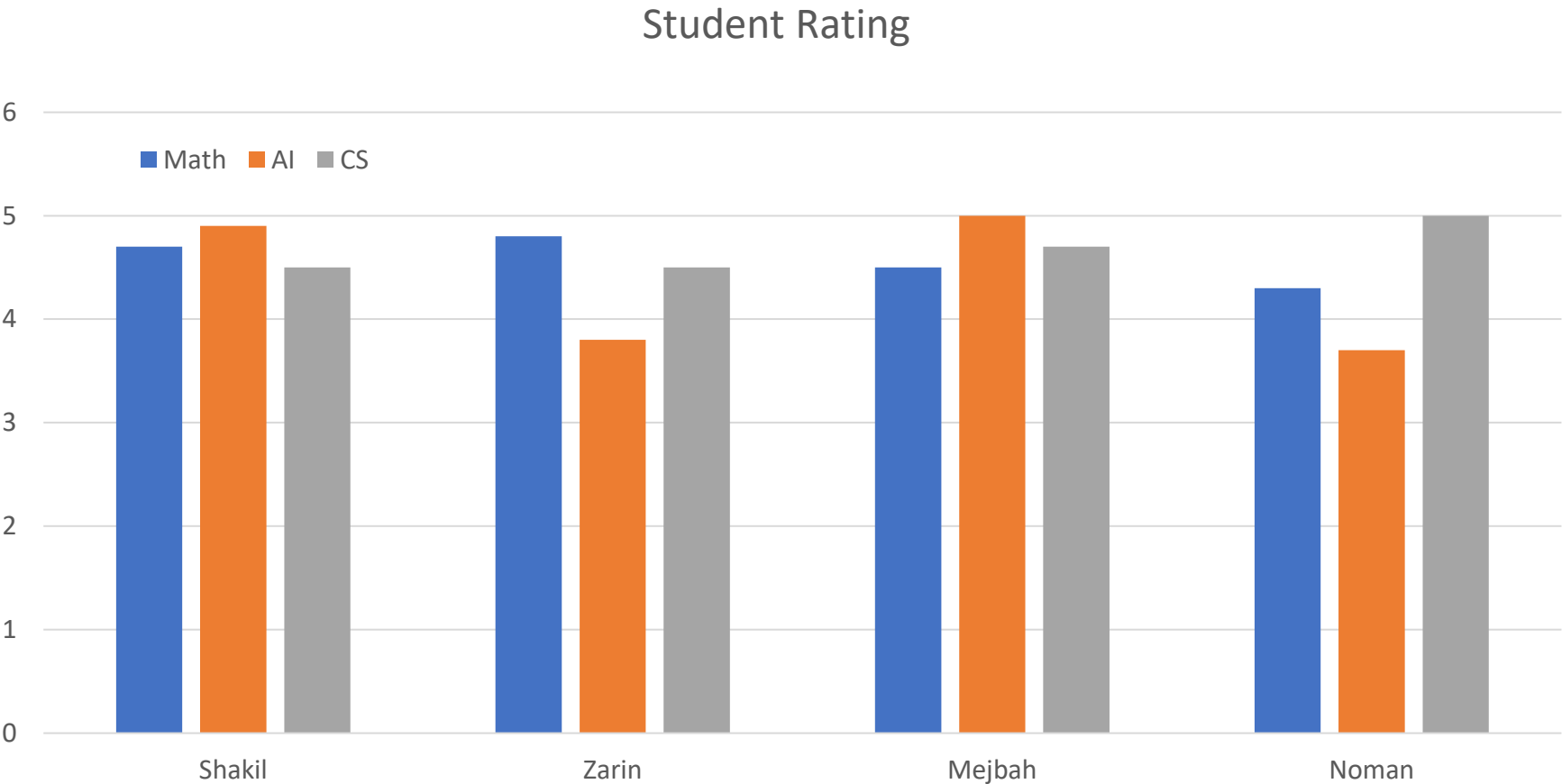
Data is defined as a collection of organized or unorganized facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, processing by humans, or some automatic means such as computers, ATMs.

The main examples of data are phone numbers, weights, prices, costs, number of items sold, product names, addresses, registration marks, etc.



What is Information?

Definition



What is Database?

Definition



A database is an organized collection of related data or information stored and accessed electronically within a computer system.

For example, MySQL, MongoDB, Oracle Database, PostgreSQL, etc. are all examples of different databases. These modern databases are managed by DBMS. Structured Query Language, or SQL as it is more widely known, is used to operate on the data in a database.

What is Data Warehouse?

Definition

A data warehouse is a centralized storage system that allows for the storing, analyzing, and interpreting of data to facilitate better decision-making.

- Snowflake
- Databricks
- Amazon Redshift
- Azure Synapse Analytics
- Google Big Query



How Much Data is in the World Today?



How Much Data is in the World Today?

1 How much data is generated every minute?

Source: Domo

 **41,666,667**

messages shared
by WhatsApp users

 **1,388,889**

video / voice calls made
by people worldwide

 **404,444**

hours of video streamed
by Netflix users

 **347,222**

stories posted by Instagram users

 **150,000**

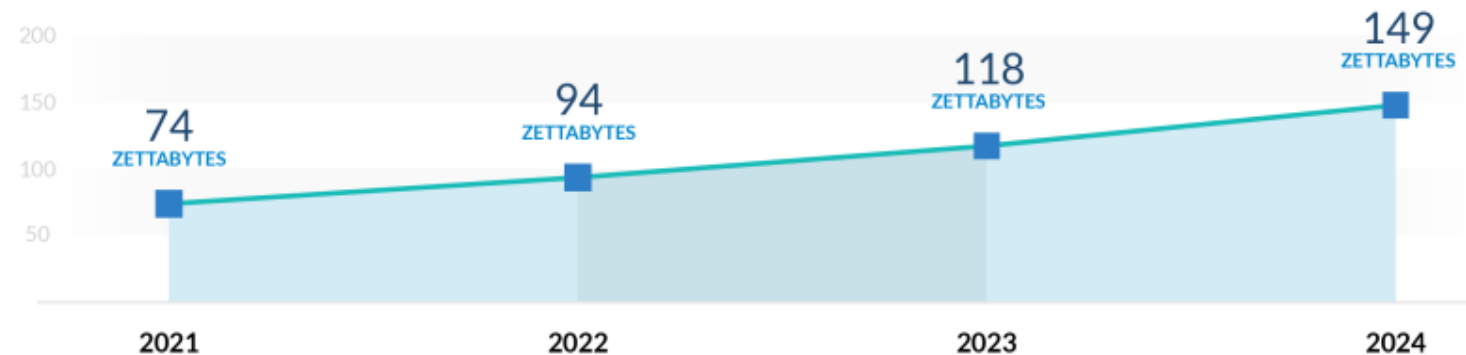
messages shared by Facebook users

 **147,000**

photos shared by Facebook users

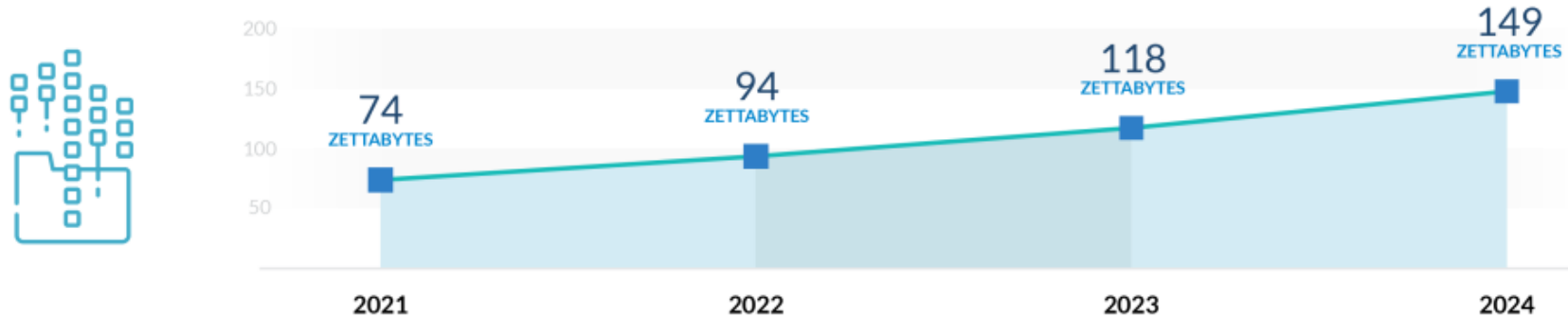
2 Estimated Data Consumption from 2021 to 2024

Source: IDC / Statista



2 Estimated Data Consumption from 2021 to 2024

Source: IDC / Statista



3 Data Growth in 2021

Sources: TechJury, Internet Live Stats, Cisco, PurpleSec

 **2 TRILLION**

searches on Google by the end of 2021

 **1.134 TRILLION MB**

volume of data created every day

 **3,026,626**

emails sent every second, 67% of which are spam

 **278,108 PETABYTES**

global IP data per month by the end of 2021

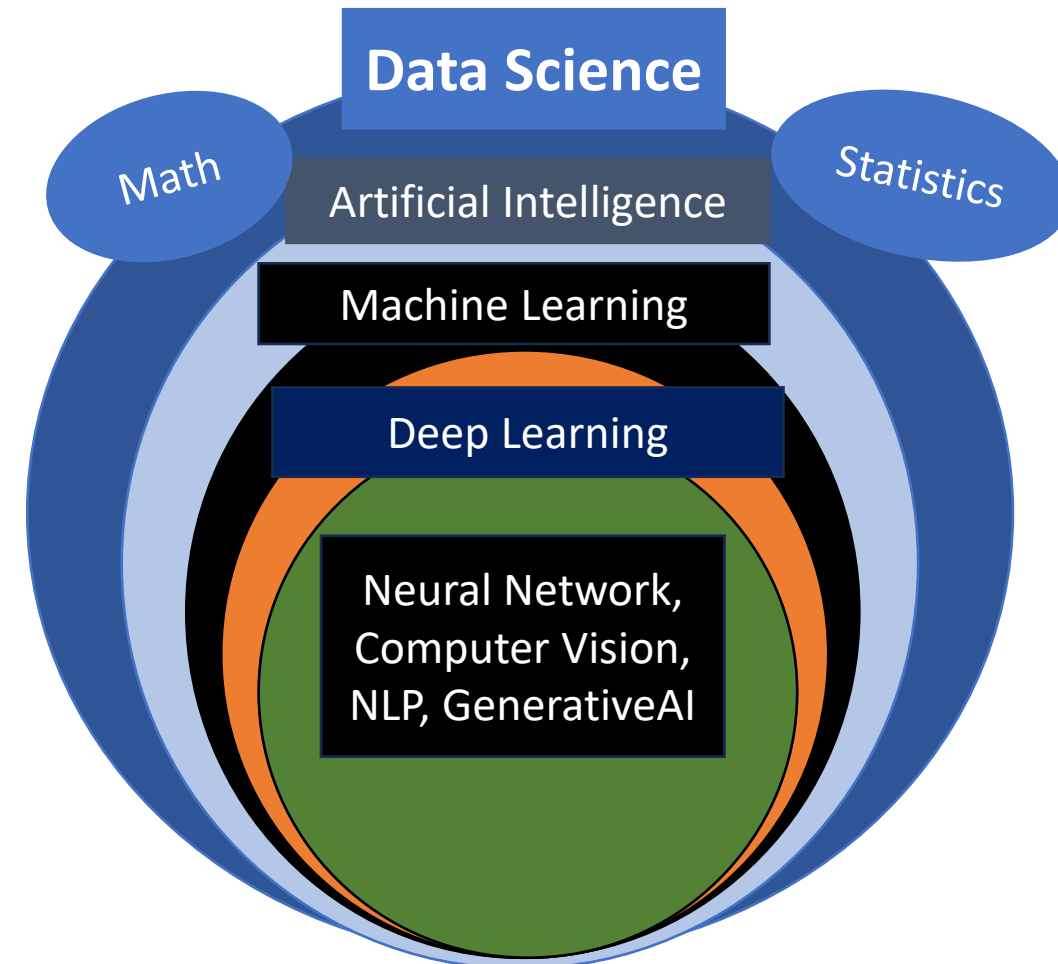
 **230,000**

new malware versions created every day

 **82%**

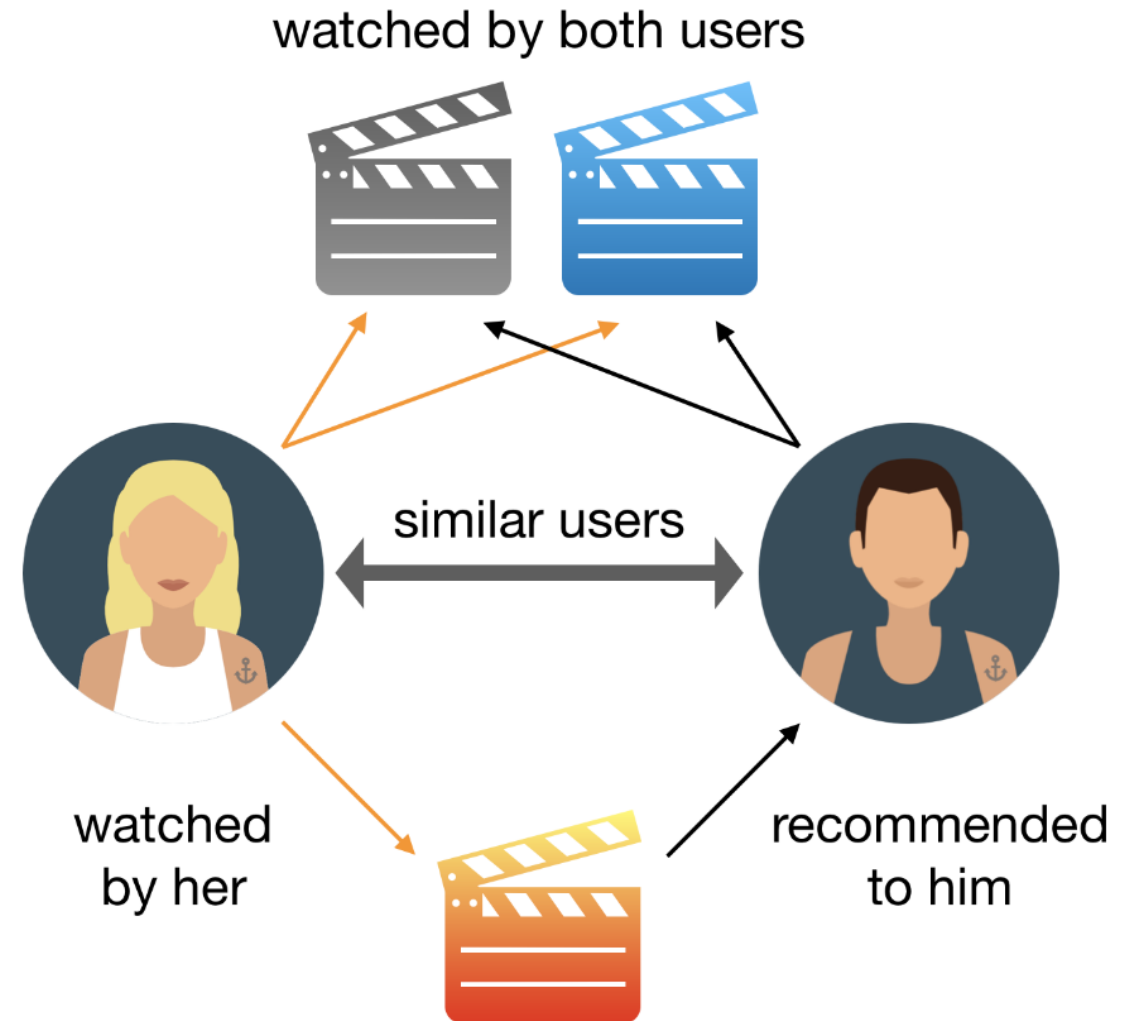
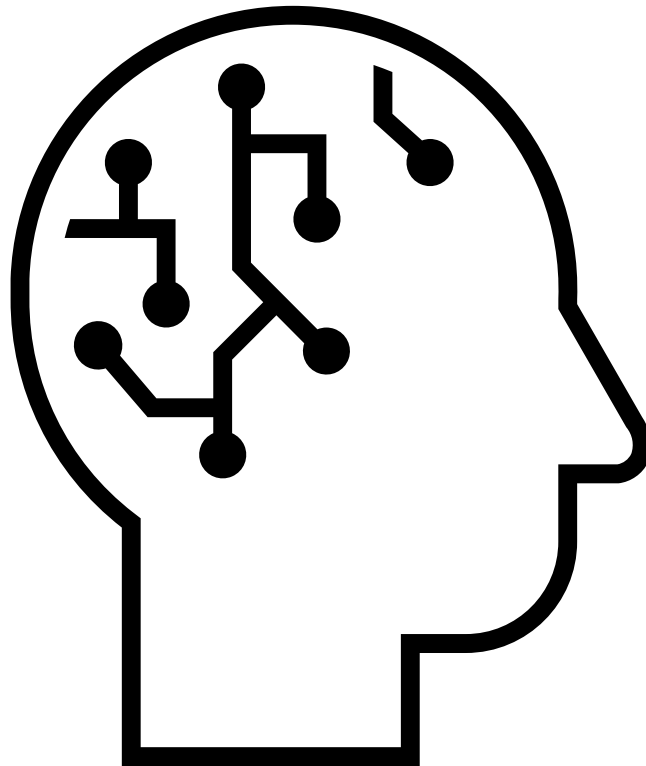
share of video in total global internet traffic at the end of 2021

- Computer Science
- Mathematics
- Statistics



What is AI?

Definition





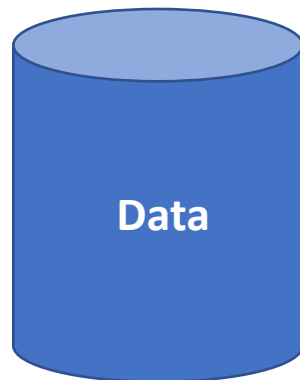
AI Chat Bot for your Business

IMPROVE YOUR
CUSTOMER SERVICE
WITH AI CHAT BOT.



What is Machine Learning?

$ML = Machine + Learning$



Data

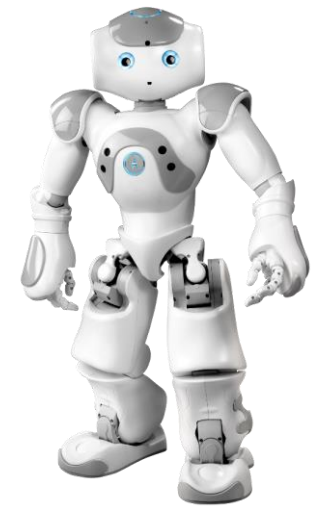
Training

Testing

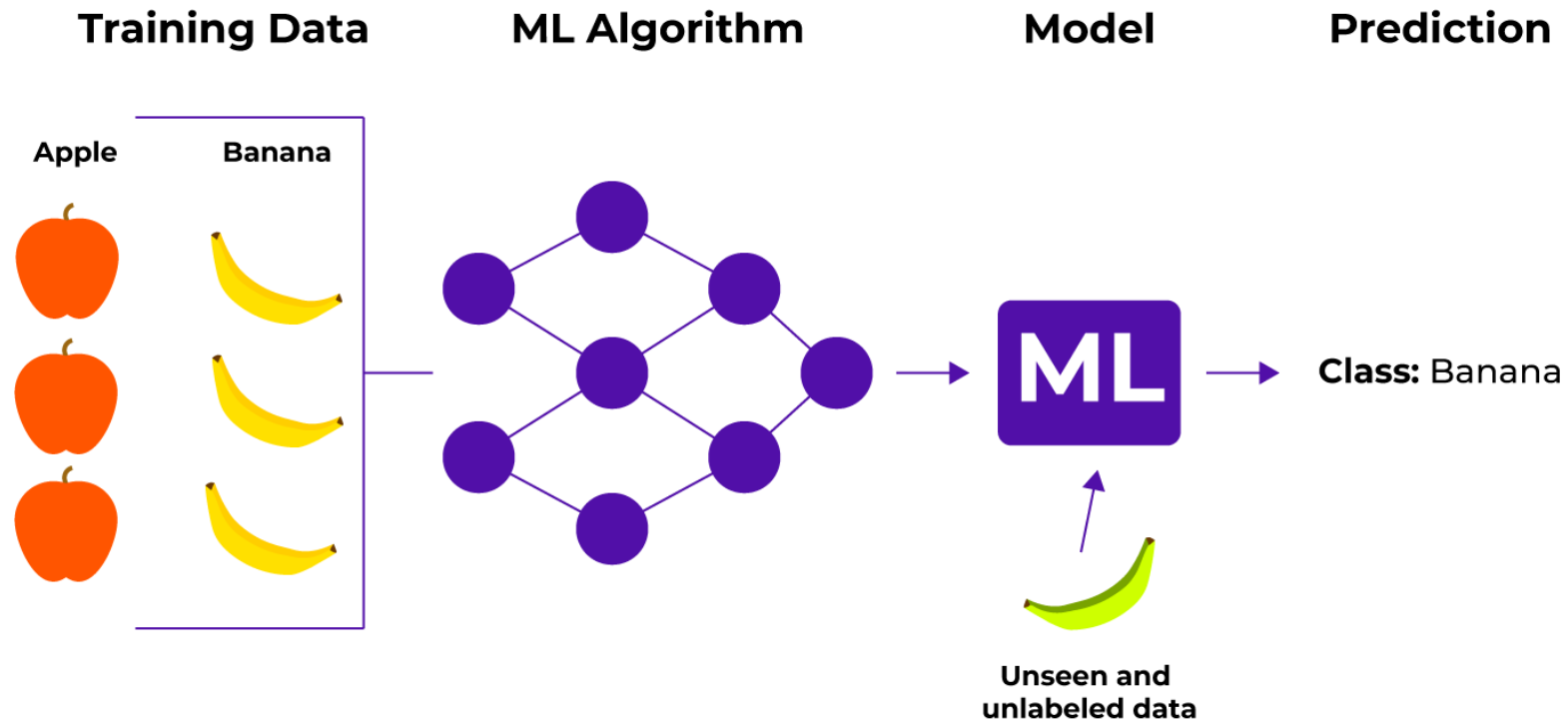


Create Artificial
Brain Using ML

Able to
Make Prediction



Artificial Intelligence



Types Machine Learning Model

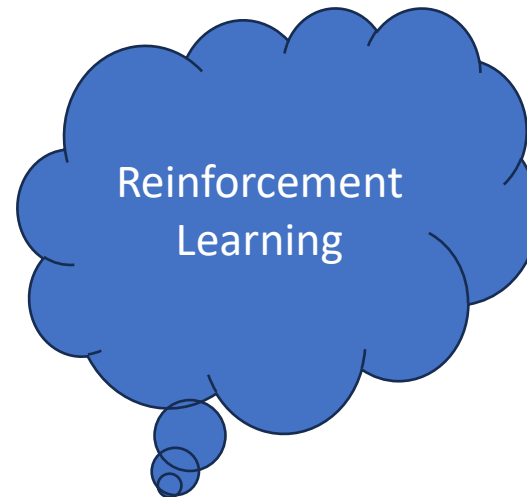
Generalized classification



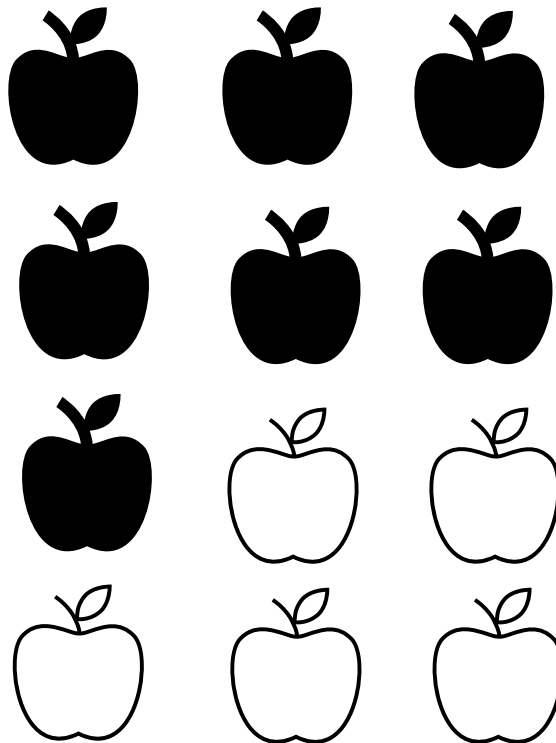
Discriminative Models



Generative Models



1. Classification Tasks

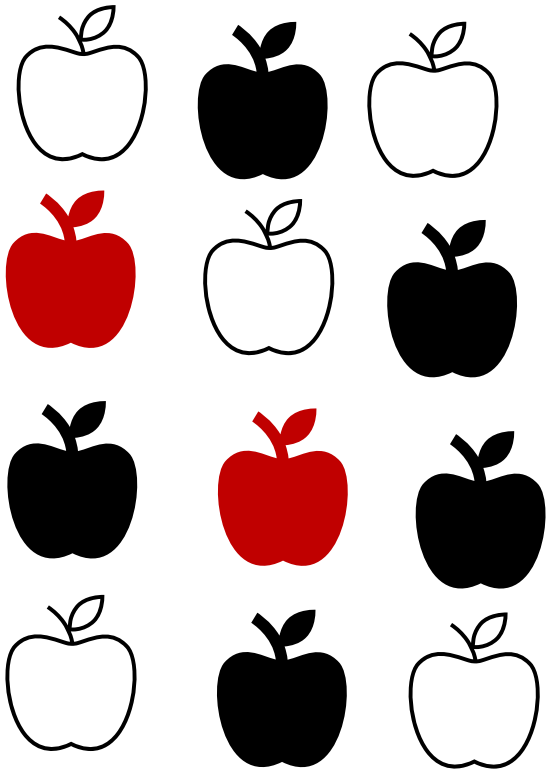


2. Regression Tasks

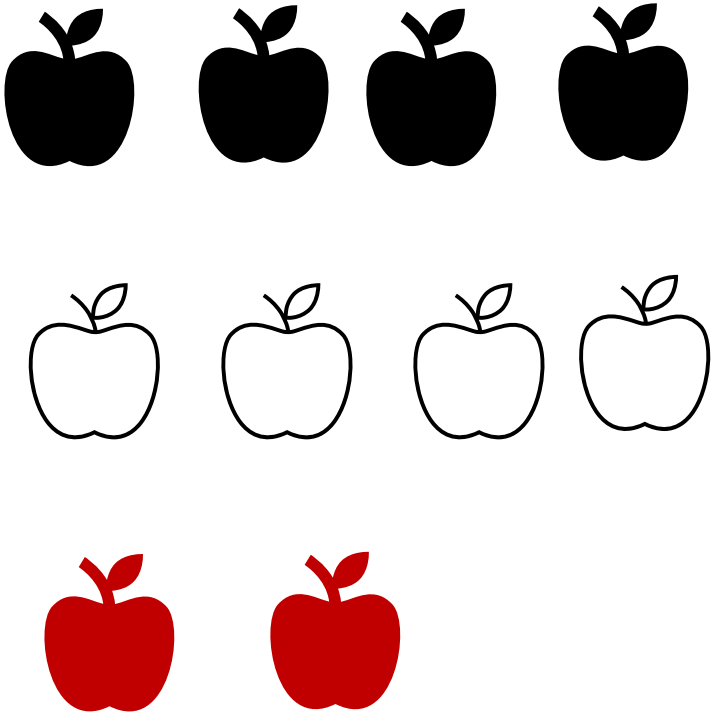


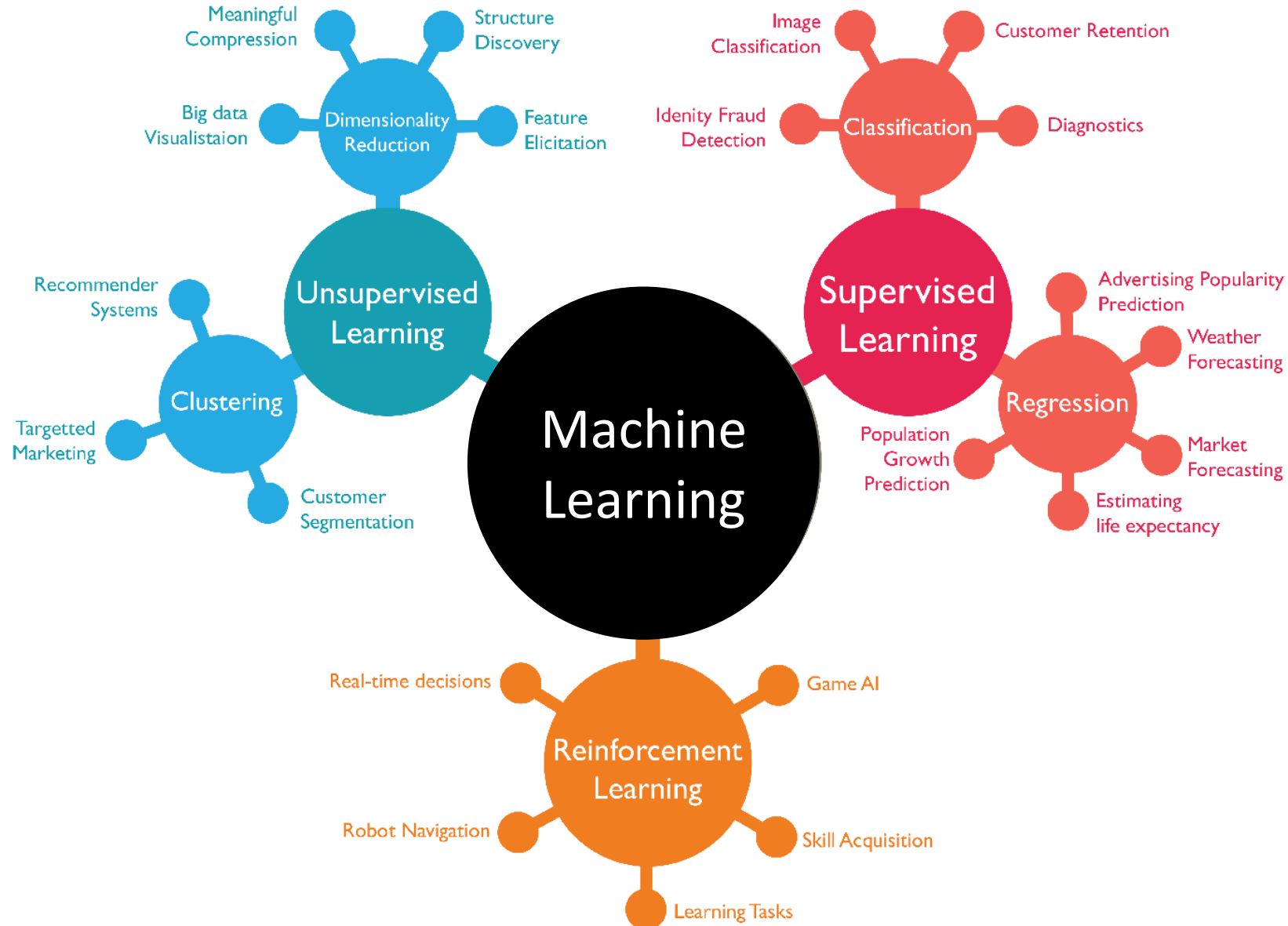


Before Cluster



After Cluster





1. Convolutional Neural Networks (CNNs)

2. Recurrent Neural Networks (RNNs)

3. Long Short-Term Memory (LSTM) Networks

4. Generative Adversarial Networks (GANs)

5. Variational Autoencoders (VAEs)

6. Transformer Networks

7. N-Gram Model

8. Bayesian Neural Networks (BNNs)



Fig: Training Model

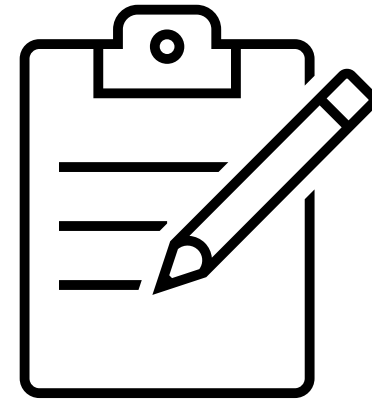
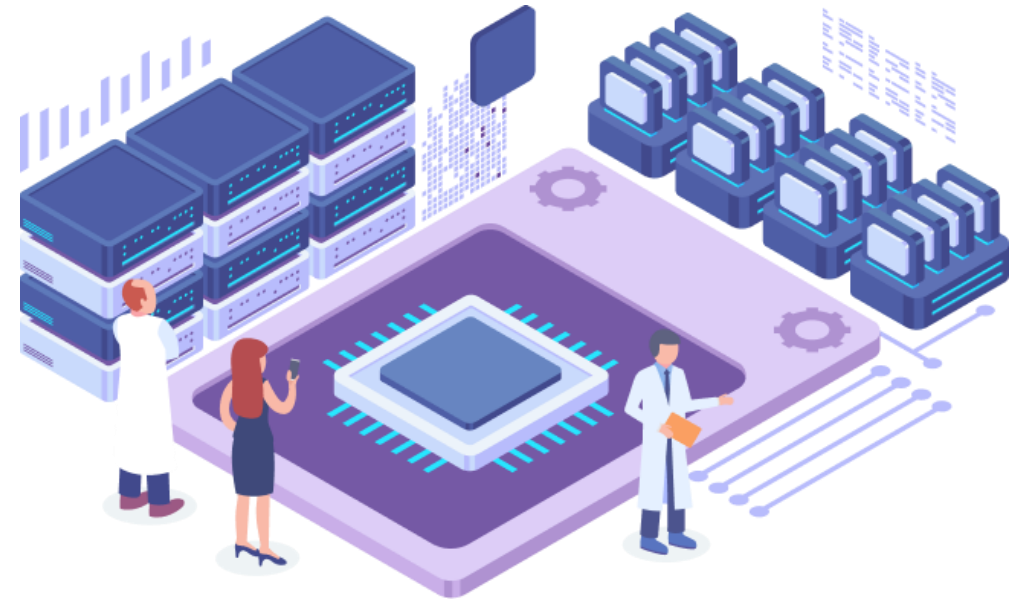


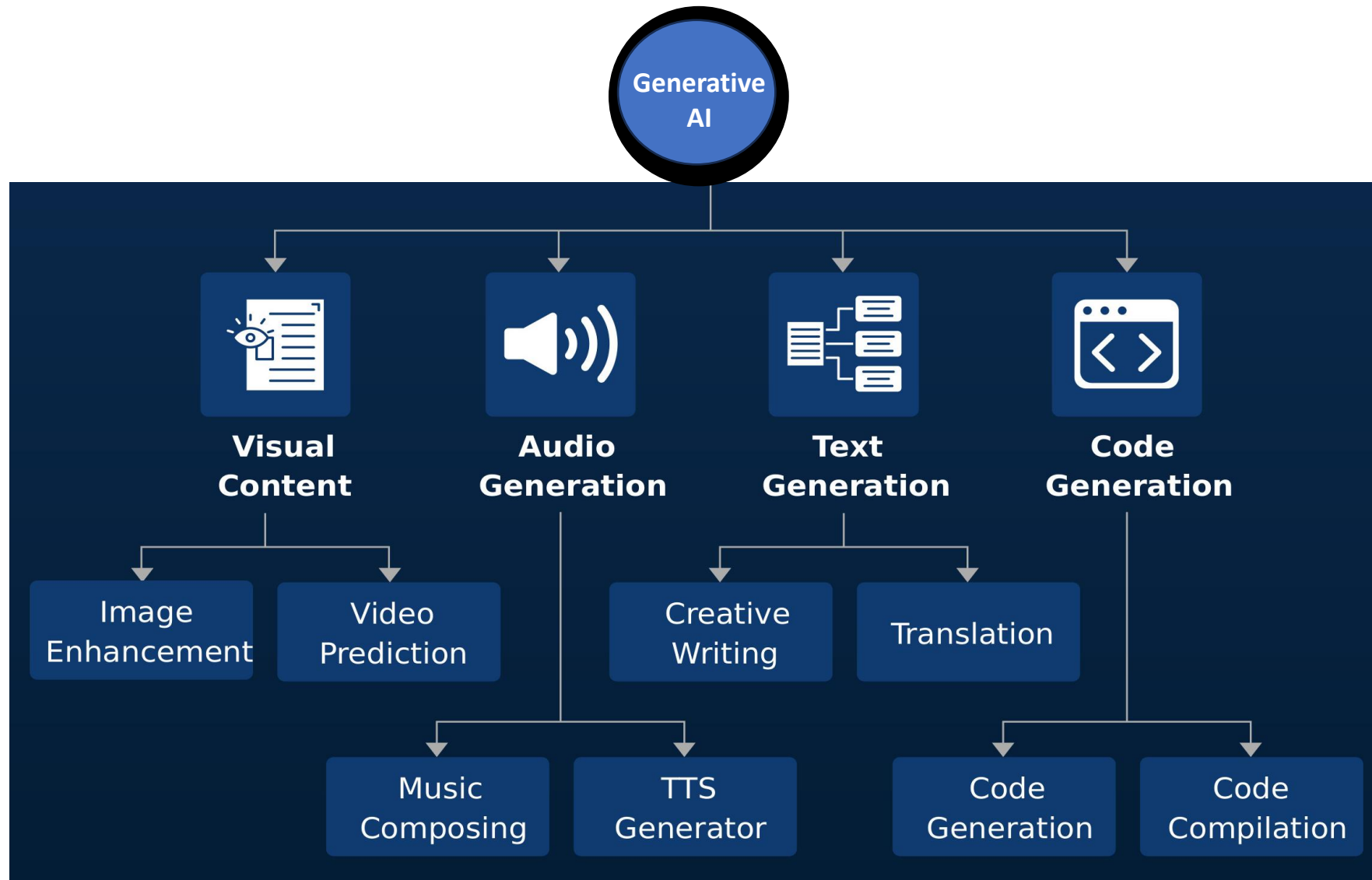
Fig: Generating Answer

Large Language Models: LLMs

Brief Discussion

Large language models (LLMs) are natural language processing computer programs that use artificial neural networks to generate text. Some notable ones are GPT-3, GPT-4, LaMDA (Bard), BLOOM, and LLaMA. LLMs power many applications, such as AI chatbots and AI search engines.



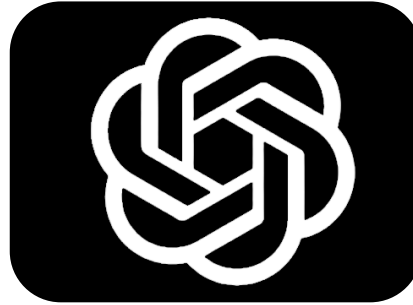


Generative AI Tools

Some Popular Tools



Chat-GPT



DALLE / DALLE 2



Google Bard AI



Midjourney AI

Recurrent Neural Networks (RNNs)

Long Short-Term Memory (LSTM) Networks

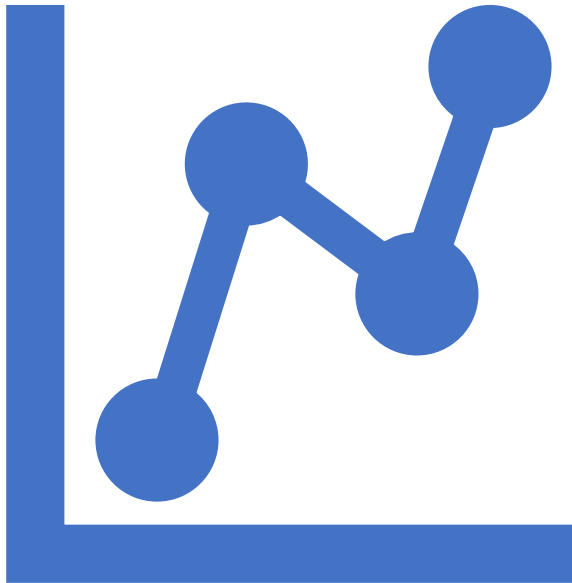
Generative Adversarial Networks (GANs)

Variational Autoencoders (VAEs)

Transformer Networks

N-Gram Model

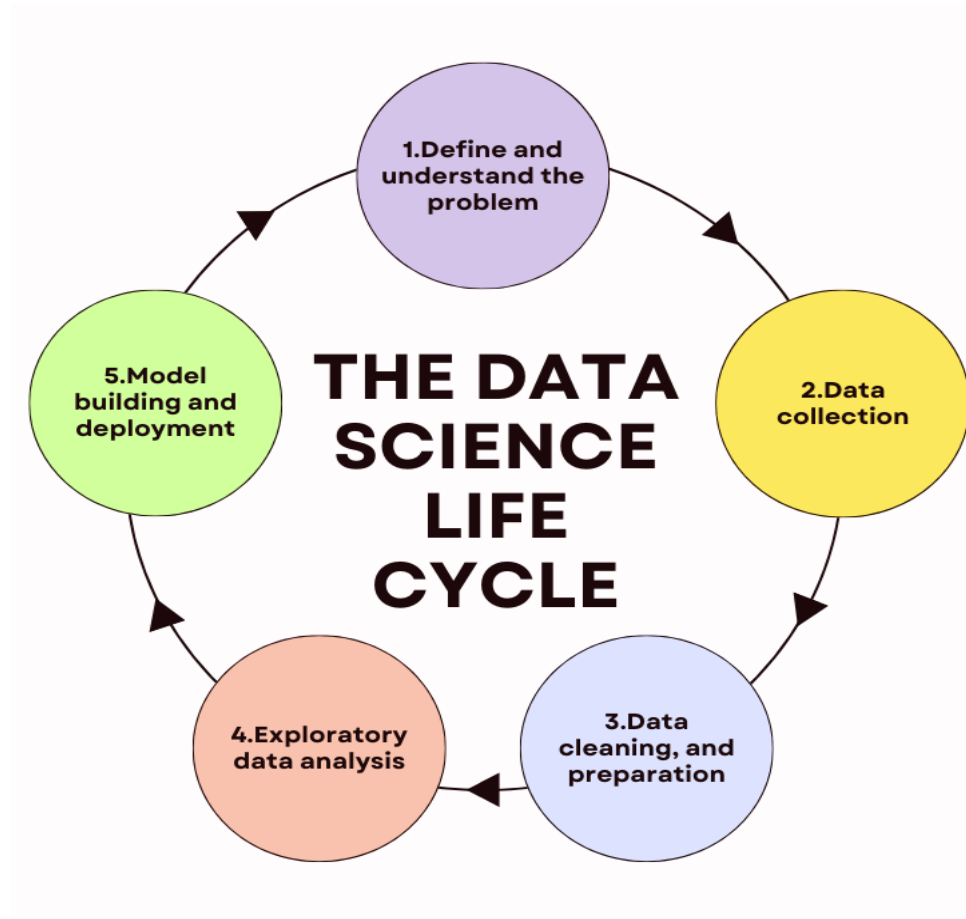
- Programming Languages: Python, Julia, R, etc.
- Statistics and Mathematics for Data Science
- Data Manipulation with Pandas
- SQL for Data Science
- Data Visualization (Excel, PowerBi)
- EDA & Feature Engineering
- Machine Learning (Statistical ML Algorithms)
- Deep Learning (Neural ML Algorithms)
- Git & GitHub
- Data Science & AI Libraries: Matplotlib, Seaborn, NumPy, Scikit-learn, Keras, TensorFlow, PyTorch
- Cloud Computing: AWS/ Azure/ Google Cloud
- Domain Knowledge
- Collaboration and Teamwork



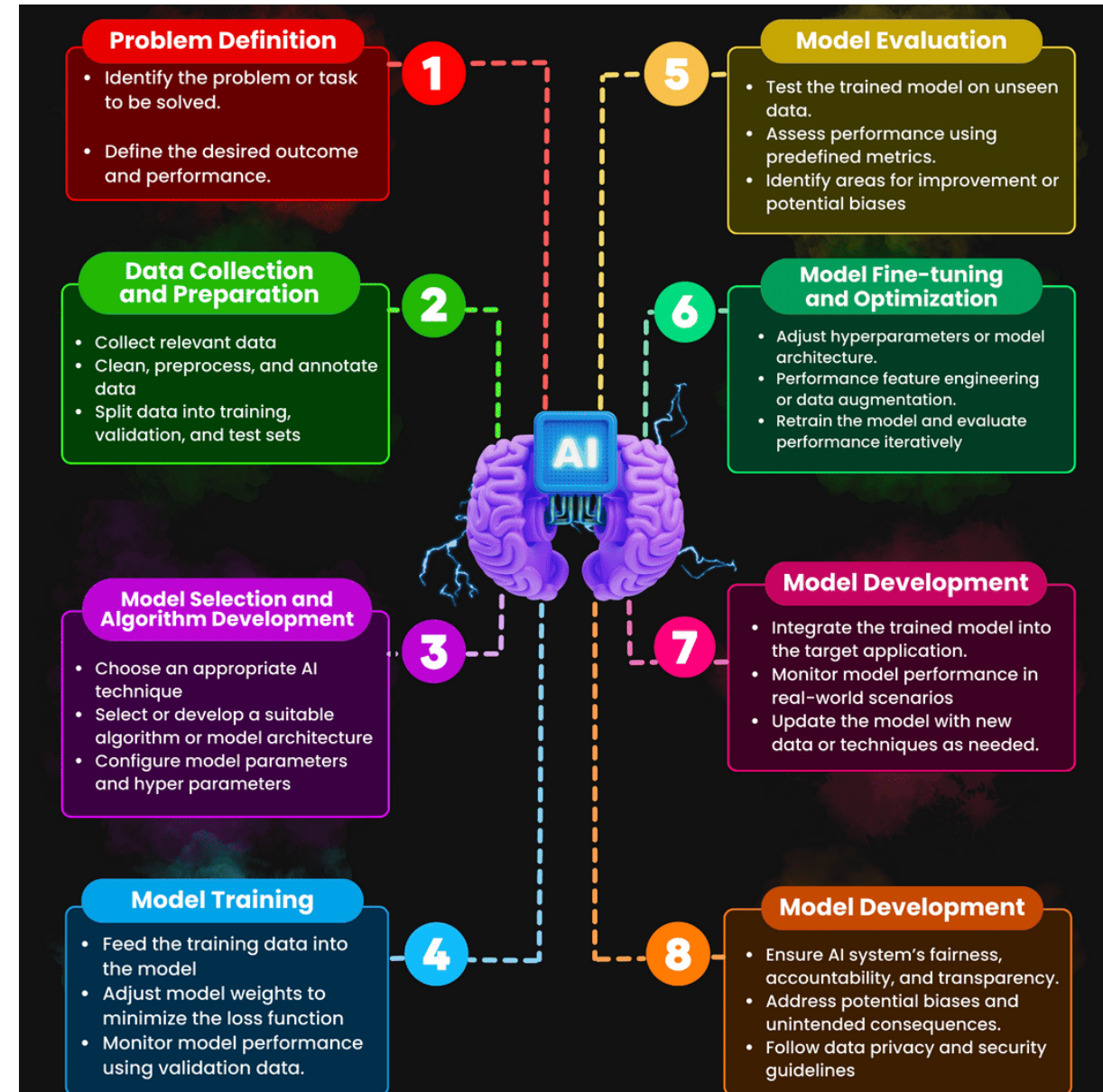
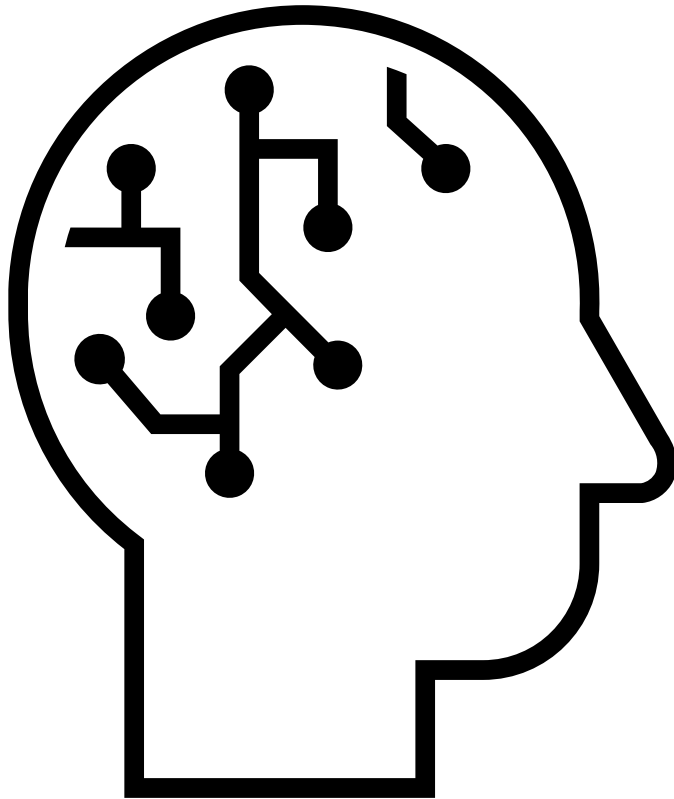
How to Use Data for a Smart Career?

How to Use Data for a Smart Career?



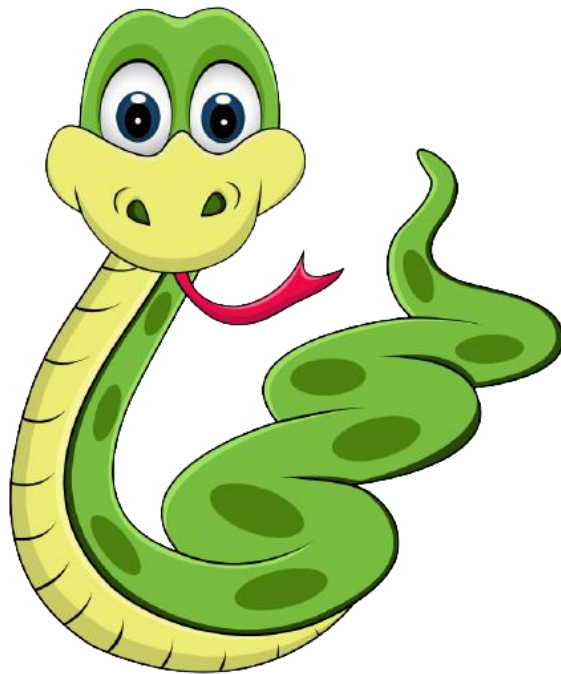


How to Build an AI Application



- Module 01: Introduction to Data Science & AI
- Module 02: Basics Python
- Module 03: Regression Algorithms
- Module 04: Feature Engineering
- Module 05: Basics of Linear Algebra
- Module 06: Statistics for Data Science
- Module 07: Classification Algorithms
- Module 08: ML Model Evaluation
- Module 09: Unsupervised Learning
- Module 10: Hyperparameter Optimization
- Module 11: Deep Learning & Neural Networks

[Download Module!](#)

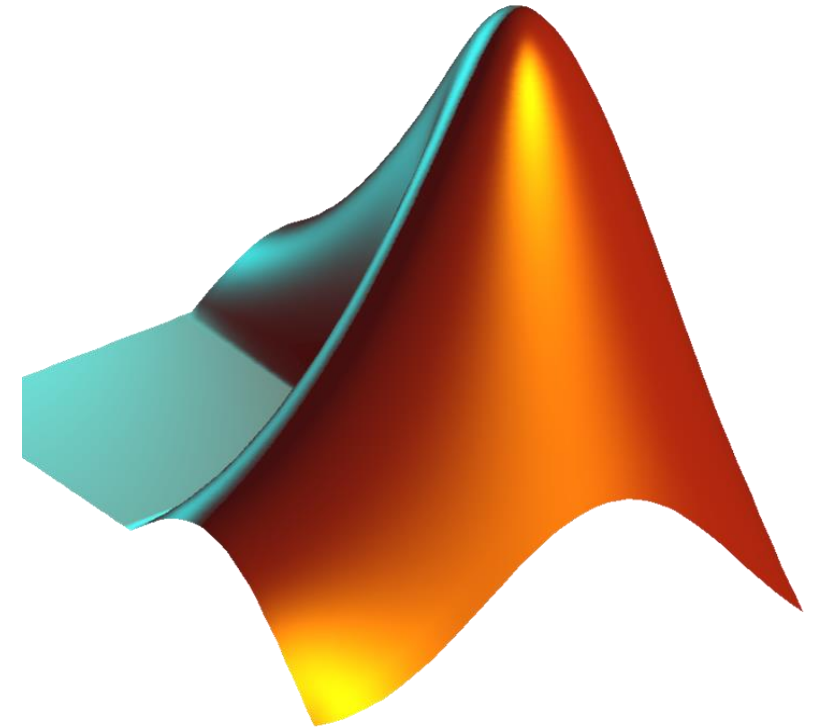


1. Easy Syntax, Flexible, Support OOP & Faster
2. Python has Machine Learning Libraries
3. Python has Data Analysis Library
4. Python has Data Frame Library
5. Python has Calculator Library
6. Python is Significant for Deep Learning
7. Keras, Tensorflow, Pytorch
8. Web (Django & Flask)
9. Open Resources

- Data Analysis
- Data Visualization
- Statistical Libraries
- Statistics and Research Methods



- Numerical Computing
- Matrix Operations
- Machine Learning Toolbox
- Deep Learning Support
- Data Visualization
- Simplicity and User-Friendly Interface
- Community Support
- Integration with Other Languages



- Data Scientist (\$139,840/year)
- Machine Learning Engineer (\$114,826/year)
- Data Architect (\$108,278/year)
- Data Engineer (\$102,864 /year)
- Business Intelligence (BI) Developer(\$81,514 /year)
- Statistician (\$76,884/year)
- Database Administrator (\$72,400/year)
- Data Analyst(\$62, 453/year)

aiQuest's all courses are available: at <https://www.aiquest.org>



[LinkedIn](#)



[GitHub](#)

kaggle

[Kaggle](#)

[Download the Books](#)

Thank you!

