9
CBSE

ARTIFICIAL INTELLIGENCE

CODE 417 | Skill Education

Supplement

<This book belongs to>

<Name> .....</Name>

<Class> .....</Class> <Section> .....</Section>

<Roll No.> .....</Roll No.> <Age> .....</Age>

<School> .....

.....</School>

Shalini Harisukh
Anjna Virmani

Ver. 2.0

ORANGE

Published by:

Orange Education

9, Daryaganj, New Delhi-110002

Phone: 43776600 8588814859
Email: info@orangeeducation.in
Website: www.orangeeducation.in

IE Code: 0511063121

Branches:
• Chennai • Guwahati

Regional Offices:
• Ahmedabad • Bengaluru • Bhopal
• Bhubaneswar • Dehradun • Hyderabad
• Jaipur • Kochi • Kolkata • Lucknow
• Mumbai • Patna • Raipur • Ranchi

Printed at:

Printed and Distributed by Orange House Pvt Ltd

ARTIFICIAL INTELLIGENCE (SUBJECT CODE 417)
CLASS – IX

Total Marks: 100 (Theory-50 + Practical-50)

OBJECTIVES OF THE COURSE:

The objective of this module/curriculum - which combines both Inspire and Acquire modules is to develop a readiness for understanding and appreciating Artificial Intelligence and its application in our lives. This module/curriculum focuses on:

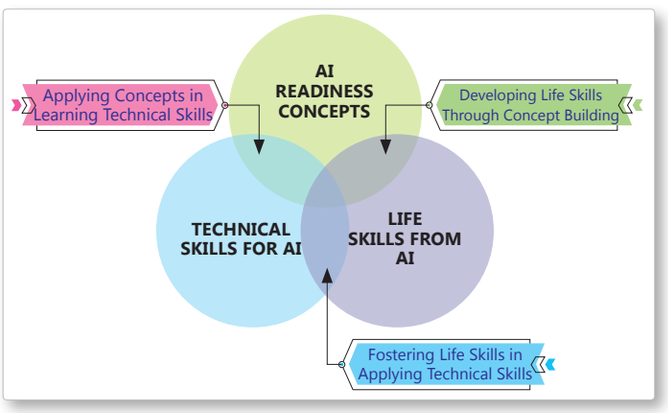
1. Helping learners understand the world of Artificial Intelligence and its applications through games, activities and multi-sensorial learning to become AI-Ready.
2. Introducing the learners to three domains of AI in an age-appropriate manner.
3. Allowing the learners to construct the meaning of AI through interactive participation and engaging hands-on activities.
4. Revisiting AI domains, project cycle and Ethics
5. Introducing the learners to the importance of Math for AI, data literacy and generative AI
6. Introducing the learners to programming skills - Basic python coding language.

LEARNING OUTCOMES:

Learners will be able to:

1. Identify and appreciate Artificial Intelligence and describe its applications in daily life.
2. Relate, apply and reflect on the Human-Machine Interactions to identify and interact with the three domains of AI: Data, Computer Vision and Natural Language Processing and Undergo assessment for analysing their progress towards acquired AI-Readiness skills.
3. Imagine, examine and reflect on the skills required for futuristic job opportunities.
4. Unleash their imagination towards smart homes and build an interactive story around it.
5. Understand the impact of Artificial Intelligence on Sustainable Development Goals to develop responsible citizenship.
6. Research and develop awareness of skills required for jobs of the future.
7. Gain awareness about AI bias and AI access and describe the potential ethical considerations of AI.
8. Develop effective communication and collaborative work skills.
9. Get familiar and motivated towards Artificial Intelligence and Identify the AI Project Cycle framework.
10. Learn problem scoping and ways to set goals for an AI project and understand the iterative nature of problem scoping in the AI project cycle.
11. Brainstorm on the ethical issues involved around the problem selected.
12. Foresee the kind of data required and the kind of analysis to be done, identify data requirements and find reliable sources to obtain relevant data.
13. Use various types of graphs to visualize acquired data.
14. Understand types of modelling.
15. Understand the importance of Math for AI.
16. Learn the concept of data literacy and generative AI
17. Acquire introductory Python programming skills in a very user-friendly format.

SKILLS TO BE DEVELOPED



SCHEME OF STUDIES:

This course is a planned sequence of instructions consisting of units meant for developing employability and vocational competencies of students of Class IX opting for skill subject along with other education subjects.

The unit-wise distribution of hours and marks for class IX is as follows.

	UNITS	NO. OF HOURS for Theory and Practical		MAX. MARKS for Theory and Practical
PART A	**Employability Skills**			
	Unit 1: Communication Skills-I	10		2
	Unit 2: Self-Management Skills-I	10		2
	Unit 3: ICT Skills-I	10		2
	Unit 4: Entrepreneurial Skills-I	15		2
	Unit 5: Green Skills-I	05		2
	Total	**50**		**10**
PART B	**Subject Specific Skills**			
		Theory	**Practical**	
	Unit 1: AI Reflection, Project Cycle and Ethics	30	25	10
	Unit 2: Data Literacy	22	28	10
	Unit 3: Math for AI (Statistics & Probability)	12	13	07
	Unit 4: Introduction to Generative AI	08	12	05
	Unit 5: Introduction to Python	01	09	08
	Total	**160**		**40**
PART C	**Practical Work**			
	Unit 5: Introduction to Python **Practical File (minimum 15 programs)**			15
	Practical Examination • Simple programs using input and output function • Variables, Arithmetic Operators, Expressions, Data Types • Flow of control and conditions • Lists * ***Any 3 programs based on the above topics***			15
	Viva Voce			5
	Total			**35**
PART D	**Project Work / Field Visit / Student Portfolio** * ***relate it to Sustainable Development Goals***			15
	Total			**15**
	GRAND TOTAL	**210**		**100**

EMPLOYABILITY SKILLS

UNIT	LEARNING OUTCOMES	THEORY	PRACTICAL
COMMUNICATION SKILLS –I	1. Demonstrate knowledge of various methods of communication	1. Methods of communication • Verbal • Non-verbal • Visual	1. Writing pros and cons of written, verbal and non-verbal communication 2. Listing do's and don'ts for avoiding common body language mistakes
	2. Identify elements of communication cycle	1. Meaning of communication 2. Importance of communication skills 3. Elements of communication cycle– (i) sender, (ii) ideas, (iii) encoding, (iv) communication channel, (v) receiver, (vi) decoding, and (vii) feedback	1. Draw a diagram of communication cycle 2. Role plays on communication process related to the sector/job role
	3. Identify the factors affecting our perspectives in communication	1. Perspectives in communication 2. Factors affecting perspectives in communication • Visual perception • Language • Past experience • Prejudices • Feelings • Environment	1. Group discussion on factors affecting perspectives in communication 2. Sharing of experiences on factors affecting perspectives 3. Sharing experiences on factors affecting communication at workplace
	4. Demonstrate the knowledge of basic writing skills	1. Writing skills related to the following: • Phrases • Kinds of sentences • Parts of sentence • Parts of speech • Use of articles • Construction of a paragraph	1. Demonstration and practice of writing sentences and paragraphs on topics related to the subject

UNIT	LEARNING OUTCOMES	THEORY	PRACTICAL
SELF-MANAGEMENT SKILLS – I	1. Describe the meaning and importance of self-management	1. Meaning of self-management 2. Positive results of self-management 3. Self-management skills	1. Identification of self-management skills 2. Strength and weakness analysis
	2. Identify the factors that helps in building self-confidence	1. Factors that help in building self-confidence– social, cultural, and physical factors 2. Self-confidence building tips– getting rid of the negative thoughts, thinking positively, staying happy with small things, staying clean, hygienic and smart, chatting with positive people, etc.	1. Role play exercises on building self-confidence 2. Use of positive metaphors/ words 3. Positive stroking on wakeup and before going bed 4. Helping others and working for community

UNIT	LEARNING OUTCOMES	THEORY	PRACTICAL
ICT SKILLS - I	1. Describe the role of Information and Communication Technology (ICT) in day-to-day life and workplace	1. Introduction to ICT 2. Role and importance of ICT in personal life and at workplace 3. ICT in our daily life (examples) 4. ICT tools– Mobile, tab, radio, TV, email, etc.	1. Discussion on the role and importance of ICT in personal life and at workplace 2. Preparing posters/collages for showing the role of ICT at workplace
	2. Identify components of basic computer system and their functions	1. Computer system– Central Processing Unit (CPU), memory, motherboard, storage devices 2. Hardware and software of a computer system 3. Role and functions of Random Access Memory (RAM) and Read Only Memory (ROM) 4. Role and functions of Central Processing Unit 5. Procedure for starting and shutting down a computer	1. Connecting the cables and peripherals to the Central Processing Unit 2. Starting and shutting down a computer 3. Group discussion on the various aspects of hardware and software
	3. Demonstrate use of various components and peripherals of computer system	1. Peripherals devices and their uses– mouse, keyboard, scanner, webcam, etc. of a computer system	1. Identification of various parts and peripherals of a computer 2. Demonstration and practice on the use of mouse 3. Demonstration and practice on the use of keyboard 4. Demonstration of the uses of printers, webcams, scanner and other peripheral devices 5. Drawing diagram of computer system and labelling it
	4. Demonstrate basic computer skills	1. Primary operations on a computer system– input, process, storage, output, communication networking, etc.	1. Identification of the various input and output units and explanation of their purposes

UNIT	LEARNING OUTCOMES	THEORY	PRACTICAL
ENTREPRENEURIAL SKILLS – I	1. Identify various types of business activities	1. Types of businesses– service, manufacturing, hybrid 2. Types of businesses found in our community (Business activities around us)	1. Prepare posters of business activities found in cities/villages, using pictures 2. Discuss the various types of activities, generally adopted by small businesses in a local community 3. Best out of waste 4. Costing of the product made out of waste 5. Selling of items made from waste materials 6. Prepare list of businesses that provides goods and services in exchange for money
	2. Demonstrate the knowledge of distinguishing characteristics of entrepreneurship	1. Meaning of entrepreneurship development 2. Distinguishing characteristics of entrepreneurship 3. Role and rewards of entrepreneurship	1. Prepare charts showing advantages of entrepreneurship over wages 2. Group discussions on role and features of entrepreneurship 3. Lectures/presentations by entrepreneurs on their experiences and success stories 4. Identify core skills of successful entrepreneur

UNIT	LEARNING OUTCOMES	THEORY	PRACTICAL
GREEN SKILLS – I	1. Demonstrate the knowledge of the factors influencing natural resource conservation	1. Introduction to environment 2. Relationship between society and environment, ecosystem and factors causing imbalance 3. Natural resource conservation 4. Environment protection and conservation	1. Group discussion on hazards of deteriorating environment 2. Prepare posters showing environment conservation 3. Discussion on various factors that influence our environment
	2. Describe the importance of green economy and green skills.	1. Definition of green economy 2. Importance of green economy	1. Discussion on the benefits of green skills and importance of green economy 2. Prepare a poster showing the importance of green economy with the help of newspaper/ magazine cuttings

SUBJECT SPECIFIC SKILLS

UNIT 1: AI REFLECTION, PROJECT CYCLE AND ETHICS

SUB-UNIT	LEARNING OUTCOMES	SESSION / ACTIVITY / PRACTICAL
AI REFLECTION	To identify and appreciate Artificial Intelligence and describe its applications in daily life.	Session: Introduction to AI and setting up the context of the curriculum ● Recommended Activity: Make a statement about lighting and LUIS will interpret and adjust the house accordingly https://aidemos.microsoft.com/luis/demo
	To recognize, engage and relate with the three realms of AI: Computer Vision, Data Statistics and Natural Language Processing.	Recommended Activity: The AI Game ● Learners to participate in three games based on different AI domains. – Game 1: Rock, Paper and Scissors (based on data) – https://next.rockpaperscissors.ai/ – Game 2: Semantris (based on Natural Language Processing - NLP) – https://research.google.com/semantris/ – Game 3: Quick Draw (based on Computer Vision - CV) – https://quickdraw.withgoogle.com/
AI PROJECT CYCLE	Identify the AI Project Cycle framework	Session: Introduction to AI Project Cycle ● Problem Scoping ● Data Acquisition ● Data Exploration ● Modelling ● Evaluation ● Deployment

	Learn problem scoping and ways to set goals for an AI project.	Session: Problem Scoping Activity: Brainstorm around the theme provided and set a goal for the AI project. ● Discuss various topics within the given theme and select one. ● Fill in the 4Ws problem canvas and a problem statement to learn more about the problem identified in the community/ society ● List down/ Draw a mind map of problems related to the selected topic and choose one problem to be the goal for the project.
	Identify stakeholders involved in the problem scoped. Brainstorm on the ethical issues involved around the problem selected.	● Activity: To set actions around the goal. ● List down the stakeholders involved in the problem. ● Search on the current actions taken to solve this problem. ● Think around the ethics involved in the goal of your project.
	Understand the iterative nature of problem scoping for in the AI project cycle. Foresee the kind of data required and the kind of analysis to be done.	Activity: Data and Analysis ● What are the data features needed? ● How will the features collected affect the problem? ● Where can you get the data? ● How frequent do you have to collect the data? ● What happens if you don't have enough data? ● What kind of analysis needs to be done? ● How will it be validated? ● How does the analysis inform the action?
	Share what the students have discussed so far.	Presentation: Presenting the goal, actions and data. Teamwork Activity: ● Brainstorming solutions for the problem statement.
	Identify data requirements and find reliable sources to obtain relevant data.	Session: Data Acquisition Activity: Introduction to data and its types. ● Students work around the scenarios given to them and think of ways to acquire data. Activity: Data Features ● Identifying the possible data features affecting the problem. Activity: System Maps ● Creating system maps considering data features identified.
	To understand the purpose of Data Visualisation	Session: Data Exploration/ Data Visualisation ● Need of visualising data ● Ways to visualise data using various types of graphical tools. Quiz Time
	Use various types of graphs to visualise acquired data.	Recommended Activities: Let's use Graphical Tools ● Selecting an appropriate graphical format and presenting the graph sketched. ● Understanding graphs using https://datavizcatalogue.com/ ● Listing of newly learnt data visualization techniques. ● Top 10 Song Prediction: Identify the data features, collect the data and convert into graphical representation. ● Collect and store data in a spreadsheet and create some graphical representations to understand the data effectively.
	Understand modelling (Rule-based & Learning-based)	Session: Modelling ● Introduction to modelling and types of models (Rule-based & Learning-based)

	Understand various evaluation techniques.	Session: Evaluation Learners will understand about new terms ● True Positive ● False Positive ● True Negative ● False Negative
	Challenge students to think about how they can apply their knowledge of deployment in future AI projects and encourage them to continue exploring different deployment methods.	Session: Deployment Recommended Case Study: Preventable Blindness. Activity: Implementation of AI project cycle to develop an AI Model for Personalized Education.
	To understand and reflect on the ethical issues around AI.	Session: Ethics Video Session: Discussing about AI Ethics Recommended Activity: Ethics Awareness ● Students play the role of major stakeholders, and they have to decide what is ethical and what is not for a given scenario. ● Students to explore Moral Machine (https://www.moralmachine.net/) to understand more about the impact of ethical concerns
	To gain awareness around AI bias and AI access.	Session: AI Bias and AI Access ● Discussing about the possible bias in data collection ● Discussing about the implications of AI technology
	To let the students analyse the advantages and disadvantages of Artificial Intelligence.	Recommended Activity: Balloon Debate ● Students divide in teams of 3 and 2 teams are given same theme. One team goes in affirmation to AI for their section while the other one goes against it. ● They have to come up with their points as to why AI is beneficial/ harmful for the society

UNIT 2: DATA LITERACY

SUB-UNIT	LEARNING OUTCOMES	SESSION / ACTIVITY / PRACTICAL
Basics of Data Literacy	● Define data literacy and recognize its importance Understand how data literacy enables informed decision-making and critical thinking ● Apply the Data Literacy Process Framework to analyze and interpret data effectively ● Differentiate between Data Privacy and Security ● Identify potential risks associated with data breaches and unauthorized access. ● Learn measures to protect data privacy and enhance data security	Session: Basics of data literacy ● Introduction to Data Literacy ● Impact of data Literacy ● How to become Data Literate? ● What are data security and privacy? How are they related to AI? ● Best Practices for Cyber Security
		Recommended Activity: Impact of News Articles Reference Videos: ● https://www.youtube.com/watch?v=yhO_t-c3yJY ● https://www.youtube.com/watch?v=aO858HyFbKI ● https://www.cbse.gov.in/cbsenew/documents/Cyber%20Safety.pdf

Acquiring Data, Processing, and Interpreting Data	• Determine the best methods to acquire data. • Classify different types of data and enlist different methodologies to acquire it. • Define and describe data interpretation. • Enlist and explain the different methods of data interpretation. • Recognize the types of data interpretation. • Realize the importance of data interpretation	Session: Acquiring Data, Processing, and Interpreting Data • Types of data • Data Acquisition/Acquiring Data • Best Practices for Acquiring Data • Features of data and Data Preprocessing • Data Processing and Data Interpretation • Types of Data Interpretation • Importance of Data Interpretation Recommended Activities: • Trend analysis • Visualize and Interpret Data
Project Interactive Data Dashboard & Presentation	• Recognize the importance of data visualization • Discover different methods of data visualization	Session: Project Interactive Data Dashboard & Presentation • Data visualization Using Tableau Reference Links • https://public.tableau.com/en-us/s/download • https://www.datawrapper.de/ Video Links: • https://www.youtube.com/watch?v=NLCzpPRCc7U • https://www.youtube.com/watch?v=_M8BnosAD78

UNIT 3: MATH FOR AI (STATISTICS & PROBABILITY)

SUB-UNIT	LEARNING OUTCOMES	SESSION / ACTIVITY / PRACTICAL

Importance of Math for AI	Analyzing the data in the form of numbers/ images and find the relation/pattern between the them.
	Use of Math in AI.
		Session: Importance of Math for AI
		• Finding Patterns in Numbers and images.
		• Uses of Math -
		  ○ Statistics
		  ○ Linear Algebra
		  ○ Probability
		  ○ Calculus

	Number Patterns Picture Analogy	Activity: • observe the number pattern and find the missing number. • To find connections between sets of images and use that to solve problems,s

Statistics	Understand the concept of Statistics in real life.	Session :
		• Definition of Statistics
		• Applications
		  ○ Disaster Management
		  ○ Sports
		  ○ Diseases Prediction
		  ○ Weather Forecast

	Application in various real life scenarios	Activity: Uses of Statistics in daily life • Students will explore the applications of statistics in real life .They collect data and can apply various statistical measures to analyze the data. Activity: Car Spotting and Tabulating Purpose: To implement the concept of data collection, analysis and interpretation. Activity Introduction: • In this activity, Students will be engaged in data collection and tabulation. • Data collection plays a key role in Artificial Intelligence as it forms the basis of statistics and interpretation by AI. • This activity will also require students to answer a set of questions based on the recorded data.

Probability	Understand the concept of Probability in real life and explore various types of events.	Session: Introduction to Probability ● How to calculate the probability of an event ● Types of events ● understand the concept of Probability using a relatable example. Exercise: Identify the type of event.
	Application in various real life scenarios	Session : Applications of Probability ● Sports ● Weather Forecast ● Traffic Estimation Exercise: Revision time

UNIT 4: INTRODUCTION TO GENERATIVE AI:

LEARNING OUTCOMES	SESSION / ACTIVITY / PRACTICAL
Students will be able to define Generative AI & classify different kinds.	Recommended Activity: ● Activity: Guess the Real Image vs. the AI-generated image
● Students will be able to explain how Generative AI works and recognize how it learns.	Session: ● Introduction to Generative AI ● Generative AI vs Conventional AI:
	Session: ● Types of Generative AI ● Examples of Generative AI
	Session: ● Benefits of using Generative AI ● Limitations of using Generative AI
● Applying Generative AI tools to create content.	Recommended Activities: ● Hands-on Activity: GAN Paint ● Generative AI tools
● Applying Generative AI tools to create content.	Session: ● Ethical considerations of using Generative AI

UNIT 5: INTRODUCTION TO PYTHON:

LEARNING OUTCOMES	SESSION / ACTIVITY / PRACTICAL
Learn basic programming skills through gamified platforms.	Recommended Activity: ● Introduction to programming using Online Gaming portals like Code Combat.
Acquire introductory Python programming skills in a very user-friendly format.	Session: ● Introduction to Python language ● Introducing python programming and its applications
	Theory + Practical: Python Basics ● Students go through lessons on Python Basics (Variables, Arithmetic Operators, Expressions, Comparison Operators, logical operators, Assignment Operators, Data Types - integer, float, strings, type conversion, using print() and input() functions) ● Students will try some simple problem-solving exercises on Python Compiler.
	Practical: Flow of control and conditions 1. Students go through lessons on conditional and iterative statements (if, for and while) 2. Students will try some basic problem-solving exercises using conditional and iterative statements on Python Compiler.
	Practical: Python Lists 3. Students go through lessons on Python Lists (Simple operations using list) 4. Students will try some basic problem-solving exercises using lists on Python Compiler.

Content from Existing Book

New	Unit / Sub Unit	Old	Page No
Unit - 1	AI Reflection, Project Cycle and Ethics		
	AI Reflection	**Unit - 1**	143-208
	AI Project Cycle	**Unit - 2**	209-254
Unit - 5	Introduction To Python	**Unit - 4**	278-346

CONTENTS

Part B Subject Specific Skills

PART-B

SUBJECT SPECIFIC SKILLS

UNIT-1

AI REFLECTION, PROJECT CYCLE AND ETHICS

Learning Outcomes

- Project Cycle
- Model Evaluation Terminologies
- What is Deployment?
- Difference between Ethics and Morals
- What is Evaluation?
- Confusion Matrix
- AI Project Cycle Mapping Template

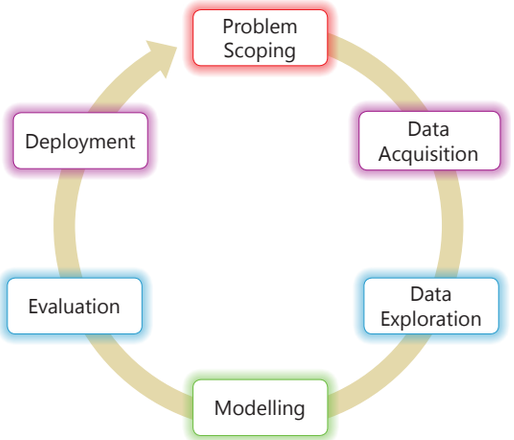
Artificial Intelligence (AI) is the simulation of human intelligence by machines, particularly computer systems. It involves algorithms that enable computers to perform tasks such as learning, reasoning, problem-solving, and understanding language. AI systems can improve their performance over time through machine learning. AI is used in various applications, from virtual assistants to autonomous vehicles.

Project Cycle

AI project cycle is the process of solving the real-life problems by converting them into an AI-based model. The project cycle framework is designed to help project managers guide their projects successfully from start to end. The purpose of the project life cycle is to create an easy-to-follow framework to guide projects. The AI project cycle provides us with an appropriate framework which can lead us towards our goal.

Stages in an AI Project Cycle

The AI project cycle involves several key stages, each building upon the previous one to develop, deploy, and maintain an AI system effectively. These stages are as follows:



- **Problem Scoping:** The first stage of an AI project cycle is problem scoping to identify the problem and have a vision to solve it. Problem scoping refers to understanding a problem and various factors which affect the problem, and finding a solution for it using AI technology. The 4W's of problem scoping are Who, What, Where, and Why. This Ws helps in identifying and understanding the problem in a better and efficient manner.
- **Data Acquisition:** The next stage of the AI project cycle is data acquisition. The term data acquisition means collecting raw data for the purpose of reference or analysis for the project. The data can be in the form of text, numbers, images, videos, or audio and it can be collected from various sources like Internet, journals, newspapers, and so on. The data acquisition system allows us to obtain valuable information about reality to improve the performance of the project.
- **Data Exploration:** Data exploration refers to the techniques and tools used to visualise data collected in data acquisition through complex statistical methods. It is the process of analysing a large dataset.
- **Modelling:** It is the design phase of the AI project cycle. In this, we select the best way to reach the solution. It requires the process of selecting the right algorithm to develop a working model for the project. In this step, different models based on the visualised data can be created and even checked for the advantages and disadvantages of the model.
- **Evaluation:** It is the testing phase of the AI project cycle, where we check if the model can achieve required goals or not. If the model is not fulfilling the requirements, the model or even the data can be changed. Once the developer feels the project is ready, the project will be put into working conditions and then deployed and handed over to the user. If the deployment stage is not reached, the project is of no use.
- **Deployment:** In this stage, we integrate the best-performing model into the production environment, setting up continuous monitoring, and maintenance to sustain performance over time.

Ai Reboot

1. List down the steps of AI project cycle.

   _____

2. Differentiate between evaluation and deployment in an AI project cycle.

   _____

Why do we need an AI Project Cycle?

An AI project cycle is essential because it provides a structured framework for developing, deploying, and maintaining AI systems. Each stage in the AI project life cycle serves a vital role. **Problem scoping** ensures that the problem is clearly defined and aligned with business objectives. The **data acquisition** and **exploration** phases help in gathering and cleaning relevant data, which is crucial for building effective models. During the **modelling** phase, appropriate algorithms are selected and optimised for the best performance. The **evaluation** phase thoroughly tests these models to ensure they meet required standards. **Deployment** integrates the model into the production environment, with monitoring to maintain its performance. The iterative nature of the cycle allows for continuous improvement, addressing new data and evolving business needs. This structured approach helps manage risks, ensures quality, and facilitates communication with stakeholders, ultimately leading to successful AI implementation.

What is Evaluation?

Model evaluation is the stage of testing the model where testing data is given to the system and the output generated is evaluated with the actual result to see the accuracy of the output and the reliability of the AI model. There can be different evaluation techniques, depending on the type and purpose of the model.

AI Reboot

1. Explain the term problem scoping in AI project Cycle.

   _____

2. Differentiate between data exploration and data acquisition.

   _____

Importance of Evaluation

Evaluation is a process that critically examines a program. It involves collecting and analysing information about a program's activities, characteristics, and outcomes. Its purpose is to make judgments about a program, to improve its effectiveness, and/or to inform programming decisions. Following are some of the advantages of evaluating a AI model:

- Evaluation ensures that the model is operating correctly and optimally.
- Evaluation is an initiative to understand how well it achieves its goals.
- Evaluations help to determine what works well and what could be improved in a program.

Model Evaluation Terminologies

Evaluation of an AI model can be done using various terminologies. Let us try to understand them with the help of a scenario.

Scenario:

An AI-based prediction model is deployed in schools. The model is supposed to predict whether the students of grade 12 will be taking board exams in the coming year or not. The model will be checking for whether there will be board exams in the coming year or not.

There are two important parameters which are used for the **evaluation** of a model. These are:

- **Prediction**: It is the output given by the AI model using a machine learning algorithm.
- **Reality**: It is the real scenario of the situation for which the prediction has been made.

Let's look at the various combinations that can be considered for the above scenario.

Case 1: Is there a Board Exam?

Predicion: Yes

True Positive

Reality: Yes

Due to COVID-19 things are becoming unpredictable. Even the conducting of the board exams totally depends on the number of active cases. We need an AI model which can predict whether the board exams will be conducted or not so that the students can timely plan their preparation and schedule to study as per the date sheet.

In the above picture, we show the possibility of board exams for grade 12 students. The model predicts a **Yes** which means the board exams will be conducted. The prediction matches with the reality: Yes, therefore, this condition is called **True Positive**.

Case 2: Is there a Board Exam?

Predicion: No

True Negative

Reality: No

There are no board exams as the numbers of COVID-19 cases have increased, hence the **Reality** is **No**. In this case, the machine too has **predicted** it correctly as a **No**. Therefore, this condition is termed as **True Negative**.

Case 3: Is there a Board Exam?

Predicion: Yes

False Positive

Reality: No

Here, the reality is that there are no board exams to be conducted as they got cancelled due to the COVID-19 number of cases increasing drastically. However, the machine has incorrectly predicted that there will be board exams for the students of grade 12. This case is termed as **False Positive**.

Case 4: Is there a Board Exam?

Predicion: No

False Negative

Reality: Yes

Here, the board has decided to conduct examinations for grade 12 students because of which the **Reality is Yes** but the machine has incorrectly **predicted** it as a **No** which means the machine predicts that there will be no board exams. Therefore, this case becomes **False Negative.**

Confusion Matrix

Confusion matrix is a tabular structure which helps in measuring the performance of an AI model using the test data. The table is made with 4 different combinations of predicted and actual values (Reality) in the form of a 2×2 matrix. The comparison between the prediction and the reality can be used to analyse the rate of success. It also gives a clear picture of which classes are being predicted correctly and incorrectly, and what type of errors are being made.

This matrix is also known as the **Error Matrix** and is used in situations where we need to evaluate the performance of the model, where it went wrong and help us in finding ways to increase the efficiency of the model. It is useful for measuring Recall, Precision, Accuracy and F1 Score.

The following confusion matrix table illustrates how the 4-classification metrics are calculated (TP, FP, FN, TN), and how our predicted value is compared to the actual value in a confusion matrix.

Confusion Matrix		**Reality**	
		Yes	No
Prediction	Yes	True Positive (TP)	False Positive (FP)
	No	False Negative (FN)	True Negative (TN)

In the confusion matrix:

- The target variable has two values: **Positive** and **Negative.**
- The **columns** represent the **actual values** of the target variable.
- The **rows** represent the **predicted values** of the target variable.

For example:

In schools, a lot of times it happens that there is no transport facility available to commute. The unavailability of school transport is a very common and prominent problem. Hence, an AI model is designed to predict if there is going to be school transport available or not.

The confusion matrix for the same is:

The Confusion Matrix	**Reality:1 (TP)**	**Reality: 0 (FP)**
Prediction:1	22	12
Prediction:0	48	18

Ai Reboot

1. Differentiate between prediction and reality.

   _____

2. What is error matrix?

   _____

ROC Curve

The Receiver Operating Characteristic (ROC) curve is a graphical representation that illustrates the performance of a binary classifier system at varying threshold values. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings.

This curve plots two parameters:

- **True Positive Rate (Sensitivity)** is the proportion of actual positive cases that are correctly identified by the classifier.

$$TPR = \frac{TP}{TP + FN}$$

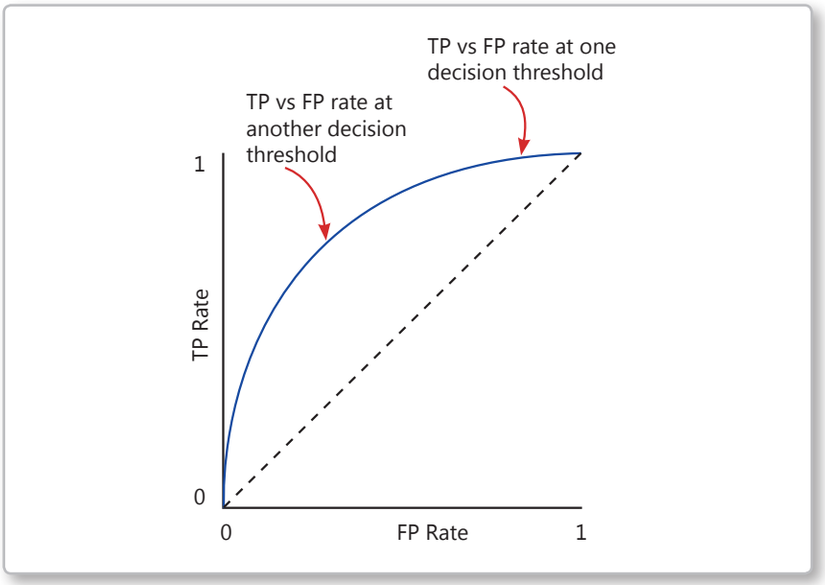
- **False Positive Rate** is the proportion of actual negative cases that are incorrectly classified as positive.

$$FPR = \frac{FP}{FP + TN}$$

To generate an ROC curve, you need to perform the following tasks:

- Vary the threshold of your classifier, usually ranging from 0 to 1, and calculate TPR and FPR at each threshold.
- Plot these TPR and FPR values on a graph. TPR is plotted on the y-axis, and FPR is plotted on the x-axis.

Lowering the classification threshold classifies more items as positive, thus increasing both false positives and true positives. The following figure shows a typical ROC curve.

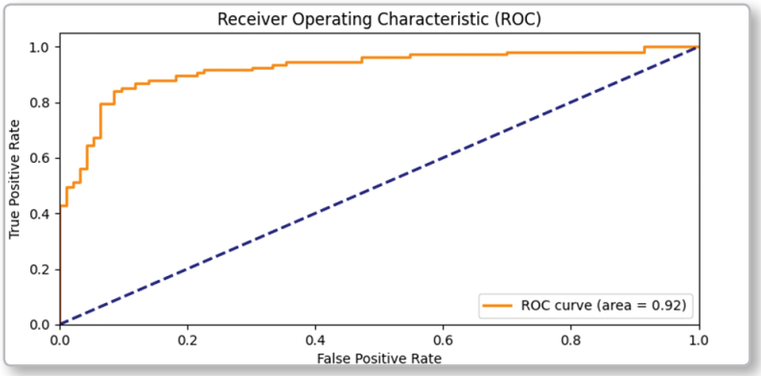


Evaluate: Exoplanet Use Case

At this particular stage, we may need to evaluate the model to find out the accuracy of the model that makes the best prediction.

ROC is a metric that is used to find out the accuracy of the model.

Figure shows the accuracy of the model using ROC Curve:



What is Deployment?

The deployment phase is the last stage of the AI project cycle and is used when the AI model is put into use in a real-world setting. This involves integrating the model into existing systems or applications, such as creating Application Programming Interfaces (APIs) or embedding it directly into software. It also includes setting up the necessary infrastructure, like servers or cloud services, to support the model. Once integrated, the model needs to be able to process new data and provide predictions. Monitoring tools are established to track the model's performance and ensure it works correctly. Logging and reporting are also important to capture data on how the model is performing and to identify any issues that might arise. This phase is crucial for making the AI model functional and useful for end-users.

AI Project Cycle Mapping Template

AI Project Cycle Mapping Template presents how different stages are related to each other and how the functions performed in every phase forms an input for the next phase.

The performed task at one stage forms the root for the next stage.

AI Project: Customer churn prediction (identifying at-risk customers who are likely to cancel their subscriptions or close/abandon their accounts.)

- **Problem Scoping**
  - ✶ **Identify the problem:** The telecommunications company wants to reduce customer churn rates.
  - ✶ **Define objectives:** Develop a predictive model to identify customers at risk of churning.
- **Data Acquisition**
  - ✶ **Gather data sources:** Collect customer demographics, usage patterns, service history, and churn status data from the company's databases.
  - ✶ **Ensure data quality:** Clean the data, handle missing values, and remove duplicates.
- **Data Exploration**
  - ✶ **Explore the data:** Analyse customer demographics, usage patterns, and churn rates through visualisations and statistical summaries.
  - ✶ **Preprocess data:** Simplify numerical features, convert categorical variables, and create new metrics like customer tenure.
- **Modelling**
  - ✶ **Select techniques:** Choose machine learning algorithms suitable for classification tasks, such as logistic regression, decision trees, and random forests.
  - ✶ **Train models:** Use the prepared data to train multiple models, adjusting hyperparameters and performing cross-validation to optimise performance.
- **Evaluation**
  - ✶ **Evaluate models:** Assess the performance of each model using metrics like accuracy, precision, recall, and F1-score.
  - ✶ **Compare models:** Compare the performance of different models to select the best-performing one for deployment.

- **Deployment**
  - ✶ **Deploy model:** Integrate the selected model into the company's customer management system to predict churn risk for new customers.
  - ✶ **Monitor performance:** Monitor the model's predictions in real-time, track churn rates, and gather feedback from customer service interactions.

In this example, each phase of the AI project cycle builds upon the outputs of the previous phase:

AI Project Cycle Mapping Template					
Problem Solving	**Data Acquistion**	**Data Exploration**	**Modelling**	**Evaluation**	**Deployment**
The telecommunications company wants to reduce customer churn rates.	Gather customer demographics, usage patterns, service history, and churn status data from company databases.	Analyse customer demographics, usage patterns, and churn rates with visualisations and statistical summaries.	Select machine learning algorithms for classification, like logistic regression, decision trees, and random forests.	Evaluate each model's performance using accuracy, precision, recall, and F1-score.	Integrate the model to predict new customer churn risk.

- **Problem Scoping:** States the problem that needs attention.
- **Data Acquisition:** Data acquisition consists of two words: Data and Acquisition. Data refers to the raw facts, figures, information, or statistics; where as, acquisition refers to acquiring data for the project.
- **Data Exploration:** It is the first step of data analysis which is used to visualise data. It generates insights used to inform modelling decisions.
- **Modelling:** Develops predictive models based on insights gained from data exploration.
- **Evaluation:** Assesses model performance by the feeding the data into the model and comparing the output with the actual answers. It is used to determine the best model for deployment.
- **Deployment:** Integrates the selected model into the company's systems for real-world usage.

The feedback loop continues as the deployed model's performance is monitored, and insights gathered are used to refine future iterations of the AI solution.

Difference between Ethics and Morals

Ethics and morals are related concepts often used interchangeably, but they have distinct meanings and applications. The word Ethics originated from the Greek word ethos. The meaning of ethos is a character. The word Morals originated from the Latin word Mos. The meaning of Mos is custom.

Aspect	Ethics	Morals
Definition	Rules provided by an external source	Principles regarding right and wrong held by an individual
Source	Institutions, organisations, societal norms	Personal beliefs, cultural norms, religious teachings
Application	Specific situations and professional practices	Personal behaviour and conduct

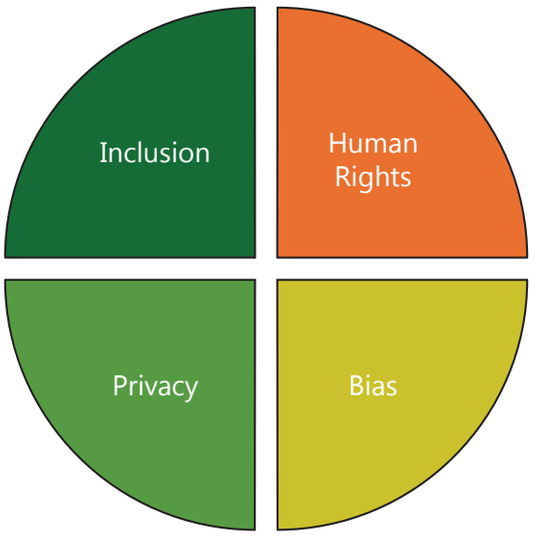
Aspect	Ethics	Morals
Objective	Maintain order and fairness in society	Foster personal integrity and align with personal values
Examples	Medical ethics, business ethics, legal ethics	Personal beliefs about honesty, integrity, kindness
Origin	External and often codified	Internal and subjective
Scope	Consistent within a profession or society	Varies between individuals
Enforcement	Enforced by external bodies (e.g., professional organisations, legal systems)	Self-governed and enforced by individual conscience
Flexibility	Can change over time to reflect new norms or societal changes	More stable over time, but can evolve with personal growth

Ethics and Personal Data

Ethics play a crucial role in handling personal data, focusing on privacy, consent, transparency, and data security. Privacy ensures that individuals' personal information is respected and protected, requiring organisations to collect, use, share, and process data in ways that maintain confidentiality. Consent involves obtaining clear and explicit permission from individuals before collecting, sharing, processing, or using their data, ensuring they are informed about how their data will be used and giving them the option to withdraw consent. Transparency means being open about data collection practices, clearly communicating what data is collected, how it is used, stored and analysed, and who it is shared with. Data security involves implementing strong measures to protect personal data from unauthorised access, breaches, and other threats, ensuring the integrity and safety of the information. These ethical principles help build trust and ensure responsible data management.

What are the Principles of AI Ethics?

Ethics in AI encompasses the moral principles, values, and guidelines that govern the development, deployment, and use of artificial intelligence systems.



- **Human rights:** This principle emphasises that AI solutions should respect, protect, and uphold fundamental human rights. This includes rights such as privacy, freedom of expression, freedom from discrimination, and the right to a fair trial. AI systems should be designed and implemented in a way that they does not infringe upon these rights and should be held accountable if they do.
- **Bias:** Bias in AI refers to the unfair or unjust treatment of individuals or groups based on characteristics such as race, gender, age, or socioeconomic status. Bias can be unintentionally introduced into AI systems through biased training data, flawed algorithms, or skewed decision-making processes. Addressing bias in AI involves identifying, mitigating, and preventing bias at every stage of the AI development lifecycle, from data collection and preprocessing to model training and deployment.
- **Privacy:** Privacy concerns the protection of individuals' personal data and their right to control how that data is collected, used, and shared. AI systems often rely on vast amounts of data, which may include sensitive information about individuals. It is essential to implement robust privacy measures, such as data anonymisation, encryption, and user consent mechanisms, to ensure that AI solutions respect individuals' privacy rights and comply with relevant data protection regulations.
- **Inclusion:** Inclusion in AI refers to ensuring that AI solutions are accessible, equitable, and beneficial for all members of society, regardless of factors, such as race, gender, disability, or socioeconomic status. This involves considering the diverse needs, perspectives, and experiences of different user groups throughout the design, development, and deployment of AI systems. Inclusive AI design aims to prevent the exacerbation of existing inequalities and to promote equal opportunities and outcomes for all individuals.

By adhering to these AI ethics principles, developers and organisations can contribute to the creation of AI solutions that are not only technically robust but also ethically sound, socially responsible, and aligned with the values and interests of society.

At a Glance

- AI project cycle is the process of converting a real-life problem into an AI-based model.
- An AI project cycle is essential because it provides a structured framework for developing, deploying, and maintaining AI systems.
- Data Exploration refers to the techniques and tools used to visualise data through complex statistical methods.
- Evaluation stage is the testing stage of the AI project cycle, where we check if the model can achieve the required goals or not.
- Confusion matrix is a tabular structure which helps in measuring the performance of an AI model using the test data.
- The Receiver Operating Characteristic (ROC) curve is a graphical representation that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.
- The Deployment phase of the AI project cycle is when the AI model is put into use in a real-world setting.
- AI Project Cycle Mapping Template presents how different stages are related to each other and how the functions performed in every phase forms an input for the next phase.
- Ethics and morals are related concepts often used interchangeably, but they have distinct meanings and applications.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. What is the primary objective of the cycle of AI projects?
   a. To create complex AI models quickly. ◯
   b. To simplify the process of converting problems from the actual world into code. ◯
   c. To provide a structured framework for managing AI projects. ◯
   d. To reduce the cost of developing AI-based remedies. ◯
2. What is the initial stage of an AI project's lifecycle?
   a. Outlining the technology specifications for the AI model. ◯
   b. Converting the real-world problem into an AI job with precise definitions. ◯
   c. Compiling and preparing the data required for the AI model's training. ◯
   d. Selecting the AI algorithm that works best for the given job. ◯
3. Which of the following metrics may be produced from a confusion matrix?
   a. Execution time and memory usage ◯ b. Recall, Precision, Accuracy, and F1 Score ◯
   c. Learning rate and Batch size ◯ d. Data acquisition and data cleaning ◯
4. What is the primary purpose of the confusion matrix while evaluating an AI model?
   a. To measure the execution time of the model ◯
   b. To identify the best algorithm for the model ◯
   c. To measure the performance of an AI model using the test data ◯
   d. To optimise the hyperparameters of the model ◯

B. Fill in the blanks.

1. The ..... involves several key stages, each building upon the previous one to develop, deploy, and maintain an AI system effectively.
2. The term ..... means collecting raw data for the purpose of reference or analysis for the project.
3. ..... refers to the techniques and tools used to visualise data through complex statistical methods.
4. ..... is the stage of the AI project cycle that is used to determine the best model for deployment.

C. State whether these statements are true or false.

1. Problem scoping means selecting a problem and finding a solution for it using AI technology. .....
2. The deployment phase of the AI project cycle involves integrating the model into the production environment and does not require any further monitoring. .....
3. Evaluation is a process that critically examines a program. .....
4. Receiver Operating Characteristic curve plots the true positive rate against the false positive rate at various threshold settings. .....

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. How many stages are there in the AI project cycle? Write their names.

Ans. There are six stages of any AI project life cycle: problem scoping, data acquisition, data exploration, modelling, evaluation, and deployment.

2. What are the parameters that are used for the evaluation of a model?

Ans. There are two important parameters which are used for the Evaluation of a model. These are:

- **Prediction:** It is the output given by the AI model using machine learning algorithm.
- **Reality:** It is the real scenario of the situation for which the prediction has been made.

B. Long answer type questions:

1. What are the difference between modelling and evaluation stages of the AI project cycle?

Ans. Modelling is the design phase of the project cycle. In this stage, we select the best way to reach the solution. It requires the process of selecting the right algorithm to develop a working model for the project. The algorithm is converted into a model. Whereas, Evaluation is the testing of the system, where we check if the model can achieve required goals or not. If the model is not fulfilling the requirements, the model or even the data can be changed. Once the developer feels the project is ready, the project will be put into working conditions and then deployed and handed over to the user. If the deployment stage is not reached, the project is of no use.

2. What are the advantages of evaluating a model?

Ans. Following are some of the advantages of evaluating a model:

- Evaluation ensures that the model is operating correctly and optimally.
- Evaluation is an initiative to understand how well it achieves its goals.
- Evaluations help to determine what works well and what could be improved in a program.
- Evaluation determines the best model for deployment.

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. What is covered by the ethical principle in terms of handling personal data?

a. Privacy ◯ b. Consent ◯

c. Transparency ◯ d. Data security ◯

2. Which ethical principle places the most focus on disclosing information about data collection methods and making it clear what data is collected and how it will be used?

a. Keeping data collection practices secret ◯

b. Collecting data without informing individuals ◯

c. Obtaining clear and explicit permission from individuals before collecting or using their data ◯

d. Sharing data with third parties without notification ◯

3. Which of the following is NOT a fundamental human right that AI solutions should respect?

a. Freedom of expression ◯ b. Right to a fair trial ◯

c. Right to own slaves ◯ d. Privacy ◯

4. What is the primary concern regarding bias in AI systems?
   a. It can be computationally expensive to address.
   b. It can lead to unfair or unjust treatment of individuals.
   c. It can slow down the development of AI models.
   d. It can make AI models less accurate.

B. Fill in the blanks.

1. ..... involves collecting and analysing information about a program's activities, characteristics, and outcomes.
2. The ..... curve is a graphical representation that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.
3. Protecting individuals' personal data and their right to control how it's used is a core principle of AI ethics called .....
4. Ensuring that AI solutions are accessible and beneficial for all members of society, regardless of background, is referred to as .....

C. State whether these statements are true or false.

1. ROC is a metric that is used to find out the accuracy of the model. .....
2. The capacity of an organisation to collect and use personal data without harming confidentiality is referred to as privacy in data ethics. .....
3. Data security is the use of strong security measures to prevent unauthorised access and breaches of personal data. .....
4. AI systems should only be designed by human rights experts to ensure they are ethical. .....

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. What measures are implemented at the deployment phase of the AI project cycle to ensure that the AI model can operate in an actual setting?
2. Why are tools for recording, reporting, and monitoring so important throughout the AI project's deployment phase?

B. Long answer type questions:

1. Explain the true positive rate and false positive rate.
2. What are the differences between ethics and morals?

AI In Life

A robotic vacuum cleaner, sometimes called a robovac or a roomba as a generic trademark, is an autonomous robotic vacuum cleaner which has a limited vacuum floor cleaning system combined with sensors and robotic drives with programmable controllers and cleaning routines. Do robot mops work? Can they be replaced by our conventional house helps?

Deep Thinking

Its vacation time, and you have a flight to catch to go for vacation. List down all examples of AI that you encounter from entering the airport till you reach your hotel room.

Ai Lab

Computational Thinking

1. Make a presentation on the stages of AI project cycle.
2. Using an online AI tool, fetch the information about principles of AI ethics. Using this information make a collage in any word processor package.

Answers

AI Quiz Section A (Objective Type Questions)

A. 1. c 2. b 3. b 4. c

B. 1. AI project cycle 2. data acquisition 3. Data exploration 4. Model evaluation

C. 1. True 2. False 3. True 4. True

Ai Ready 1

1. What is the meaning of deployment? Can this term be used only for AI projects?

2. How is prediction used in our daily life? Give an example.

3. List any two examples of TRUE positive cases that exist in our day-to-day life.

4. Explain the term "Inclusion" as one of the principles of AI ethics. Give an example.

UNIT-2

DATA LITERACY

Learning Outcomes

- Defining Data Literacy
- Why is Data Literacy Essential?
- How to Become Data Literate?
- Data Security and Privacy
- Data Acquisition/Acquiring Data
- Data Processing and Data Interpretation
- Types of Data Interpretation
- Using Tableau for Data Presentation
- Data Pyramid and Its Different Stages
- Impact of Data Literacy
- Data Literacy Process Framework
- Acquiring, Processing, and Interpreting Data
- Usability, Features, and Preprocessing of Data
- Methods of Data Interpretation
- Importance of Data Interpretation

Data refers to any collection of raw facts, figures, or statistics that can be stored and processed by a computer. It can be in different forms like numbers, text, images, audio, and video etc.

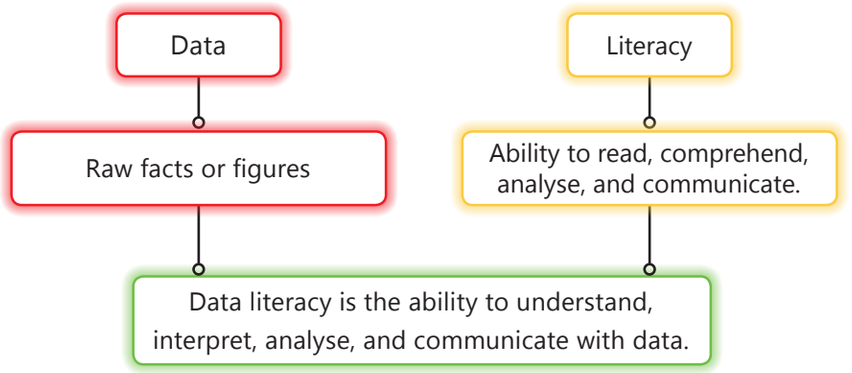
Defining Data Literacy

Literacy refers to the ability to read, comprehend, and use information effectively.

Data + Literacy = Data Literacy

Data literacy means knowing how to understand, work with, and talks about data. It's about being able to collect, analyse, and show data in ways that make sense.

Data literacy is essential because it enables individuals to make informed decisions, think critically, solve problems, and innovate.



Experiential Learning

Video Session

Scan the QR code or visit the following link to understand the Data Literacy:

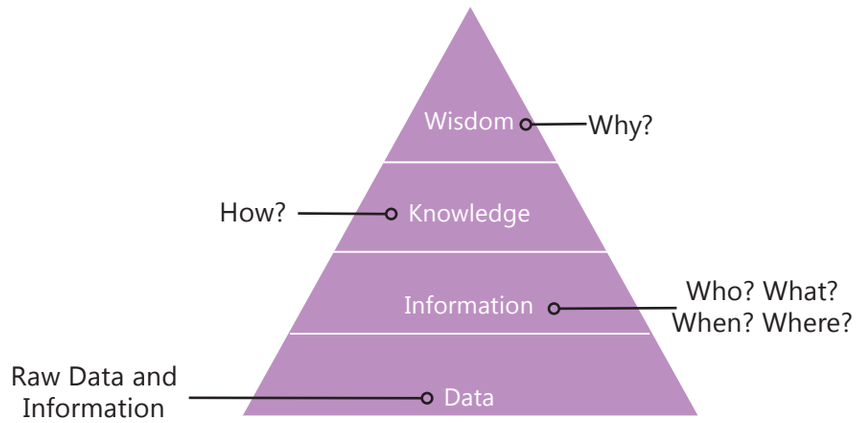
https://www.youtube.com/watch?v=yhO_t-c3yJY

1. What are the takeaway from the given video?

2. What is the definition of data and data literacy, and how do these concepts impact our ability to understand and use information effectively?

Data Pyramid and Its Different Stages

The data pyramid is a conceptual model that illustrates the hierarchical structure of data processing, depicting the progressive transformation of raw data into actionable wisdom. It starts with raw **data**, which initially has no use. Through processing and analysis, this data evolves into meaningful **information**, then **knowledge**, and ultimately **wisdom**. This transformation enables informed decision-making and a deeper understanding of the world around us.

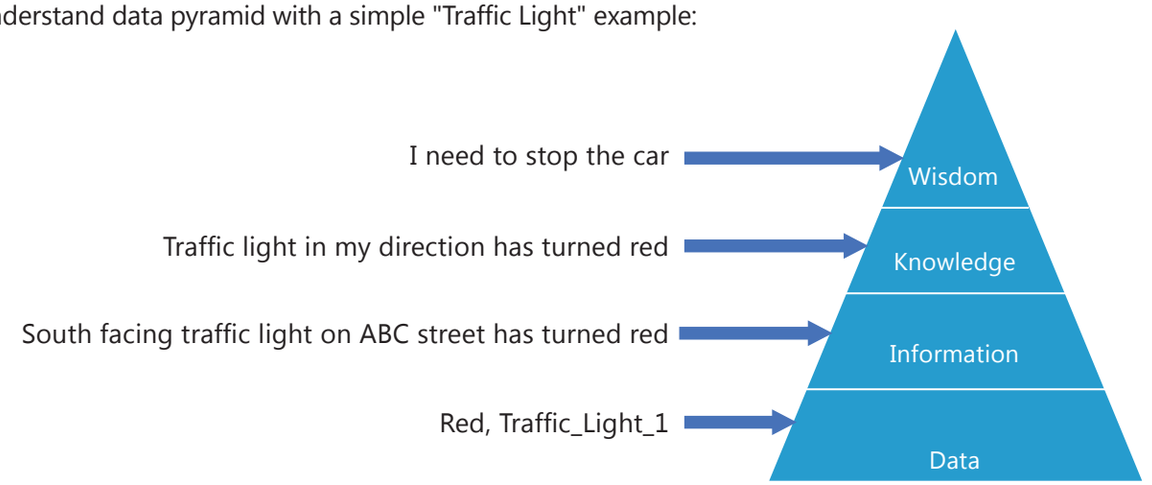


Different Stages of the Data Pyramid

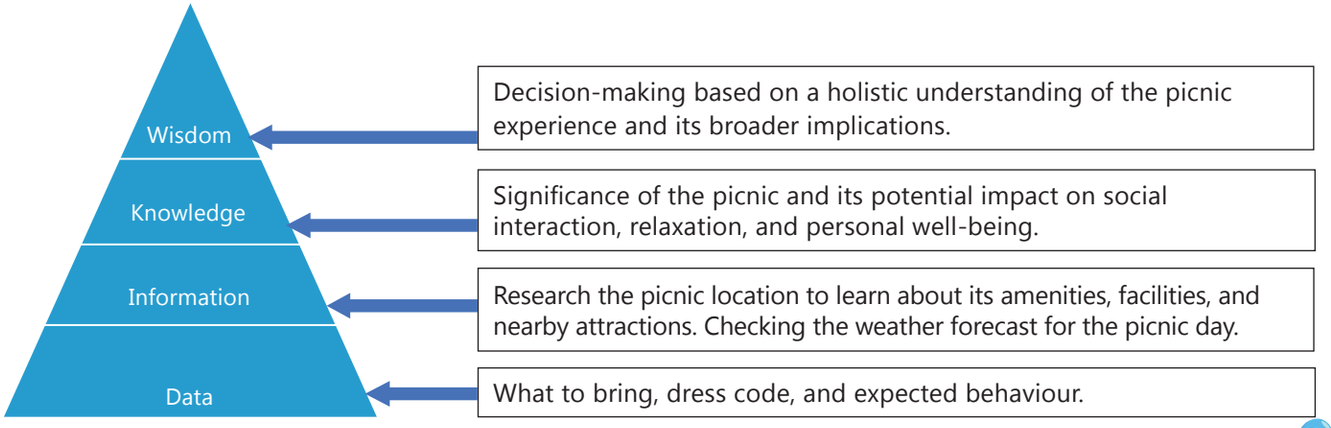
The DIKW pyramid represents the relationships between data, information, knowledge and wisdom. Each building block is a step towards a higher level - first comes data, then is information, next is knowledge, and finally comes wisdom. Each step answers different questions about the initial data and adds value to it. Let's learn about them.

- **Data (Base Level)**
  - In this stage data is in its most basic form, unprocessed and unstructured.
  - It has no meaning and is not very useful in this form.
- **Information**
  - It is a processed data that collectively carries a logical meaning.
  - It is obtained by analysing raw data to make it easier to measure, visualise and analyse for decision-making.
  - By asking relevant questions about 'who', 'what', 'when', 'where', etc., we can derive valuable information from the data and make it more useful for us.
- **Knowledge**
  - It is useful information that leads to a deeper understanding.
  - It represents a more profound comprehension of **how** things happen.
  - It is the ability to use information to achieve desired output.
- **Wisdom (Top Level)**
  - It represents the highest level of understanding. Wisdom is knowledge applied in action.
  - It is the ability to understand why things are happening in a particular way.
  - It involves critical thinking to interpret data and make good consistent decisions.

Let's understand data pyramid with a simple "Traffic Light" example:

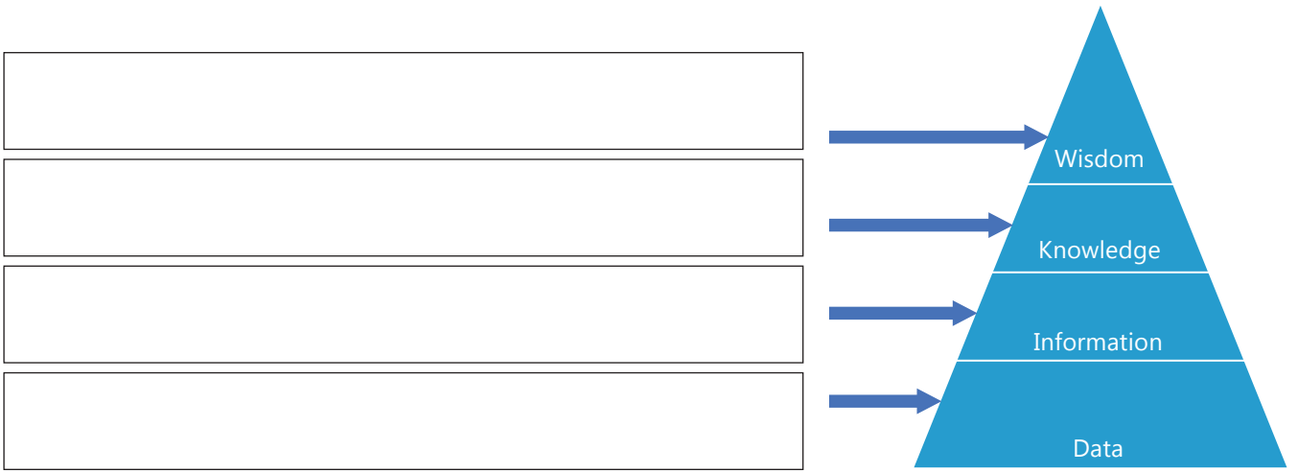


Let us see another example of creating a data pyramid for planning a picnic as a student after receiving the official letter from school:



Ai Task

Can you try another example of creating a data pyramid for preparing a speech on the occasion of "Technology Day"?



Sudha attended three consecutive presentations and rated them as follows:

- the first was 'outstanding'
- the second was 'poor'
- the third was 'satisfactory'

Can you filter the data from this statement? Are the ratings of the same type?

Why is Data Literacy Essential?

Data literacy can equip individuals with skills and knowledge to improvise in a tech and data-driven world. There are countless reasons why data literacy is critical to an organisation. Some of these are:

- Data literacy enhances **decision-making ability** in individuals based on evidence. Based on sources of data, emerging trends and interpretations, individuals can make decisions that are data-driven.
- Data literacy is able to cultivate **critical thinking** skills to understand and explore data's implications by questioning assumptions, reaching logical conclusions, identifying patterns, and evaluating evidence and data accuracy.

- Data literacy helps in **analytically producing solutions** to problems that help people in developing critical thinking skills. It enables user to tackle complex problems and derive meaningful relevance.
- Data literacy fuels **innovation** by providing tools and techniques to explore data from different perspectives. It helps in innovating to meet the requirements of emerging trends and market demands.

Impact of Data Literacy

Data literacy has an immense impact on various aspects of society like business, education, healthcare, and public policy as given below:

- **Business:** It can lead to growth and innovation and helps companies maintain their competitive edge in a tech-focused world. Data literate employees can effectively analyse data to gain insights into market trends, customer behaviours, and operational performance.
- **Education:** It empowers the teaching-learning process. In the field of education, data literacy is essential for both students and educators. Students can better understand and analyse information, while educators can use data to tailor instruction and identify areas where students may need additional support. Students can engage more deeply with course material, particularly in STEM fields.
- **Healthcare:** Healthcare professionals can use data to improve diagnostics, treatment plans, and patient monitoring. Hospitals and clinics can use data to optimise resource allocation, reduce waste, and improve operational efficiency.
- **Public policy:** Policymakers can use data to design, implement, and evaluate policies more effectively. Data literacy promotes transparency, allowing the public to hold policymakers accountable through data-driven evidence.
- **Social equity:** Data literacy can highlight disparities in areas such as education, healthcare, and employment and can promote social equity. It helps ensure that resources are distributed effectively to areas of greatest need.

Ai Task

Problem Solving & Logical Reasoning

Let us do an activity to understand the impact of data literacy.

Impact of News Articles (Select any trending news)

Session Preparation Logistics:

For a class of 40 Students [Pair Activity]

Materials Required:

Item	Quantity
Online Data Sources Clues	NA
Computers	20

Purpose:

The purpose of this activity is to engage participants in various scenarios that involve collecting data and analysing its sources. Emphasising the importance of validating data sources, the aim is to instil the concept of data literacy. By understanding how authentic data sources contribute to reliable and unbiased decision-making, participants will develop critical skills for navigating and interpreting data effectively.

Brief: [Pair Activity]

Participants will search the internet for data sources, extracting key information to support their decisions.

Author of the Source	Weblink to the Source	How was the situation described by the Source	Key figures in the source

You have to rank the sources of the news articles from most accurate to least, state reasons for your choice.

Rank	Data Source	Remarks

So, we can conclude that every data tells a story, but we must be careful before believing the story.

How to Become Data Literate?

Data literate is a person who can interact with data to understand the world around them and derive meaningful information from data. Some key points that help you to become a data literate are as follows:

- **Data identification and sourcing:** Identify the source of data to find whether the data is reliable or not.
- **Understand the basics**: Learn the concepts of data, types of data and how it can be used as not all data is suitable for every kind of analysis.
- **Learn data analysis tools**: There are many data analysis apps available that can be learned in order to understand the impact of the right data. Analysis involves using statistical tools or software to interpret the data. This can include calculating simple averages for more complex tasks.
- **Gain statistical knowledge**: Statistics play a vital role in data literacy. It's one of the vital components that must be learned before you dive into the data-driven world.
- **Use data visualisation**: Understand the techniques of data visualisation graphics, and charts. Tools like tableau, matplotlib, and python can be used effectively.
- **Learn data manipulation**: Understanding how to manipulate data to meet the requirements is also one of the key factors. Methods like filtering, sorting, grouping, and omitting are essential for extracting insights from large dataset.
- **Practise data cleaning**: Learning to remove data redundancy and data inaccuracy is essential to be data literate. This may include dealing with missing values, removing outliers, or transforming data into a format suitable for analysis.

Ai Reboot

1. What does the term "data literacy" mean?

   _____

2. What makes data literacy indispensable?

   _____

Ai Task

Creativity & Innovativeness

Let us do an activity:

Scenario: Buying a video game online.

Data literacy helps people research products while shopping over the Internet.

How do you decide the following things when you are shopping online?

1. Which is the cheapest product available?

   _____

2. Which product is liked by the users the most?

   _____

3. Does a particular product meet all the requirements?

   _____

4. When will the product delivered?

   _____

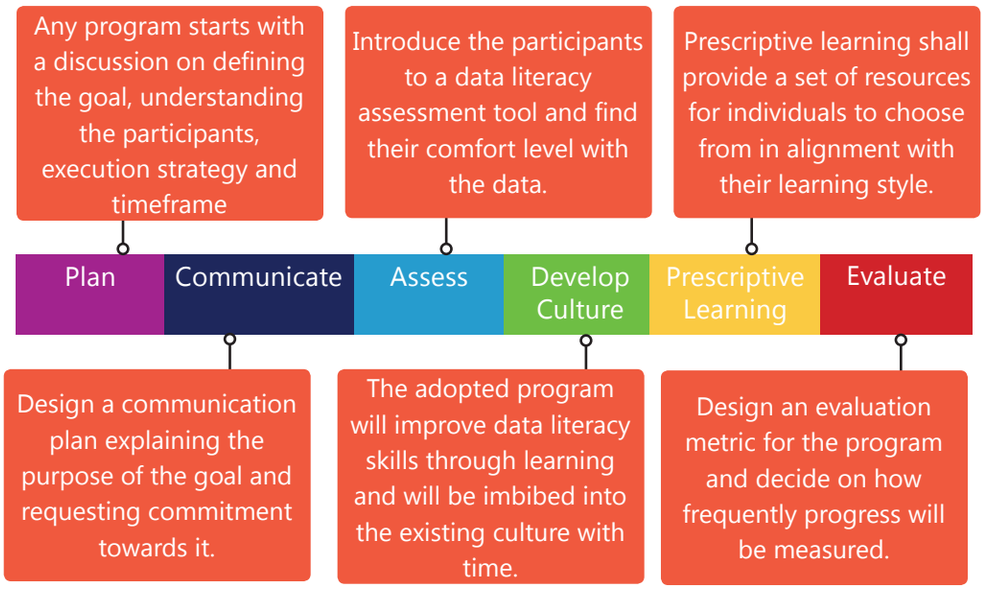
A data literate person can:

- Filter the category as per the requirement – If the budget is low, select the price filter as low to high.
- Evaluate the reviews and ratings to have a better understanding of the product by checking the user ratings of the products.
- Check for specific requirements in the product.

Data Literacy Process Framework

The data literacy framework provides a comprehensive and structured approach to develop the necessary skills for using data efficiently and with all levels of awareness. Each level builds upon the previous one, fostering a deeper and more understanding ability to work with data.

Here are the typical levels of awareness in a Data Literacy Process Framework:



Plan

Planning sets a clear roadmap and structured approach to enhance data literacy across different levels of awareness. You can:

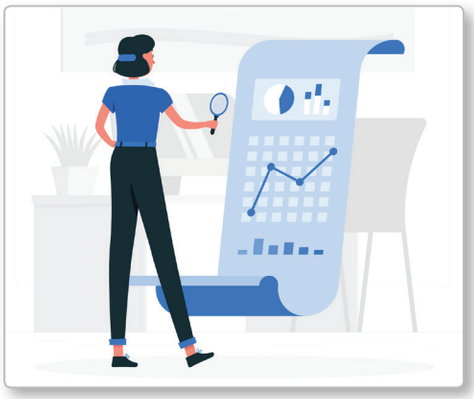
- define the specific and measurable goals.
- develop a timeline and milestones for achieving these goals.
- identify and allocate resources needed (e.g., budget, tools, personnel).

Communication

Clear and consistent communication about data literacy will ensure an efficient data literacy framework within an organisation. This involves:

- designing a well-formulated communication plan to explain the purpose of the goal and requesting commitment from the team towards it.
- sharing stories and case studies that demonstrate the positive impact of data literacy. It will help in motivating and encouraging the team.
- monitoring continuously the effectiveness of communication efforts and make adjustments as needed. It will help in minimising the risk of any associated costs.

Assess



Introducing participants to data literacy assessment tools and finding out how comfortable they are with data is crucial in a data literacy program. This process will help to:

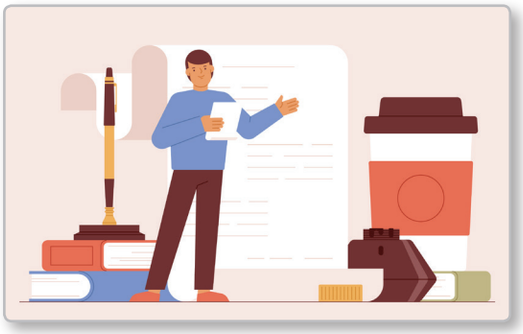
- see what skills they already have
- identify what they need to learn
- create customised training plans for them

Develop Culture

To integrate data literacy skills into the organisational culture it is important to make data-driven decision-making a fundamental part of everyday work. This will help:

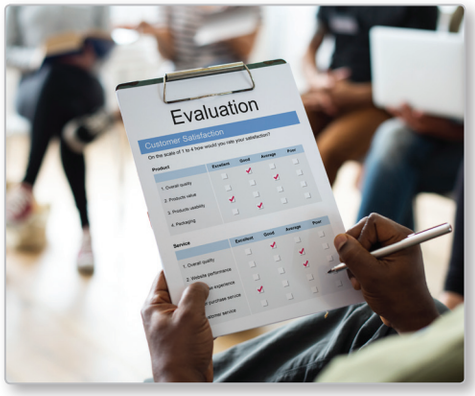
- leaders play a crucial role in fostering this data-driven culture in an organisation.
- in training programs to build data literacy skills through learning across all levels.
- in encouraging a collaborative environment that will lead to imbibing this new culture into the existing culture with time.

Prescriptive Learning



By implementing a prescriptive learning approach, organisations can provide a set of diverse resources that align with individual learning styles. This approach ensures that there is:

- customised learning journeys tailored according to different people(for example different educational background) based on individual needs and preferences.
- a variety of learning materials that cater to different learning styles and help in easier grasping of concepts.
- enough leverage or advantage to the learners to progress at their own pace, accommodating their schedules and learning speeds.
- create an environment that make learners to feel comfortable and gain new skills in an environment to supports continuous learning and encourages self-directed exploration.
- each participant can choose the materials and methods that work best for them, leading to more effective learning and greater improvement in data literacy skills over time.



Evaluate

Designing an evaluation metric for the data literacy program involves creating a structured framework to assess participants' progress and the effectiveness of the program overall. It helps to:

- improve participants' overall data literacy skills.
- establish clear criteria to measure the success of the data literacy program and individual participant growth.
- establish a schedule for assessing participant progress to monitor their development over time.

Data Literacy Framework—An Iterative Process

This means the development and enhancement of data literacy skills are not static or one-time event. Instead, they evolve through continuous cycles of learning, application, and refinement.

- **Learning**
  - ✶ Learning is the initial stage where individuals acquire new knowledge and skills related to data literacy.
  - ✶ Individuals engage in various learning activities such as formal training sessions, online courses, reading materials, and hands-on workshops to gain insights into data concepts, tools, and methodologies.
- **Application**
  - ✶ Application involves putting acquired knowledge and skills into practice in real-world contexts.
  - ✶ Individuals apply what they have learned to analyse real datasets, solve data-related problems, and make informed decisions.
  - ✶ They are engaged in data projects, experiments, or simulations to gain practical experience and develop a deeper understanding of data concepts.
- **Refinement**
  - ✶ Refinement focuses on reflecting the past experiences, identifying areas for improvement, and enhancing data literacy skills over time.
  - ✶ Feedback from peers, mentors, supervisors, and outcomes of data-related activities informs the refinement process, guiding individuals to adjust their practices accordingly.

Ai Reboot

List down the levels of data literacy process framework.

Data Security and Privacy

The terms data security and data privacy are often used interchangeably, but they mean different things. Data privacy determines who can access the data, while data security involves tools and policies to restrict access to the data.

Data Privacy — It is governing how data is collected, shared, and used.

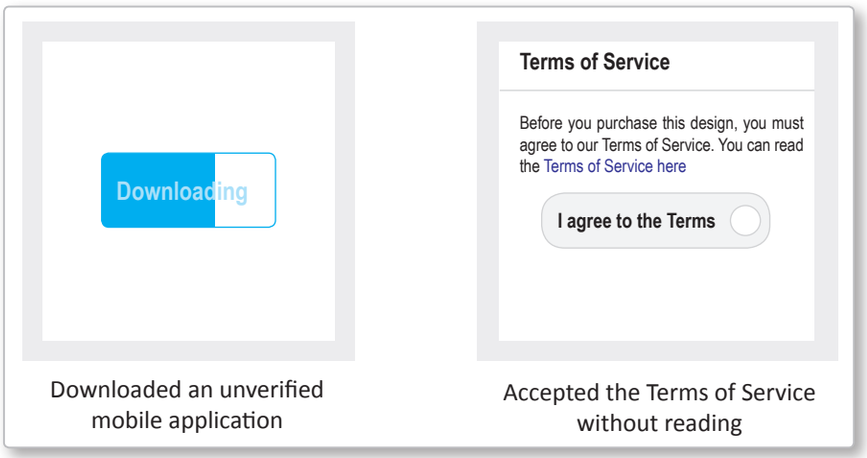
Data Security — It is protecting data from attackers who might want to misuse it.

What is Data Privacy?

Data privacy referred to as information privacy is concerned with the proper handling, processing, storage, and usage of sensitive data including personal data and other confidential data, such as certain financial data and intellectual property data, to meet regulatory requirements as well as protecting the confidentiality and immutability of the data.

So, when we talk of data privacy, it is expected that any platform or individuals who have access to this information ensure that data is used in a way that respects and fulfils legal requirements and compliance of handling privacy rights. These include how the data is collected and shared for usage and who all have access to data.

Given below are two examples which may compromise our data privacy.



Downloaded an unverified mobile application

Accepted the Terms of Service without reading

Why is Data Privacy Important?

Data privacy is important because:

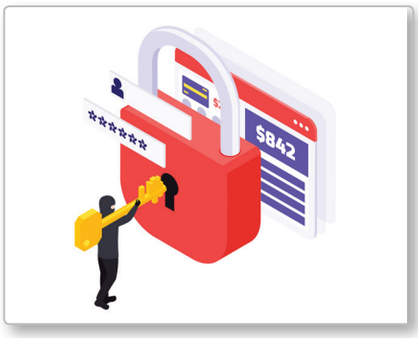
- a data breach at a government agency can put top secret information in the hands of an enemy country.
- a data breach at a hospital can put personal health information in the hands of those who might misuse it.
- a data breach at a corporation can cause put proprietary data in the hands of a competitor.
- a data breach at a school can inconvenience to the parents, by getting continuous calls from tuition and coaching centers cause annoyance and stress.

The following best practices can help you ensure data privacy:

- Understanding what data you have collected, how it is handled, processed, used, and where it is stored.
- Only necessary data required for a project should be collected.
- User consent while data collection must be of utmost importance.

What is Data Security?

Data security involves safeguarding digital information from unauthorised access, corruption, or theft throughout its entire lifecycle. Essentially, security means protecting anything from theft and misuse. Data security is related to securing sensitive data. It aims to prevent unauthorised access, theft, or corruption of data, regardless of whether the data is personal or not. Systems and networks must be established to prevent malicious and fraudulent activities from harming, destroying, misusing, or stealing crucial digital data.



Why is Data Security Important?

Cyber attacks are becoming more frequent as a result of the growing volume of data stored in the cloud. The best course of action given the volume of traffic being produced is to regulate and secure the transmission of private or sensitive data everywhere that it is known to exist. Avoid entering sensitive information, such as your address, PAN, or Aadhar number on unrecognised and unsafe websites.

The most possible reasons why data security is more important now are:

- A constant fear cyberattacks affect all people.
- The fast-technological changes will boom cyberattacks.
- A persistent fear everyone is impacted by cyberattacks.
- Rapid technical advancements will increase the frequency of cyberattacks.

Types of Data Security Controls

Different types of data security controls are as follows:

- **Strong Passwords:** A strong password is a combination of atleast 8 characters with upper and lower-case letters, numbers, and special characters that is difficult for unauthorised individuals or automated programs to guess or crack. It is a very basic step that one should take and never share the same with even the most trusted ones. Avoid using birth dates, anniversary dates, and common combinations of numbers.

  Some examples of strong passwords are: m#P52s@ap$V, "N4&vQ2! p".

- **Authentication:** It also refers to multi-factor authentication (MFA) and is an additional security layer in online data systems. After a user enters their password to log in, MFA requires them to provide one or more additional forms of authentication to verify their identity. This could include one-time generated code as a security token in smartphones or emails or a fingerprint or facial recognition, or Passwords or PINs.

- **Access Controls:** Access controls refer to the security measures and protocols to restrict access to sensitive data, ensuring that only authorised individuals or entities can view, modify, or interact with it. This reduces the risk of unauthorised access by limiting the number of users who can interact with sensitive data.

- **Data Backup:** Data backup refers to the process of creating copies of data to ensure that it can be restored in the event of data loss due to natural disasters, accidents, cyber-attacks, or other unexpected events. Sometimes physical backup media is used to secure in access-controlled environments. Another method to secure data can be the cloud backup which is considered more reliable.

- **Encryption:** Encryption is a security technique that transforms readable data (plaintext) into an unreadable format (ciphertext) using an algorithm and an encryption key. This process ensures that only authorised individuals with the correct decryption key can access the original data. Encrypted data is meaningless if captured by attackers.

- **Data Disposal:** Data disposal refers to the process of securely destroying or deleting data that is no longer needed to prevent unauthorised access, recovery, and misuse. Proper data disposal practices are essential to ensure that sensitive and confidential information does not fall into the wrong hands. Paper documents, CDs, DVDs, and other physical media can be shredded to render them unreadable. It is also referred as data erasure or data destruction.

- **Firewall and Antivirus Software:** Using firewall and antivirus software can stop and alert users of any suspicious activity happening on their devices. With the timely updated versions of the same, can go a long way in ensuring data security. Firewalls use pre-configured rules to inspect all the packets entering and exiting a network and, therefore, help stop malware and other unauthorised traffic from connecting to devices on a network.

- **Data Masking:** It obscures data so that, even if criminals exfiltrate it, they can't make sense of what they stole. Unlike encryption, which uses encryption algorithms to encode data, data masking involves replacing legitimate data with similar but fake data. This data can also be used by the company in scenarios where using real data isn't required, such as for software testing or user training.

- **Training:** Corporates must take up regular Data Security sessions of their staff to sensitise them about following the data protection processes being implemented and the importance of doing so. Making them conscious of suspicious emails, links that they might receive, not leaving their devices unlocked when unattended, keeping software's up to date and not sharing passwords, are some of the things that can be taken up.
- **Audits and Testing of Security System:** Regular audits and testing of security policies, integrated malware protection, firewalls, Wi-Fi connections security, Hardware-based security, checking applications security, email security and compliance also play very important role in maintaining data privacy and providing data security.
- **Other Basic Preventions:** Being aware of surroundings and threats from insiders, complying with security regulations which might be shared by entrusted agencies or bodies which track online cyber activities all across the world are few other ways to provide cyber security.

Differences between Data Security and Data Privacy

Data Privacy	Data Security
Data privacy ensures the ethical and lawful use of data.	Data security ensures the protection of data from unauthorised access and breaches.
It focuses on how data is collected, used, shared, and stored so that the rights of individuals over their data is protected.	It focuses on safeguarding personal data, business data, intellectual property, and many more from various threats.

How are Data Security and Data Privacy related to AI?

Data security and data privacy are crucial components of Artificial Intelligence (AI).

Data Security in AI

AI systems often rely on vast amounts of data for training and operation. Unauthorised access and tampering could lead to inaccurate AI models and compromised outcomes. Many AI applications process sensitive data, such as personal, financial, or health-related information. Strong data security measures can stop data breaches and unauthorised access.

Data Privacy in AI

Data privacy brings the ethical use of AI. This ensures that AI systems comply with data privacy laws and regulations (such as GDPR, CCPA) to help protect individuals' rights and maintain public trust. AI systems must ensure that data is collected, shared, and used in ways that users have explicitly consented to, maintaining transparency and trust.

Best Practices for Cyber Security

Cybercrime is undoubtedly one of the fastest-growing crimes in the world and it continues in all industries. To stay protected from cyberattacks, one needs to be aware of the most up-to-date cyber security tips and best practices.

Cyber security involves protecting computers, servers, mobile devices, electronic systems, networks, and data from harmful attacks. The best practices for cyber security are constantly evolving to keep up with the cyber threats. Reference Links:

Interdisciplinary

Video Session

Scan the QR code or visit the following link to understand the Internet safety tip for online security:

https://www.youtube.com/watch?v=aO858HyFbKI

What are some key takeaways from the video session on online security, specifically regarding best practices for protecting personal information?

Ai Task

Ethical & Moral Reasoning

Refer the given link and answer the following questions:

https://www.cbse.gov.in/cbsenew/documents/Cyber%20Safety.pdf

1. What are the key online security practices discussed in the PDF that can help protect personal information from cyber threats?

2. What essential digital etiquette guidelines were highlighted in the PDF to ensure respectful and effective online communication?

The following are the list of **Do's** and **Don't 's** to be followed for best practices of cyber security:

Do's	Don't's
• Use strong, unique passwords with a mix of characters for each account. • Activate Two-Factor Authentication (2FA) for added security. • Download software from trusted sources only and scan files before opening. • Prioritise websites with "https://" for secure logins. • Keep your browser, OS, and antivirus updated. • Adjust social media privacy settings for limited visibility to close contacts. • Always lock your screen when away. • Connect only with trusted individuals online. • Use secure Wi-Fi networks. • Report online bullying to a trusted adult immediately. • Do use privacy settings on social media sites to restrict access to your personal information.	• Avoid sharing personal info like real name or phone number. • Don't send pictures to strangers or post them on social media. • Don't open emails or attachments from unknown sources. • Ignore suspicious requests for personal info like bank account details. • Keep passwords and security questions private. • Don't copy copyrighted software without permission. • Avoid cyberbullying or using offensive language online. • Don't respond to phone calls or emails asking for confidential data. • Don't leave wireless or Bluetooth turned on when not in use.

Ai Reboot

1. How are data security and data privacy related?

   _____

2. List down the practices that can help you ensure data privacy.

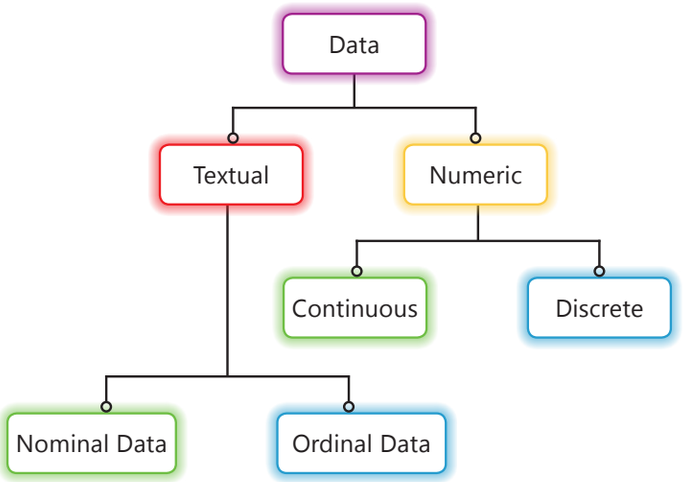
   _____

Acquiring, Processing, and Interpreting Data

Working with data involves three key steps: **acquiring, processing, and interpreting**. First, gather data from sources like surveys and databases. Next, process it by cleaning and organising it for accuracy. Finally, analyse the data to find patterns and insights that help make informed decisions.

Types of Data

In statistics, various types of data are gathered, analysed, interpreted, and presented. These data consist of individual factual pieces recorded for analysis. Data analysis involves interpretation and presentation, producing statistics to get some meaningful insight from that data. Data classification and handling are crucial processes that use multiple tags and labels to define data, ensuring its integrity and confidentiality. Artificial Intelligence is crucial, with data serving as its foundation. We come across different types of data and information every day.



Data can be broadly classified under Textual data and Numeric Data as explained.

Textual Data (Qualitative Data)

Textual data is the information that is written or expressed using words and language. It includes things like articles, books, emails, messages, and any other written content. Instead of numbers, it's made up of letters, words, and sentences that convey meaning and information. Qualitative data is also called Categorical data.

Example: "Learning AI is fun"

Qualitative data is further categorised into two categories that includes,

- Nominal data
- Ordinal data

Nominal Data

It consists of categories or names that cannot be ordered or ranked. Nominal data is often used to categorize observations into groups, and the groups are not comparable. Examples of nominal data include gender (Male or Female), and blood type (A, B, AB, O).

Ordinal Data

It consists of categories that can be ordered or ranked. Ordinal data is often used to measure opinions, where there is a natural order to the responses. Examples of ordinal data include education level (Elementary, Middle, High School, College), job position (Manager, Supervisor, Employee), etc.

Numeric Data (Quantitative Data)

Numerical data means information that's in numbers, not words or descriptions. It's often called quantitative data because it's collected as numbers and can be used for math and stats. For instance, if you know the total number of workers and how many are men, you can figure out how many are women by subtracting. This ability to do math with numerical data makes it great for doing statistics and analysing data.

For example, Marks, Temperature, Height, Weight, etc.

Numeric data can be further classified as:

Continuous Data	Discrete Data
Continuous data can take as a numeric value given within a range.	Discrete data refers to distinct single values. It consists of whole numbers without decimal parts that represent distinct categories or values.
Continuous data is measurable.	Discrete data is countable.
This type of data can be infinitely subdivided and often includes decimal points.	Discrete data cannot be subdivided meaningfully.
Often used to analyse using statistical techniques such as mean, median, standard deviation, and correlation.	It is used to analyse using frequency distributions, bar charts, and probability distributions.
Example: Dimensions of classroom, Height, Weight, Temperature, Time, etc.	Examples: Number of girls and boys in class, Number of subjects in class 9th, Count of anything.

Qualitative Data versus Quantitative Data

Quantitative Data	Qualitative Data
Data is depicted in numerical terms	Data is not depicted in numerical terms.
Can be shown in numbers and variables like ratio, percentage, and more	Could be about the behavioural attributes of a person, or things.
Example: 100%, 1:3, 123	Examples: loud behaviour, fair skin, soft quality, and more.

AI Domains and Type of Data

Various types of data are utilised across different domains to train models, make predictions, and generate insights. Here are the types of data commonly used in three key domains of AI.

Natural Language Processing (NLP)

Natural Language Processing (NLP) is a field of computer science and a subfield of artificial intelligence that aims to make computers understand human language. It all about teaching, training computers to understand and work with human language. Types of data used in NLP are:

- **Textual data**: This includes a wide range of written text, such as articles, books, emails, social media posts, web content, PDF files, etc.
- **Audio data:** Audio recordings of spoken language, which are transcribed into textual data.

Computer Vision

Computer Vision is a field of artificial intelligence (AI) that uses machine learning and neural networks to teach computers to derive meaningful information from digital images, videos and other visual inputs. It is like giving eyes to computers. It helps them look at pictures and videos from the real world and understand what they're seeing. With Computer Vision, computers can figure out what's in a picture or video, just like we do. They can recognise objects, people, and even actions happening in videos.

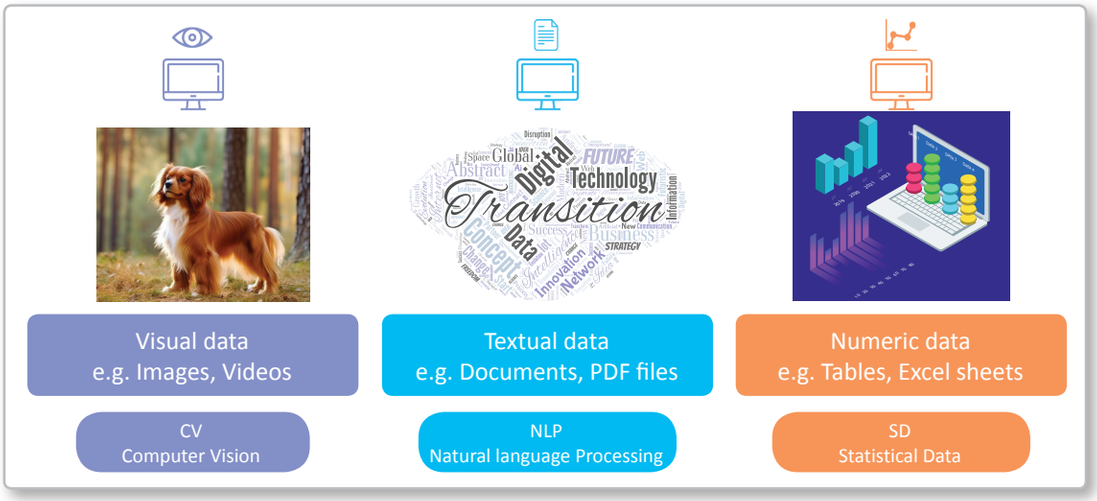
Types of data used in computer vision include:

- **Image data**: Digital images captured by cameras or satellite imagery, medical scans, and surveillance footage.
- **Video data**: Video data captured using camera

Statistical Data

Statistical data analysis involves collecting, analysing, and interpreting data to discover patterns and insights. It uses structured and time series data for tasks such as hypothesis testing, regression analysis, and predictive modeling. This domain is essential for making data-driven decisions in business, science, and social sciences. Types of data used in statistical data are:

- **Numeric data:** Data taken from tables, Excel sheets, etc.
- **Time series data:** Data points collected or recorded at specific time intervals, such as stock prices, weather data, sensor readings, and economic indicators.



Ai Task

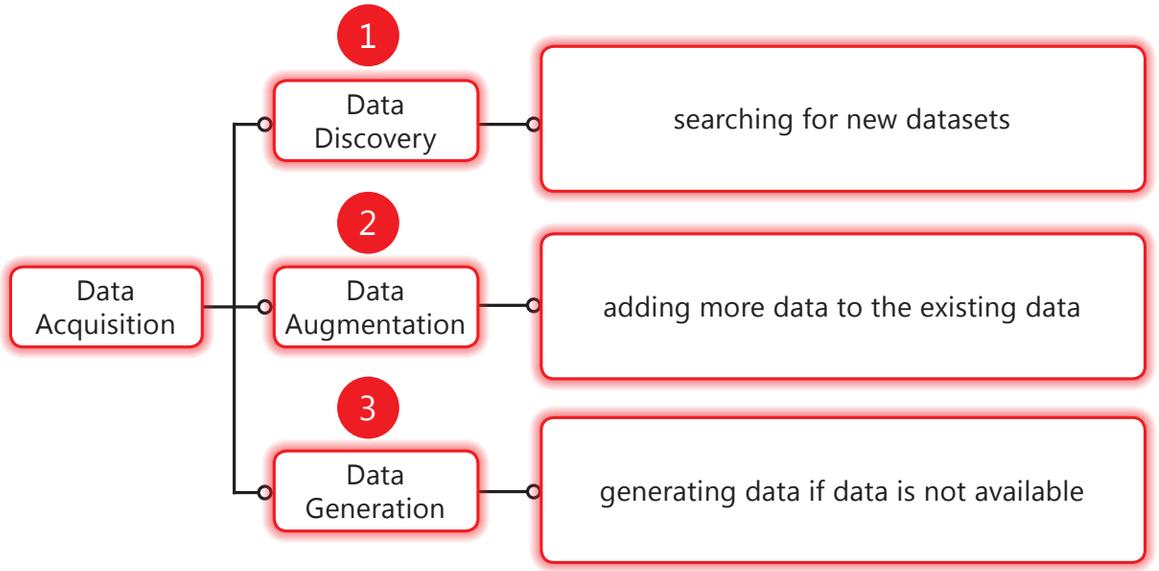
Experiential Learning

Let us now do an exercise to categorise the given data as textual data (qualitative data) or numeric data (quantitative data)

Temperature
Gender
Show size
Comment on social media
Favourite colour
Newspaper article
Population number in a state
Email
Heart rate
Weight of a person

Data Acquisition/Acquiring Data

Data acquisition, also known as acquiring data, refers to the procedure of gathering data like raw facts, figures or statistics from relevant sources either for reference or for analysis needed in AI projects. This involves searching for datasets suitable for training AI models. The process typically comprises three key steps and plays a crucial role in obtaining and preparing data for analysis.



Let's say we want to collect data for making a CV model for a self-driving car.

The three key steps involved in data acquisition are given below:

Step 1: Data Discovery

Data discovery is about hunting for valuable information in different places, checking if it's good quality, and making sense of what we find. In the above example:

- We will require pictures of roads and the objects on roads.
- We can search and download this data from the Internet.

Step 2: Data Augmentation

Data augmentation is the process of increasing the amount and diversity of data. We do not collect new data, rather we transform the already present data. Data augmentation means increasing the amount of data by adding copies of existing data with small changes. The image given here does not change, but we get data on the image by changing different parameters like colour, rotation, flipping and brightness. New data is added by slightly changing the existing data.

In the above example:

- We apply flipping and rotation transformation to create variations of the original images.
- We also simulate occlusions such as objects partially blocking the view to train the model to handle obstructed scenarios.

Step 3: Data Generation

Data generation refers to generating or recording data using sensors. Recording temperature readings of a building is an example of data generation. Recorded data is stored in a computer in a suitable form.

In the above example, of self-driving car. Data acquisition is done for creating fake driving scenarios with different road conditions, traffic patterns, weather, and lighting to cover many possible situations.

Ai Task

Experiential Learning

Visualise that you are in a big, mysterious forest and searching for hidden treasure.

Write four observations you will be making for finding the treasure and categorise them under the heads viz: data discovery, data augmentation and data generation.

Observations	Categories
1.	.....
2.	.....
3.	.....
4.	.....

Sources of Data Acquisition

Data can be acquired from various sources, classified as either primary or secondary.

Primary Data Sources

The data generated from the experiment is an example of primary data. Some of the sources for primary data include surveys, interviews, experiments, etc. Here is an Excel sheet showing the data collected for students of a class.

	A	B	C	D	E	F
1	**Name**	**Gender**	**Address**	**Phone**	**Percentage**	
2	Ridham	Female	A-12 Ramesh Nagar	5555555555	92	
3	John	Male	K-15 Pitampura	5557777888	89	
4	Rohit	Male	G-2 Faridabad	2323232323	90	
5	Gaurav	Male	L-3 Karolbagh	3987678931	87.5	
6						

Secondary Data Sources

Secondary data sources are the external sources for collecting data, rather than generating it personally. Some sources for secondary data collection are: Published Literature, Government Publications, Market research reports, etc.

Best Practices for Acquiring Data

1. **Set Clear Goals:** Understand why you need the data and what you want to achieve; specify the type, format, and detail level required.
2. **Identify Data Sources:** Use primary data you collect yourself (surveys, interviews) and secondary data from others (reports, databases).
3. **Evaluate Sources:** Ensure data sources are trustworthy, relevant, accurate, and current; get necessary permissions and respect privacy.
4. **Collect and Prepare Data:** Use surveys, interviews, sensors, and web scraping; clean data by fixing errors, removing duplicates, and anonymising.
5. **Validate, Document, and Store:** Cross-check and sample for accuracy, keep detailed records and meta-data, store data securely, and regularly update it while following laws and regulations.

Checklist of factors that make data good or bad

Here's a checklist of factors that can help determine whether data is of good quality (good) data or poor quality (bad) data:

Good Data

- Information is well structured
- It is accurate
- It is consistent
- It is cleanly presented
- Contains information which is relevant to our requirement

Bad Data

- Information is scattered
- Contains a lot of incorrect values
- Contains missing and duplicate values
- It is poorly presented
- Contains information which is not relevant to our requirement

Data Acquisition from Websites

The process of collecting data from websites using software is called Data Scraping. It is a common method for extracting information from websites. It is commonly known as Web Scrapping.

Just like you might copy text from a book or from your friend, data scraping involves copying information from websites. But instead of doing it manually, we use special tools or programs to do it automatically. These tools can navigate websites, find the information we want, and copy it into the required format.

We scrape websites to get data needed for different reasons. It can be collecting prices for market research, news articles for analysis, or customer reviews for a product.

While web scraping is not illegal, using data without permission is illegal. Think of web scraping like picking fruit from someone else's garden without their permission and it is also about what you do with the fruit afterwards

Using data with permission is legal and ethical, just like getting permission from the owner of garden to take fruit. It's all about respecting the rights of the website owner and following the rules.

Ethical Concerns in Data Acquisition

While gathering data and choosing datasets, certain ethical issues can be addressed before they occur:

Bias	Take steps to understand and avoid any preferences or partiality in data
Consent	Take necessary permissions before collecting or using an individual's data
Transparency	Explain how you intend to use the collected data and do not hide intentions
Anonymity	Protect the identity of the person who is the source of data
Accountability	Take responsibility for your actions in case of misuse of data

Usability, Features, and Preprocessing of Data

Data is indeed a collection of information gathered through various means such as observations, measurements, research, surveys or analysis. This information can include a wide range of elements like facts, numbers, names, figures, or descriptions of things. To make data easier to understand and analyse, it is often organised into formats such as graphs, charts, or tables.

Usability of Data

Let's take an example of completing a school project. You need clear instructions, a neat workspace, and accurate information. Similarly, using data effectively relies on its clarity, organisation, and accuracy. There are three primary factors determining the usability of data:

1. **Structure of Data:** Defines how data is stored. Data needs to have a clear structure. It should be organised in a way that makes sense so that it can be used effectively.

   Like when your mother starts cooking your favourite food she ensures before cooking that all ingredients are available and are put in order for smooth and organised cooking.

   For example:

   Marks of a students arranged in a spreadsheet.

Student ID	Class	Section	Name	Percentage
10187	12	D	Rohit Rawat	72%
10013	12	B	Ashish Gupta	85%
10143	12	C	Vishal Garg	65%
11919	11	E	Chandan Bhatia	89%
10578	12	C	Ruchi Sharma	91%
10143	12	C	Vishal Garg	65%
11518	11	B	Deepak Vashisht	81%
11213	11	A	Deepti Verma	95%
10311	10	C	Vashali Gurung	93%
11095	11	A	Misha Malhotra	83%

Spreadsheet – Good structure
Data is stored in a sheet with the details of each individual stored according to a set of rules.

Rohit Rawat a student with ID 10187 of Class 12 Section D has scored 72%.

Text document – Poor structure
Data is stored in a text document with no set of organising rules.

2. **Cleanliness:** Clean data should not have duplicates, missing values, outliers, and other anomalies so that its reliability and usefulness for analysis are not affected. In the given example, cleaning of data removes the duplicate values.

Student ID	Class	Section	Name	Percentage
10187	12	D	Rohit Rawat	72%
10013	12	B	Ashish Gupta	85%
10143	12	C	Vishal Garg	65%
11919	11	E	Chandan Bhatia	89%
10578	12	C	Ruchi Sharma	91%
10143	12	C	Vishal Garg	65%
11518	11	B	Deepak Vashisht	81%
11213	11	A	Deepti Verma	95%
10311	10	C	Vashali Gurung	93%
11095	11	A	Misha Malhotra	83%

Student ID	Class	Section	Name	Percentage
10311	10	C	Vashali Gurung	93%
11213	11	A	Deepti Verma	95%
11095	11	A	Misha Malhotra	83%
11518	11	B	Deepak Vashisht	81%
11919	11	E	Chandan Bhatia	89%
10013	12	B	Ashish Gupta	85%
10578	12	C	Ruchi Sharma	91%
10143	12	C	Vishal Garg	65%
10187	12	D	Rohit Rawat	72%

3. **Accuracy:** Accuracy is same as reliability so it indicates how well the data matches real-world values. Accurate data closely reflects actual values without errors, enhancing the quality and trustworthiness of the dataset. When your measurement is accurate, it makes your data really good. It's like having a gold star on your homework—it shows you did a great job!

In the example given below, we are comparing data gathered for measuring the weight of 12 eggs in a box in grams.

Weight of 12 Eggs in a Box (gms)	Average
1116	
840	
1080	720
1200	
720	
Inaccurate Data	**Average**

Weight of 12 Eggs in a Box (gms)	Average
600	
840	
660	720
780	
720	
Accurate Data	**Average**

Ai Task

Creativity & Innovativeness

Open a website https://www.kaggle.com. Kaggle is like a playground for data enthusiasts! It's an online platform where people from all over the world come together to play with data, learn new things, and compete in data science competitions.

Do this

The Titanic competition on Kaggle is a classic and beginner-friendly challenge that introduces you to the basics of data analysis and machine learning. The goal is to predict whether a passenger survived the Titanic shipwreck based on factors like age, gender, ticket class, and more.

Explore:

Kaggle provides tutorials and notebooks to help you get started with the Titanic competition. You can find them under the "Notebooks" tab on the competition page.

Features of Data

Data features are also called the characteristics or properties of the data. They describe each piece of information in a dataset. They define what each data point represents and help us make sense of the data. For example,

- In a table of student records, features could include things like the student's name, age, or grade.
- In a photo dataset, features might include properties like the colour present in each image, the resolution, brightness, or the presence of certain objects.

These features help us understand and analyse the data. In AI models, we need two types of features: Independent and Dependent.

Independent

Independent variables (sometimes called predictor variables) are those that are used to generate predictions about or to account for the variation in the dependent variable (the goal). These features are the input to the model—they're the information we provide to make predictions.

Dependent

The dependent variable is the variable about which predictions or explanations are being sought. These features are the outputs or results of the model—they're what we're trying to predict. For example, imagine we're building an AI model to predict students' final exam grades based on various factors.

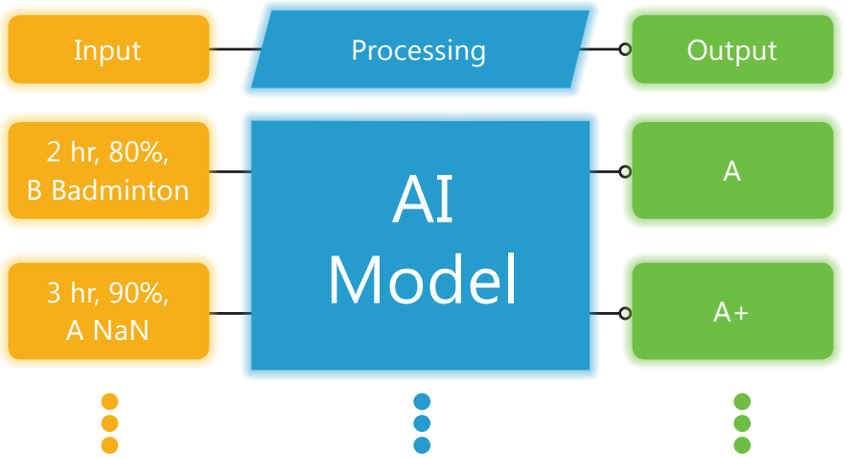
The independent features would include:

- **Study hours:** The number of hours a student spends studying.
- **Attendance:** Whether the student attended classes regularly or not.
- **Previous grades:** The grades the student received in previous exams.
- **Extracurricular activities:** Participation in extracurricular activities, such as sports or clubs.

The dependent feature, in this case, would be:

- The final exam grade—the outcome or prediction that the model gives us.

Together, they help us understand and improve student outcomes using AI-driven predictions.



- The independent variable is the cause. Its value is independent of other variables in your study.
- The dependent variable is the effect. Its value depends on changes in the independent variable.

Data Preprocessing

Data preprocessing is an essential phase in the machine learning process that prepares datasets for effective machine learning applications. It is the process of detecting and correcting (or removing) corrupt or inaccurate records from a dataset. It includes multiple processes to clean, transform, reduce, integrate, and normalise data.

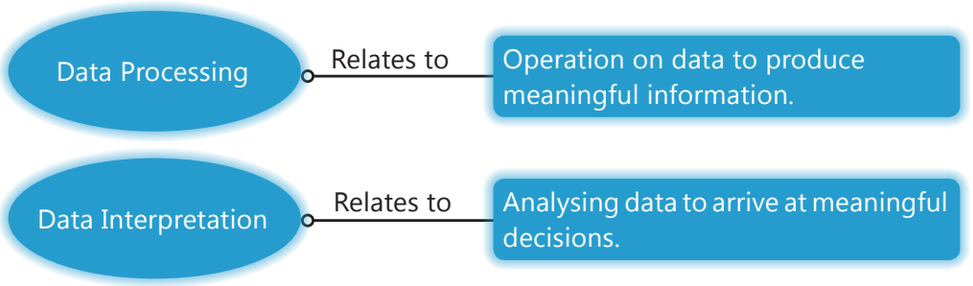
Data Processing and Data Interpretation

Data processing means preparing and analysing raw information to train models or predict outcomes, including tasks like cleaning and training. Data interpretation in AI involves analysing model outputs to understand patterns, refine models, and make informed decisions.

Observe and answer the following:

- How many large lollipops are there in the given picture?
- If each large lollipop represents 5 units of sweetness, how much total sweetness do the three lollipops represent?
- Among the small round candies, which colour appears most frequently?
- What is the ratio of green round candies to blue round candies?

- How many pink round candies are there in the image?



Data Processing

Data processing involves tasks to refine raw data for analysis or application, including cleaning, organising, transforming, and summarising information. It ensures data accuracy, relevance, and accessibility for effective decision-making and analysis. It is crucial across various sectors like business, science, and technology, facilitating better utilisation of data assets. Data processing helps computers understand raw data. Use of computers to perform different operations on data is included under data processing.

Data Interpretation

Data interpretation is the process of making sense of data by analysing it to uncover patterns, trends, and insights. It involves examining the data to understand its meaning, implications, and significance, helping to inform decision-making and draw conclusions. It is the process of making sense out of data that has been processed. The interpretation of data helps us answer critical questions.

Process of Data Interpretation

The steps in the process of data interpretation are as follows:

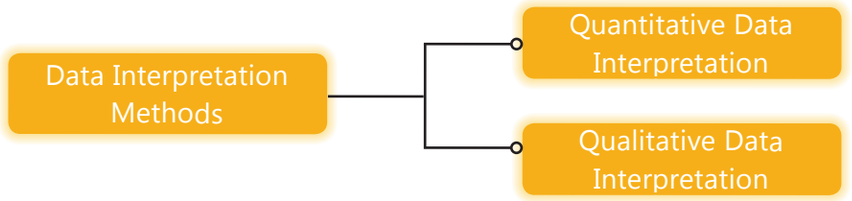
- **Acquire:** This initial step involves gathering raw data from diverse sources such as surveys, databases, or sensors. It ensures that all relevant information is collected to provide a comprehensive dataset for analysis.
- **Process:** Once the data is collected, it undergoes cleaning and organisation to remove errors, inconsistencies, or irrelevant information. This step ensures that the data is in a standardised format and ready for further analysis.
- **Analyse:** In this phase, the cleaned and organised data is scrutinised to identify patterns, correlations, or trends. Statistical methods, algorithms, or data visualisation techniques may be employed to extract meaningful insights from the data.
- **Interpret:** After analysing the data, the results are interpreted to derive actionable insights or conclusions. This involves understanding the implications of the analysis findings in the context of the problem or question at hand.
- **Present:** The final step involves presenting the interpreted findings in a clear and engaging manner. This could include visualisations such as tables, graphs or charts, along with concise summaries, to effectively communicate the insights derived from the data analysis.

These steps make sure that working with data is organised, complete, and useful, so that organisations can make smart choices based on the data.

Methods of Data Interpretation

Data interpretation is the process of making sense out of a collection of data that has been processed. This collection may be present in various forms like bar graphs, line charts and tabular forms and other similar forms.

There are two ways to interpret data-



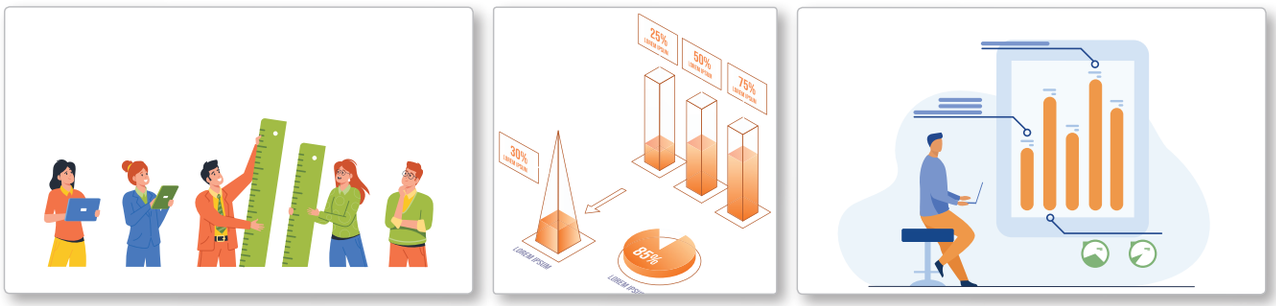
Quantitative Data Interpretation

It is the process of analysing and understanding quantitative or numeric data. This type of data often comes from surveys, experiments, and numerical measurements. Quantitative data provides statistical insights and helps in identifying patterns and trends. It requires statistical methods and techniques like mean, median, standard deviation, etc. to interpret the data. The interpretation of quantitative data focuses on measurable outcomes and numerical relationships. It helps us answer questions like "when," "how many," and "how often".

For example: (how many) numbers of likes on the Instagram post.

Data Collection Methods in Quantitative Data Interpretation

Data collection methods in quantitative data interpretation involve systematic techniques like surveys and experiments to gather numerical data. These approaches ensure data accuracy, facilitating reliable analysis and inference across various fields such as social sciences and healthcare.



Following are some application areas of quantitative data interpretation:

- **Interviews:** Quantitative interviews play a key role in collecting information.
- **Polls:** A poll is a type of survey that asks simple questions to respondents. Polls are usually limited to one question.
- **Observations:** Quantitative data can be collected through observations in a particular time period.
- **Longitudinal studies:** A type of study conducted over a long time.
- **Survey:** Surveys can be conducted for a large number of people to collect quantitative data.

Steps to Quantitative Data Analysis

The four steps involved in quantitative data analysis are:

1. **Relate measurement scales with variables:** Understand the type of data you have and match it with the appropriate measurement scale. For example, if you are looking at student grades, use ordinal (A, B, C) or ratio (numeric scores) scales.
2. **Connect descriptive statistics with data:** Use statistical measures to summarise and describe your data.

   For example, if you have test scores for a class of students, calculate the mean score to know the average performance, the median to understand the central tendency, and the standard deviation to see how varied the scores are.
3. **Decide a measurement scale:** Choose the appropriate scale to measure your data based on the type of variable and the level of detail required. For example, if you measuring weight, use a ratio scale because weight can be zero and can be measured precisely.
4. **Represent data in an appropriate format:** Display your data in a way that makes it easy to understand and interpret. For example, if you want to show the distribution of test scores in a class, you might use a histogram. If you want to show the relationship between study hours and test scores, a scatter plot would be appropriate.

Qualitative Data Interpretation

It is the process of analysing and understanding non-numeric data. This type of data is unstructured and often comes from interviews, surveys, observations, or textual content. Qualitative data tells us about the emotions and feelings of people. Qualitative data interpretation is focused on insights and motivations of people.

Data Collection Methods in Qualitative Data Interpretation

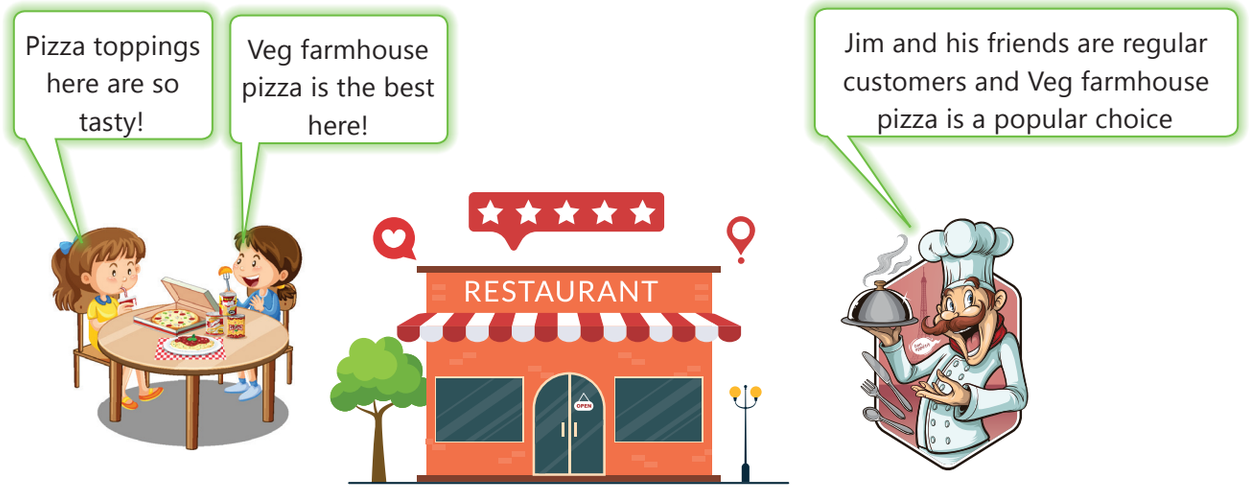
Data collection methods in qualitative data interpretation involve techniques such as interviews and observations to gather rich, descriptive data for nuanced analysis, fostering a deeper understanding of complex human experiences and behaviours. Some methods are as follows:

- **Record keeping:** This method utilises documents that are reliable and well curated and other similar sources of information as the data source that are verified and maintained. It is similar to going to a library.
- **Observation:** In this method, data is collected by observing the participants, their behaviour and emotions, carefully,
- **Case studies:** In this method, data is collected from case studies.
- **Focus groups:** In this method, data is collected after a group discussion on topics of relevance.
- **Longitudinal studies:** In this data collection method, data is collected on the same data source repeatedly over an extended period of time.
- **One-to-one interviews**: In this method, data is collected using a one-to-one interview.

Open Ended Surveys and Questionnaires

They allow organisations to collect views and opinions from respondents without meeting in person.

Reviews by customers – Qualitative data



AI Task

Problem Solving & Logical Reasoning

Let's do a small activity based on identifying trends.

- Visit the link: https://trends.google.com/trends/?geo=IN (Google Trends)
- Explore the website
- Check what is trending in the year 2024 – Global
  - ✶ Make a list of trending sports (top 5)
  - ✶ Make a list of trending movies (top 5)
- Check what is trending globally in the year 2024

List of trending athletes (top 5)	List of trending movies (top 5)

Steps to Qualitative Data Analysis

The five steps involved in qualitative data analysis are:

1. **Collect data:** Gather qualitative data through various methods to understand people's experiences, opinions, or behaviours. This is done through interviews, surveys, observations, or documents. For example, a researcher interviews patients about their experiences with a new healthcare app, recording their responses for further analysis.
2. **Organise and connect the qualitative data:** Prepare and arrange the collected data in a systematic way to make it easier to work with. For example, the researcher transcribes the recorded interviews into text documents and organises them by participant or interview date.

3. **Set a code to the data collected:** Assign labels or codes to different parts of the data to identify themes, patterns, or categories. For example, the researcher reads through the interview transcripts and highlights sections discussing "ease of use," "technical issues," and "benefits of the app," tagging them with corresponding codes.
4. **Analyse your data for insights:** Examine the coded data to identify deeper patterns, relationships, and insights. For example, The researcher groups codes related to "ease of use" and "technical issues" into a broader theme of "user experience" and analyses how these themes impact overall user satisfaction with the app.
5. **Reporting on insights derived from analysis:** Present the findings clearly, using quotes and visual aids to support your conclusions and recommendations. For example, the researcher writes a report highlighting the main themes along with positive and negative feedback.

Difference Between Qualitative and Quantitative Data Interpretation

Qualitative Data Interpretation	Quantitative Data Interpretation
Categorical	Numerical
Provides insights into feelings and emotions	Provides insights into quantity
Answers how and why	Answers when, how many, or how often
Methods – Interviews, Focus Groups	Methods – Assessment, Tests, Polls, Surveys
Example question – Why do students like attending online classes?	Example question – How many students like attending online classes?

Ai Task

Coding & Computational Thinking

Word Puzzle

Instructions:

- Partner with a person to play the game.
- There will be three rounds of Word puzzle.
- After 3 rounds, answer the questions given on the next slide.

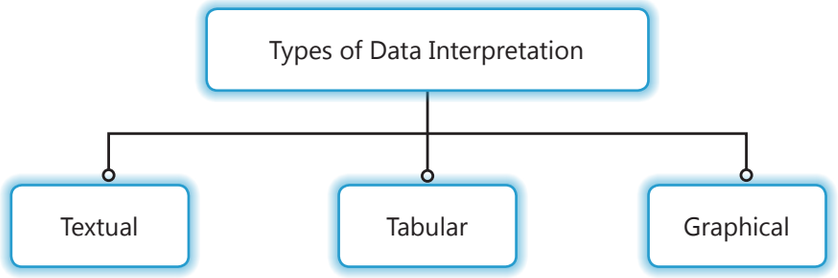
Now answer the following questions:

- Who won round one?
- Who won round two?
- Who won round three?

If you answered any of the above questions, you collected data!

Types of Data Interpretation

There are three ways in which data can be presented:



Textual DI

Data is put into words, like in a paragraph, which works well for small amounts of data that can be easily understood. But for larger amounts, this type of presentation may not be the best because it can get too complicated. For instance, a paragraph might describe how a company's sales went up in the first quarter, and how many units of each product they sold, as well as improvements in customer satisfaction.

Tabular DI

Data is organised systematically in rows and columns within a table, facilitating structured representation. In the example given below, the title of the table, "Students Marks Analysis," provides a descriptive overview of the table's content, summarising the analysis of student marks within the table.

Table Students' Marks Analysis

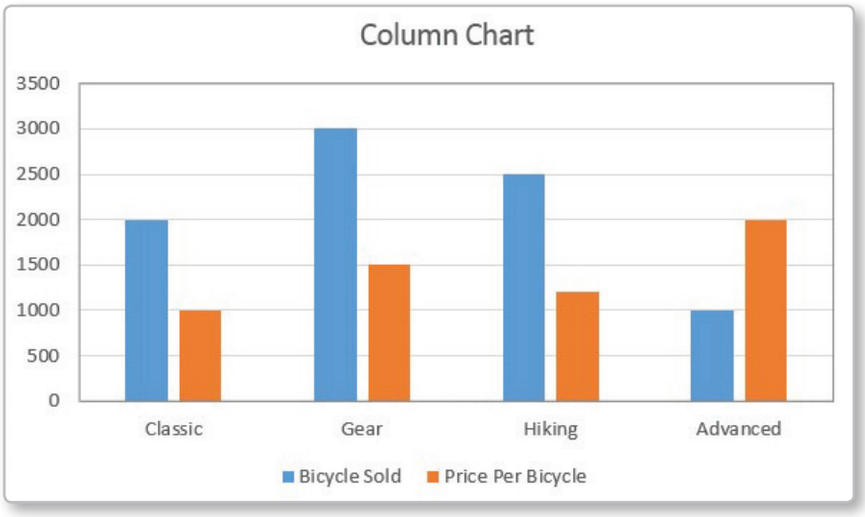
Student Name	English	Maths	Science	Percentage
Rehan	85	90	88	87.67%
Hemant	78	82	80	80.00%
Deepak	92	88	91	90.33%
Lucky	74	76	78	76.00%
Ramesh	88	84	85	85.67%
Saima	90	89	92	90.33%
Abdul	80	78	79	79.00%
Seema	85	87	86	86.00%

Graphical DI

Some of the graphs include bar graphs, line graphs, pie charts, and scatter plots, which help in visualising trends, relationships, and distributions within the data.

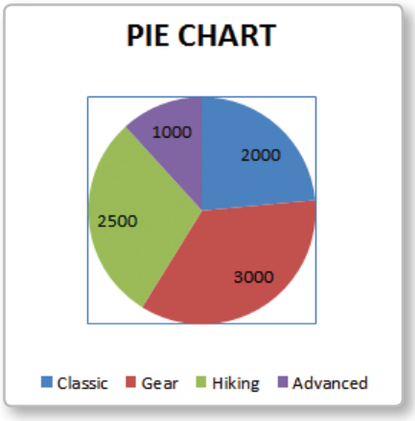
Bar Graphs

In a Bar Graph, data is represented using vertical and horizontal bars.



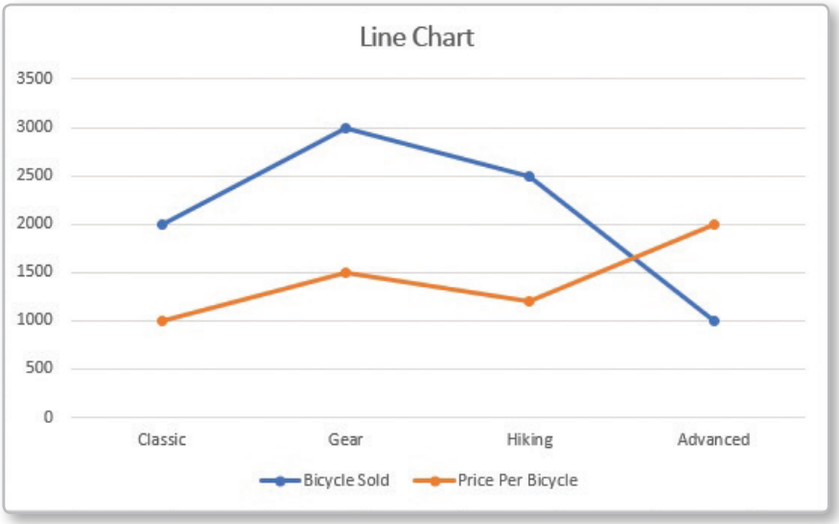
Pie Charts

Pie charts resemble pies, with each slice representing a portion of the whole pie assigned to different categories. These circular charts are divided into sections, and the size of each section corresponds proportionally to its value within the dataset.



Line Graphs

A line graph connects data points to illustrate changes in quantity over time, aiding in visualising trends and patterns.



AI Reboot

Quiz Time: AI Quiz

Session Preparation

Logistics: For a class of 40 Students [Pair Activity]

Materials Required:

ITEMS	QUANTITY
COMPUTERS	20

Brief:

The following are questions for the quiz. You can either go for a Pen/Paper Quiz or you can visit any open-sourced, free, online portal; one of which is Kahoot, and create your quiz there. For Kahoot: Go to https://kahoot.com/ and create your login ID. Then, add your own kahoot in it simply by adding all the given questions into it. Once created, you can initiate the quiz from your ID and students can participate in it by putting in the Game pin.

Quiz Questions

1. What are the basic building blocks of qualitative data?
   a. Individuals b. Units c. Categories d. Measurements
2. Which among these is not a type of data interpretation?
   a. Textual b. Tabular c. Graphical d. Raw data
3. Quantitative data is numerical in nature.
   a. True b. False
4. A bar graph is an example of ..... data interpretation.
   a. Textual b. Tabular c. Graphical d. None of the above
5. ..... relates to the manipulation of data to produce meaningful insights.
   a. Data processing b. Data interpretation
   c. Data analysis d. Data presentation

Importance of Data Interpretation

Data interpretation is crucial as it transforms raw data into actionable insights, guiding informed decision-making. By analysing and understanding data, organisations can uncover trends, patterns, and relationships, enabling them to optimise strategies, mitigate risks, and drive growth.

Informed Decision Making

A decision is only as good as the knowledge it is based on. It means when we analyse data, we get a clearer picture of what's going on. This helps us make decisions that are more likely to lead to success.

For example, if the average height of students is known, the school can custom design the chairs and tables according to the requirements of the class.

Reduced Cost

Identifying needs can lead to a reduction in cost. It means by knowing what's necessary, we can cut down on waste. We can use resources more efficiently and not spend money on things that aren't important.

For example, restaurant owner could decide to drop/modify some dishes of the menu which aren't popular or have got bad reviews.

Identifying Needs

We can identify the needs of people by data interpretation. It means understanding what people want or require by looking at the information we have.

For example, in a Pizza Shop, there are possibilities that Veg Farmhouse Pizza is a popular choice among the age group 8-10.

Using Tableau for Data Presentation

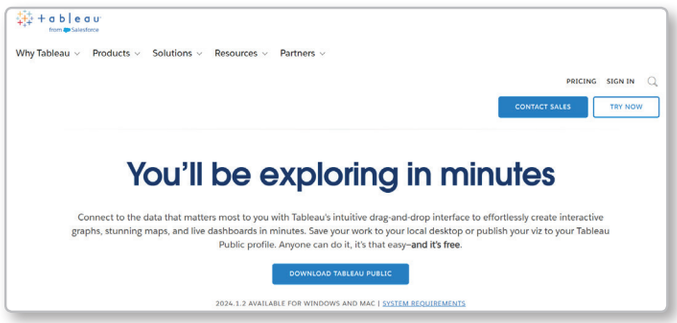
Using Tableau for data presentation involves connecting to various data sources, creating diverse visualisations, and enabling interactive features. It supports sharing and collaboration, offers advanced analytics capabilities, and promotes best practices for clear and effective data communication.

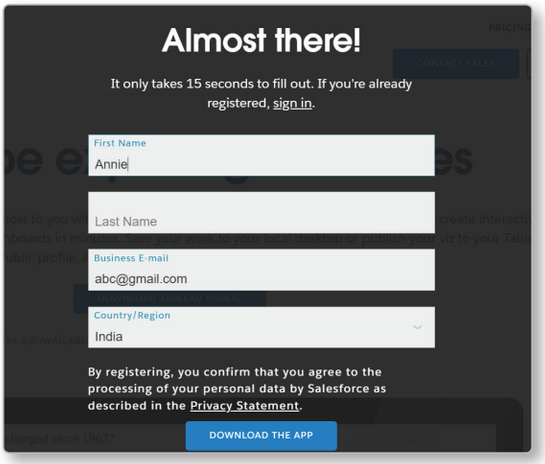
What is Tableau?

Tableau is a powerful data visualisation and business intelligence tool for visualising and analysing data in order to aid in business choices. It takes in data and produces various charts, graphs, maps, dashboards, and stories.

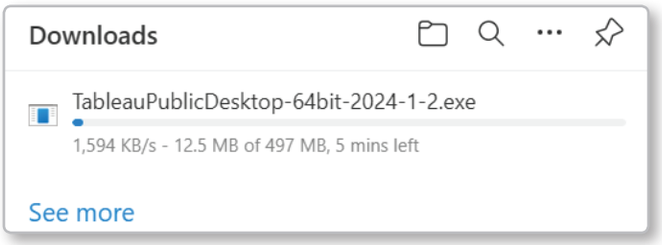
Steps to Download Tableau

1. Visit the link https://public.tableau.com/en-us/s/download
2. Click on the **DOWNLOAD TABLEAU PUBLIC BUTTON.** It will display the given screen where you enter your details.



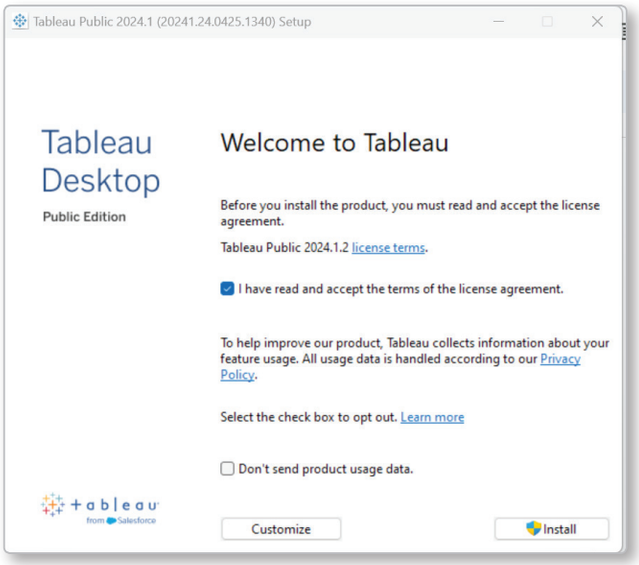


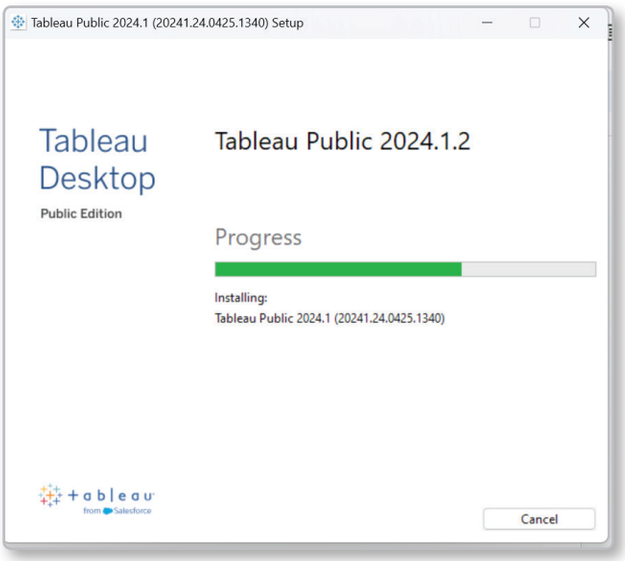
3. Click on the **DOWNLOAD THE APP** button to begin with the download process as shown below:



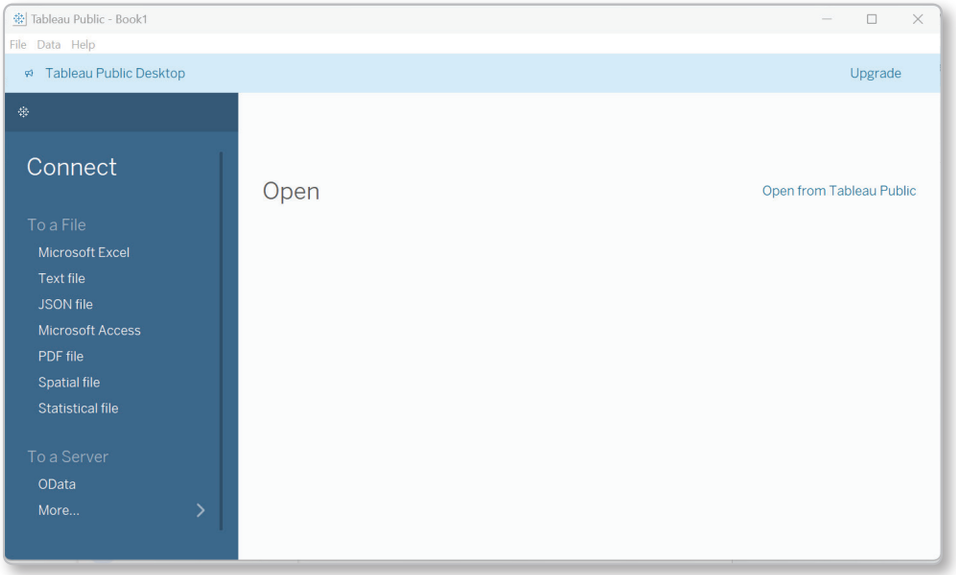
4. After finishing with the downloading of the files, double-click the installer of the **Tableau Public Desktop**. The **Tableau Public 2024.1 Setup** wizard opens.

5. Select the **I have read and accept the terms of the license agreement** check box to accept the terms of the license agreement.
6. Click on the **Install** button. Process of installation begins as shown below:





As soon as the installation process is over the application is ready for use.



You can also open the Tableau application by double-clicking the shortcut icon of the Tableau application on the Desktop.

Creating a Bar Graph Using Tableau

The steps to draw a Bar Chart in Tableau are as follows:

1. Create an Excel file and save it as **student.xlsx** with the following data:

	A	B	C	D	E
1	**Student**	**Subject**	**Score**		
2	Aarav	Math	85		
3	Aarav	Science	90		
4	Aarav	English	88		
5	Priya	Math	78		
6	Priya	Science	82		
7	Priya	English	80		
8	Rahul	Math	92		
9	Rahul	Science	85		
10	Rahul	English	87		
11	Sanya	Math	88		
12	Sanya	Science	91		
13	Sanya	English	89		
14					

2. Double-click on the **Tableau app** shortcut icon on the **Desktop**.

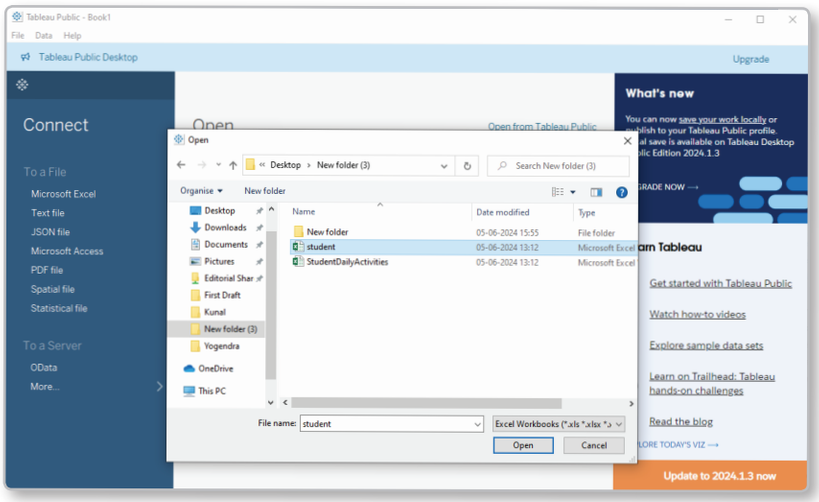
   The **Tableau** app opens.

3. Select the **Microsoft Excel** option from the **Connect** pane to access the Excel data that is used for visualising the representation in Tableau. The **Opens** dialog box appears.

4. Navigate the location where the Excel file is stored.

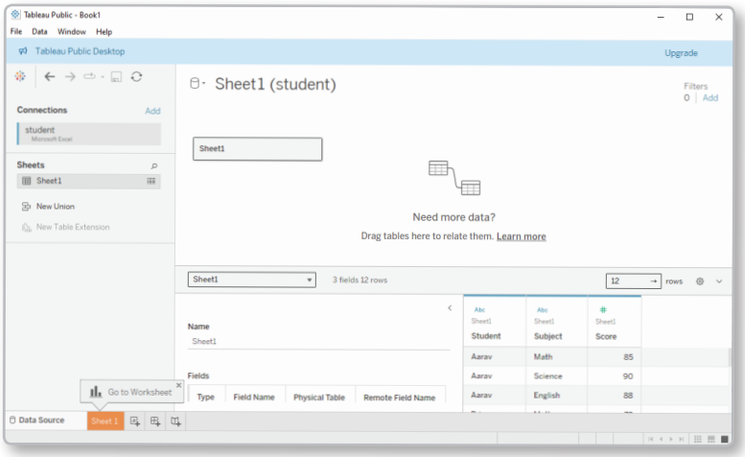
5. Select the **student.xlsx** file

6. Click on the **Open** button.

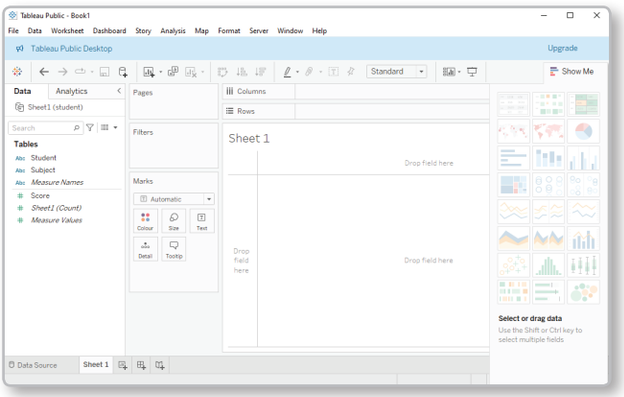


   The data of the Excel file is displayed in the **Data Source** window.

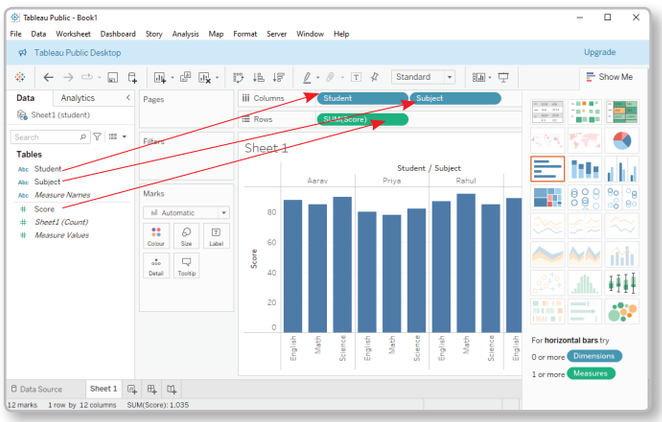
7. Click on **Sheet 1** in the **Sheet** tab.



The data of the **Sheet 1** of **student.xlsx**. In the Data pane, you will see the fields **Student**, **Subject**, and **Score**.



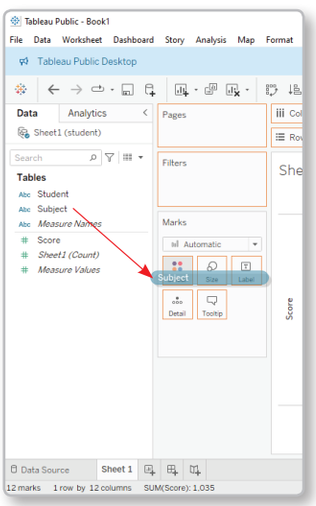
8. Drag the **Student** field to the **Columns** shelf.
9. Drag the **Score** field to the **Rows** shelf.
10. Drag the **Subject** field to the **Columns** shelf. Tableau generates a bar graph, as shown below:



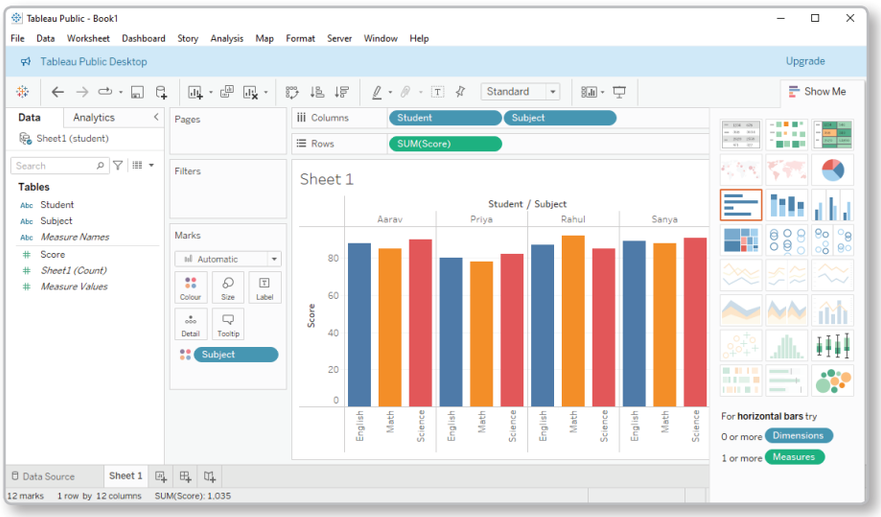
You can sort the bars in graph in ascending or descending order by clicking the **Ascending** or **Descending** option in the toolbar.

Changing Colour of the Graph

To make a colourful bar graph, drag the **Subject** field from the **Data** pane and place over the **Color** option in the **Marks** card.



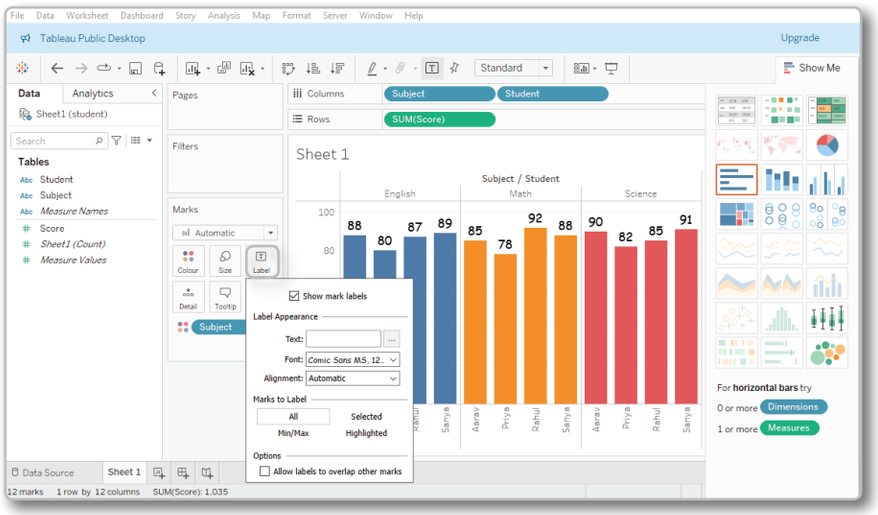
A colourful bar graph is generated, as shown below:



Changing Label of the Graph

The steps to change the label of the graph are as follows:

1. Click on the **Label** option in the **Marks** card. This opens up a box that allows us to change the font type, size, colour, etc.
2. Select the **Show mark labels** check box to show the labels on the graph.
3. Change the font type and size according to your requirement. In this case, we have changed the font size to **12** and the font to **Comic Sans MS**. The labels are displayed in the graph with the specified font type and size.

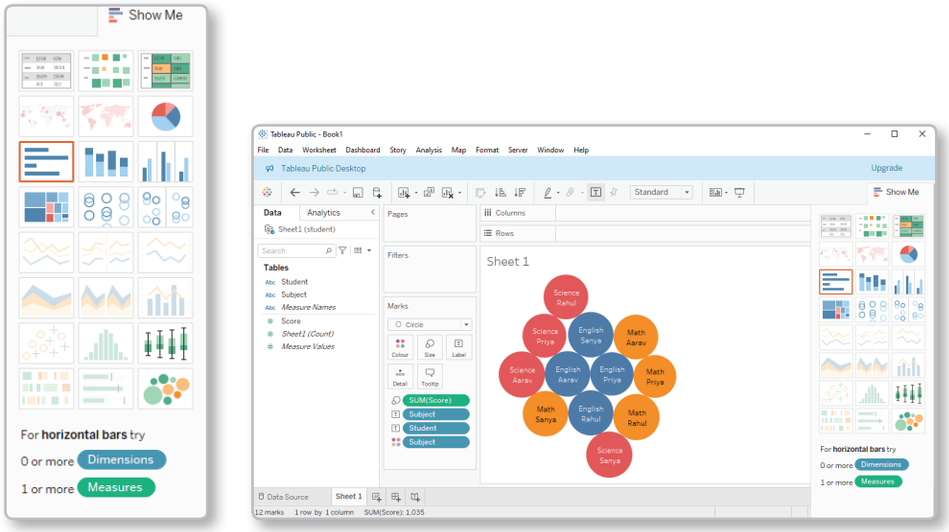


Change the Graph Type

Tableau often auto-selects the graph type based on the data. If the default graph type does not suit your data, you can change it accordingly. Perform the following steps to change the graph type:

1. Click on the **Show Me** button. The **Show Me** panel appears.

2. Select the desired graph type from the **Show Me** panel. In this case, we have selected the **Packed Bubbles** chart type. The type of the graph will be changed to **Packed Bubbles** chart type.



Duplicating a Chart

The steps to duplicate a chart are as follows:

1. Right-click the sheet in the **Sheet tab** whose chart you want to duplicate.
2. Select the **Duplicate** option from the context menu. A duplicate sheet is added in the Sheet tab.

Save and Share a Workbook

Once you have created a graph, you can save your workbook. The steps to save your Tableau Public workbook are as follows:

1. Click on the **File** → **Save** as option from the menu bar. The **Save as** dialog box opens.
2. Navigate the location where you want to save you workbook.
3. Type the name of the workbook in the **File name** text box.
4. Click on the **Save** button. The tableau public workbook is saved with the specified name.

You can export the tableau workbook as package by selecting the **Export Packaged Workbook** options in the **File** menu.

You can also share your tableau workbook by saving it to **Tableau Public** or **Tableau Server** if you have access to those services. For this, you need to select the **Save to Tableau Public** option in the **File** menu.

Experiential Learning

Video Session

Scan the QR code or visit the following link to understand about tableau:

https://www.youtube.com/watch?v=NLCzpPRCc7U

How Ben solved his business related issue using Tableau?

AI Task

Your favourite songs

- Think about songs! Which songs do you like to listen to? Which songs do you love to sing?
- Do you have a favourite song, artist, album, or playlist?
- Let's start thinking about the different aspects of a song, like instruments played and lyrics.
- Do your favourite songs have anything in common?
- Maybe your favourite music falls within the same genre.
  - ✶ A genre refers to the different styles of music.
  - ✶ Common genres include hip-hop, pop, alternative, and rock.
  - ✶ Classifying songs by genre, and other traits allows us to see trends in our favourite music.
  - ✶ All of this information is valuable data that we can count, summarise, and present!

Instructions

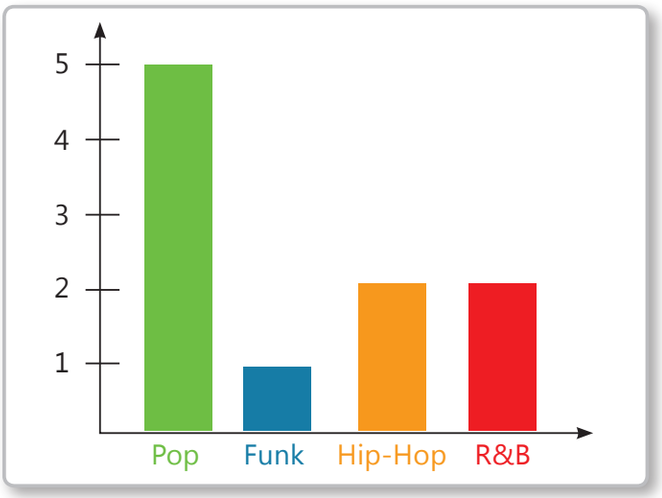
- Draw a grid with 6 columns as shown.
- Title the first column Song Name, then write down the names of 5-10 of your favourite songs.
- For this activity, we're going to collect data about the Album, Artist, Genre, Year, and Song Length.
- Add the headings to your table.
- Fill out the table by looking up each song on Google, Spotify, or Apple Music.

Song Name	Album	Artist	Genre	Year	Song Length
Blinding Lights					
Savage love					
Watermelon Sugar					
Happy					
Panini					
Cake by the Ocean					
7 Rings					
24K Magic					
Put Your Records On					
Since U Been Gone					

Let's visualise

- Count the number of songs that fall into each genre.
- Make a bar chart to visualise the number of songs within each genre using your counting. Colour each bar a different colour.
- You will get a graph as shown in the image.

- Using data visualisation, can you tell which genre has the most songs?



At a Glance

- Data refers to any collection of raw facts, figures, statistics, or information that can be stored and processed by a computer. It can be in different forms like numbers, text, images, audio, and video etc.
- Literacy refers to the ability to read, comprehend and use information effectively.
- To integrate data literacy skills into the organisational culture it is important to make data-driven decision-making a fundamental part of everyday work.
- Designing an evaluation metric for the data literacy program involves creating a structured framework to assess participants' progress and the effectiveness of the program overall.
- Data privacy referred to as information privacy is concerned with the proper handling of sensitive data including personal data and other confidential data.
- Data security is the practice of protecting digital information from unauthorised access, corruption, or theft throughout its entire lifecycle.
- Strong password is a combination of atleast 8 characters with upper and lower-case letters, numbers, and special characters that are difficult for unauthorised individuals or automated programs to guess or crack.
- Encryption is a security technique that transforms readable data (plaintext) into an unreadable format (ciphertext) using an algorithm and an encryption key.
- Data disposal refers to the process of securely destroying or deleting data that is no longer needed to prevent unauthorised access, recovery, and misuse.
- Using Firewall and Antivirus software can stop and alert users of any suspicious activity happening on their devices.
- Cyber security involves protecting computers, servers, mobile devices, electronic systems, networks, and data from harmful attacks.
- Data augmentation is the process of increasing the amount and diversity of data. We do not collect new data, rather we transform the already present data.
- Data scraping is the process of collecting data from websites using software.
- Data processing refers to the manipulation and transformation of data into useful information through various techniques and methods.
- Data interpretation involves examining the data, identifying patterns, trends, and relationships, and translating the findings into actionable information or decisions.
- Data analysis is to examine each component of the data in order to draw conclusions.
- Tableau is a powerful data visualisation and business intelligence tool for visualising and analysing data in order to aid in business choices.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Which of the following is not an example of data?
   a. Audio ◯ b. Video ◯
   c. Text ◯ d. Hardware ◯

2. The ..... illustrates the progressive transformation of raw data into actionable wisdom.
   a. Data ◯ b. Data literacy ◯
   c. Data pyramid ◯ d. Information ◯

3. The data pyramid begins with ..... .
   a. Top level ◯ b. Raw data ◯
   c. Information ◯ d. Knowledge ◯

4. Data literacy enhances ..... ability in individuals based on evidence.
   a. Programming ◯ b. Understanding ◯
   c. Decision-making ◯ d. Information ◯

5. Designing an ..... metric for the data literacy program involves creating a structured framework.
   a. Mathematical ◯ b. Logical ◯
   c. Skill ◯ d. Evaluation ◯

6. ..... is about hunting for valuable information in different places, checking if it's good quality, and making sense.
   a. Data discovery ◯ b. Data investigation ◯
   c. Data quality ◯ d. Data literacy ◯

7. ..... can recognise objects, people, and even actions happening in videos.
   a. NLP ◯ b. Data science ◯
   c. Data ◯ d. Computer vision ◯

8. Which of the following is not the method of data collection in qualitative data interpretation?
   a. Record keeping ◯ b. Observation ◯
   c. Case studies ◯ d. Driving ◯

9. ..... should be organised in a way that makes sense so that it can be used effectively.
   a. Data ◯ b. Knowledge ◯
   c. Privacy ◯ d. Ethics ◯

10. ..... should not have duplicates, missing values, outliers, and other anomalies so that its reliability and usefulness for analysis are not affected.
    a. Text data ◯ b. Clean data ◯
    c. Visual data ◯ d. Ethical ◯

B. Fill in the blanks.

1. ..... refers to any collection of raw facts, figures, statistics, or information that can be stored and processed by a computer.
2. ..... is a person who can interact with data to understand the world around them.
3. ..... can equip individuals with skills and knowledge to improvise in a data-driven world.
4. The ..... provides a comprehensive and structured approach to develop the necessary skills for using data efficiently with all levels of awareness.
5. ..... means increasing the amount of data by adding copies of existing data with small changes.
6. The data generated from the experiment is an example of ..... .
7. Tableau is a powerful data visualisation and ..... tool.
8. In ..... data is represented systematically in the form of rows and columns.
9. Quantitative data interpretation is made on ..... data.
10. In ..... data collection method, data is collected on the same data source repeatedly over an extended period of time.

C. State whether these statements are true or false.

1. Data literacy framework is an iterative process. .....
2. Data analysis is used to examine each component of the data in order to draw conclusions. .....
3. Qualitative data tells us about the numbers of the data. .....
4. Data from people's experience is a data collection method in quantitative data. .....
5. In a pie graph, data is represented using vertical and horizontal bars. .....
6. Data interpretation helps in making informed decisions by providing a clearer picture of the situation. .....
7. Descriptive statistics are not used in quantitative data analysis. .....
8. Reporting is one of the steps involved in qualitative data analysis. .....

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. From the given image answer the following questions:

A	B	C	D	E	F

In the given image typical levels of awareness in a data literacy process framework are shown.

Identify the labels marked as A, B, C, D, E, F.

Ans. A-Plan
B-Communicate
C-Assess
D-Develop Culture
E-Prescriptive Learning
F-Evaluate

2. What is prescriptive learning?

Ans. By implementing a prescriptive learning approach, organisations can provide a set of diverse resources that align with individual learning styles.

3. What is data acquisition? Give an example.

Ans. Data acquisition, also known as acquiring data, refers to the procedure of gathering data like raw facts, figures or statistics from relevant sources either for reference or for analysis as needed in AI projects. For example: collection of data to predict the topper of the school in the upcoming board examination.

4. Name the sources of data acquisition.

Ans. Sources for acquiring data can be primary or secondary.

5. Why is it important for data to be clean?

Ans. Clean data is free from duplicates, missing values, and anomalies, ensuring its reliability and usefulness for analysis.

6. What role does Tableau play in data presentation?

Ans. Tableau is a powerful tool for visualising and analysing data, aiding in business decisions by creating various charts, graphs, maps, and dashboards.

7. Why is it important to choose the appropriate measurement scale for data analysis?

Ans. Choosing the correct scale ensures accurate measurement and representation of data, leading to valid analysis results.

8. Describe a scenario where identifying needs through data interpretation can lead to reduced costs.

Ans. A restaurant owner could use customer feedback data to modify or remove unpopular dishes from the menu, reducing waste and costs.

9. What are some common features of bar graph and line graph?

Ans. Both graph types represent data visually, with bar graphs using bars and line graphs using points connected by lines to show changes over time.

10. Give an example of a situation where record keeping would be a useful data collection method.

Ans. Using library documents as reliable and curated sources of information for data collection.

B. Long answer type questions:

1. Give the impacts of data literacy in education and business?

Ans. Data literacy has an immense impact on various aspects of society like business, education, healthcare, and public policy, some of them are given here:

- **Business:** It improves the decision making skills of a person. Data-literate employees can effectively analyse data to gain insights into market trends, customer behaviours, and operational performance.
- **Education:** It empowers the teaching-learning process. Students can engage more deeply with course material, particularly in STEM fields.

2. List down the ways that will help you to become data literate.

Ans. Here is a guide to help you become data literate:

- **Understand the Basics:** Start from learning the concepts of data, types of data and how it can be used.
- **Learn Data Analysis Tools:** There are many data analysis apps available that can be learned in order to understand the impact of right data.
- **Gain Statistical Knowledge:** Statistics play a vital role in data literacy. Its one of the vital components that must be learned before you dive into the data driven world.
- **Use Data Visualisation:** Understand the techniques of data visualisation such as Graphics and Charts. Tools like Tableau, matplotlib, python can be used effectively for this purpose.

- **Learn Data Manipulation:** Understanding how to manipulate data to meet the requirements is also one of the key factors. Methods like filtering, sorting, grouping and omitting are essential for extracting insights from large data set.
- **Practise Cleaning:** Learning to remove data redundancy and data inaccuracy is essential to be data literate.

3. What are the steps involved in Data Acquisition?

Ans. The three key steps involved in Data Acquisition are as follows:

i. **Data discovery:** It is about hunting for valuable information in different places, checking if it's good quality, and making sense of what we find.

ii. **Data augmentation:** It is the process of increasing the amount and diversity of data. We do not collect new data, rather we transform the already present data. Data augmentation means increasing the amount of data by adding copies of existing data with small changes.

iii. **Data generation:** It refers to generating or recording data using sensors. Recording temperature readings of a building is an example of data generation. Recorded data is stored in a computer in a suitable form.

4. Explain the importance of having a clear structure in data and provide examples of good and poor data structure.

Ans. Clear structure in data ensures it is organised logically, facilitating efficient analysis and interpretation. For example, marks of students arranged in a spreadsheet is a good structure, whereas a poor structure example if the student records were stored in a disorganised manner, with inconsistent naming conventions or missing attributes, it would impede data analysis and decision-making processes.

5. Why is data privacy important?

Ans. It is important because:

- A data breach at a government agency can put top secret information in the hands of an enemy state.
- A breach at a hospital can put personal health information in the hands of those who might misuse it.
- A breach at a corporation can put proprietary data in the hands of a competitor.
- A breach at a school can cause inconvenience to the parents, such as constant calling from tuition and coaching centers that leads to disturbance.

6. Explain the term computer vision and the type of data used in this ?

Ans. Computer Vision is like giving eyes to computers. It helps them look at pictures and videos from the real world and understand what they're seeing. With Computer Vision, computers can figure out what's in a picture or video, just like we do with our eyes. They can recognise objects, people, and even actions happening in videos.

Types of data used in computer vision include:

- **Image Data:** Digital images captured by cameras or satellite imagery, and medical scans.
- **Video Data:** Video data captured using camera, and surveillance footage.

Unsolved Questions

SECTION A (Objective Type Questions)

AiQuiz

A. Tick (✓) the correct option.

1. Data literacy is able to cultivate ..... skills to understand and explore data's implications by questioning assumptions.

a. critical thinking ○ b. programming ○

c. awareness ○ d. probability ○

2. Data literacy fuels ..... by providing tools and techniques to explore data from different perspectives.
   a. errors ◯ b. comprehension ◯
   c. innovation ◯ d. repetition ◯
3. ..... enables users to tackle complex problems and derive meaningful relevance.
   a. Mathematics ◯ b. Trends ◯
   c. Project cycle ◯ d. Data literacy ◯
4. Data literacy has an impact on which of the following?
   a. Public policy ◯ b. Cooking ◯
   c. Driving ◯ d. Jogging ◯
5. By implementing a ..... learning approach, organisations can provide a set of diverse resources that align with individual learning styles.
   a. modern ◯ b. prescriptive ◯
   c. planned ◯ d. latest ◯
6. The process of collecting data from websites using software is called ..... .
   a. Data analysis ◯ b. Data reference ◯
   c. Data literacy ◯ d. Data scraping ◯
7. ..... is the process of increasing the amount and diversity of data.
   a. Data augmentation ◯ b. Data filtering ◯
   c. Data processing ◯ d. Data modelling ◯
8. Digital images captured by cameras or satellite imagery, medical scans, and surveillance footage is ..... .
   a. Text data ◯ b. Numeric data ◯
   c. Computer vision ◯ d. Audio data ◯
9. ..... is the process of making sense out of a collection of data that has been processed.
   a. Data interpretation ◯ b. Data scrapping ◯
   c. Data validation ◯ d. Data handling ◯
10. The ..... is the first step of data handling.
   a. Plan ◯ b. Acquire ◯
   c. Present ◯ d. Process ◯

B. Fill in the blanks.

1. ..... refers to the ability to read, comprehend and use information effectively.
2. ..... is essential because it enables individuals to make informed decisions, think critically, solve problems, and innovate.
3. ..... is a conceptual model that illustrates the hierarchical structure of data processing.
4. ..... data sources are the external sources for collecting data, rather than generating it personally.
5. ..... refers to generating or recording data using sensors.
6. ..... is all about teaching computers to understand and work with human language.
7. ..... are also called the characteristics or properties of the data.
8. Application involves putting acquired knowledge and ..... into practice in real-world contexts.

9. ..... is a combination of upper and lower-case letters, numbers, and special characters that are difficult for unauthorised individuals or automated programs to guess or crack.
10. ..... refers to the process of creating copies of data to ensure that it can be restored in the event of data loss due to natural disasters, accidents, cyber-attacks, or other unexpected events.

C. State whether these statements are true or false.

1. Data analysis is a common method for extracting information from websites. .....
2. Recording temperature readings of a building is an example of data generation. .....
3. Data augmentation is the process of increasing the amount and diversity of data. .....
4. Secondary data includes data taken from surveys, interviews, experiments, etc. .....
5. Good data has information scattered. .....
6. Data analysis is a common method for extracting information from websites .....
7. Clean data should not have duplicates. .....
8. Accurate data closely reflects actual values with errors. .....
9. Data features are also called the characteristics of the data. .....
10. Data interpretation involves collecting the data. .....

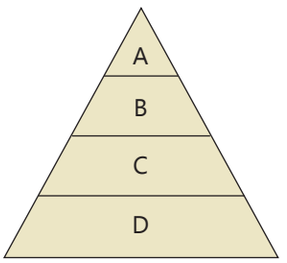
D. Match the following:

1. Computer Vision	a. Data scraping
2. NLP	b. Image Data
3. Textual Data	c. Dataset search
4. Sources of data	d. Data history
5. Data Discovery	e. Audio Data
	f. Qualitative Data

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. Image of a data pyramid is given here. Based on it, answer the following questions:



   i. What is a data pyramid?
   ii. Name the areas marked as A, B, C, and D in the data pyramid.
   iii. Give one important feature of each level.
2. Why is data literacy essential? List any two factors.
3. How can pie charts be useful in data presentation?
4. How does informed decision-making benefit from data interpretation?
5. What is the difference between qualitative and quantitative data?
6. What is the purpose of data processing?
7. What is Kaggle and how can it be useful for data enthusiasts?
8. Define the term "Data Literacy Process Framework."
9. Why is the data literacy framework an iterative process?
10. Give any two best practices that can help you ensure data privacy.

B. Long answer type questions:

1. What are the ethical concerns while doing data acquisition?
2. Why is data security important?
3. Define the term "Data Backup".
4. How is data security related to AI?
5. List any three best practices of cyber security.
6. Explain with example the two types of numeric data.
7. Explain in short the types of data used in three domains of AI.

C. Competency-based/Application-based questions:

Problem Solving & Logical Reasoning

1. Your teacher has asked students to give the choice of at least 3 co-curricular activities from the given list:

   a. Painting
   b. Music - Western
   c. Music - Indian
   d. Dance - Western
   e. Dance - Indian
   f. Best out of waste
   g. English Theatre
   h. Hindi Theatre

   You're provided with a dataset containing errors, duplicates, and missing values. How would you approach organising and cleaning this data to ensure its reliability and usefulness for analysis?

   Outline the steps you would take to organise and clean the dataset, ensuring that it is free from errors, duplicates, and missing values. Additionally, describe any methods or techniques you would use to address these issues and ensure the dataset's reliability and usefulness for analysis.

2. The following dataset represents the students' academic performance, identify which features in the dataset would be considered independent variables and which would be dependent variables in predicting students' final exam grade.

Stud Name	Study Hours	Attendance	Previous Grades	Extracurricular Activity	Final Exam Grade
Aarav	5	Yes	B	Sports	A
Riya	7	Yes	A	Debate Club	B
Arjun	6	No	C	Music Club	C
Ishan	4	Yes	B	Drama Club	B
Anaya	8	Yes	A	None	A

3. Examine the following datasets:

   Quantitative Data:

   Number of petals on a flower:
   Rose: 32
   Lily: 24
   Sunflower: 89
   Tulip: 16

   Height of flowers (in centimeters):
   Rose: 45
   Lily: 55
   Sunflower: 120
   Tulip: 30

   Qualitative Data:

   Colour of flowers:
   Rose: Red
   Lily: White
   Sunflower: Yellow
   Tulip: Pink

   Fragrance intensity:
   Rose: Strong
   Lily: Mild
   Sunflower: None
   Tulip: Moderate

These datasets contain both quantitative data (number of petals and height) and qualitative data (colour and fragrance intensity) for different types of flowers (Rose, Lily, Sunflower, Tulip).

a. Discuss the differences between quantitative and qualitative data interpretation.

b. Describe the methods and techniques commonly used for interpreting quantitative and qualitative data, highlighting their respective strengths and limitations.

4. You are tasked with analysing the performance of a company's sales across different regions over the past year. How would you utilise data visualisation techniques to present this information effectively to the company's stakeholders during a quarterly review meeting? Describe the types of visualisations you would use and explain how they would help convey the sales trends and patterns to the audience.

Ai In Life

1. Prepare a questionnaire using Padlet.com, to know how data literacy is helpful in education.
2. Make a presentation to depict the "Data Literacy Process Framework".

Ai Deep Thinking

1. In today's digital age, data has become an incredibly valuable resource, much like gold was during the gold rush era. In the context of Artificial Intelligence (AI), data is the raw material that fuels AI systems, How?
2. Who first said that Data is a new gold? Is data more precious than the gold? Justify

Ai Lab

Experiential Learning

Ask students to collect data of different coloured objects in the Lab and record it in a spreadsheet. Create a basic bar chart to visualise the collected data using spreadsheet software. Later, ask students to present their bar charts, followed by a brief discussion on the importance of data quality and ethical considerations in AI.

Answers

AI Quiz Section A (Objective Type Questions)

A. 1. d 2. c 3. b 4. c 5. d 6. a 7. d 8. d 9. a 10. b

B. 1. Data 2. Data literate individual 3. Data literacy
4. Data literacy framework 5. Data augmentation 6. Primary data 7. Business intelligence
8. Tabular DI 9. Numeric 10. Longitudinal studies

C. 1. True 2. True 3. False 4. False
5. False 6. True 7. False 8. True

Ai Ready 2

Answer the following questions:

1. Why do you think there's a need to educate students about data literacy?

2. Why is data privacy important? Give an example.

3. What are the 3 C's of data literacy?

4. What is meant by cyber attack? Give an example of it.

UNIT-3

MATHS FOR AI (STATISTICS & PROBABILITY)

Learning Outcomes

- How are Maths and AI Related?
- Statistics
- What is Probability in Statistics?
- Essential Mathematics for AI
- Application of Statistics
- Applications of Probability

Mathematics is crucial to artificial intelligence (AI) because it provides the theoretical foundation and practical tools for developing, assessing, and optimising AI systems. It is the backbone of AI algorithms and models, empowering machines to process, analyse, and interpret vast amounts of data.

In general, mathematics guarantees that AI models are precise, effective, and able to resolve challenging issues.

How are Math and AI related?

Mathematics and AI are interconnected fields, with mathematics supplying the theoretical foundations for many AI algorithms. Patterns are repeating designs or sequences that can be observed in numbers, shapes, images, languages, or objects in our surroundings. They follow a specific order or arrangement, making them easily recognisable. Mathematics aids in the study of these patterns. These patterns allow you to solve puzzles. They help identify an order or arrangement in lists of images or numbers. They are present everywhere around us.

Patterns in Numbers

Number patterns, often simply referred to as patterns, are sets of numbers that follow a specific order or rule. There are various types of number patterns, including Fibonacci, geometric, algebraic, and arithmetic patterns.

Some tricks are given below for finding the patterns in numbers:

- **Number Sequences:** Look for a sequence where each number is related to the one before it in a specific way. This could be:
  - ✶ **Adding** a constant value (e.g., 1, 4, 7, 10... +3 each time)
  - ✶ **Multiplying** by a constant value (e.g., 2, 4, 8, 16... x2 each time)
  - ✶ **Following a formula** (e.g., 2n + 1: 3, 5, 7, 9...)
- **Series:** Analyse a series of numbers to identify the underlying rule for example square series, cube series, etc.
- **Even or Odd:** Is the sequence made of even numbers (2, 4, 6...) or odd numbers (1, 3, 5...)?
- **Prime or Composite:** Are the numbers prime (only divisible by 1 and itself) or composite (divisible by more than two numbers)?
- **Fibonacci Sequence:** This famous sequence starts with 0 and 1, and each subsequent number is the sum of the two preceding numbers (0, 1, 1, 2, 3, 5, 8...).

Patterns in Images

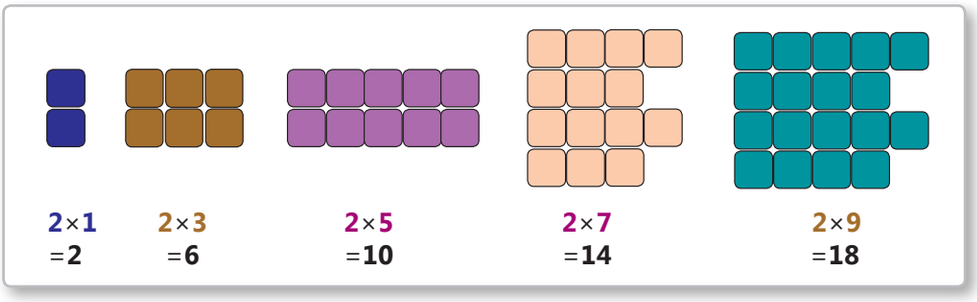
Patterns in images refer to recurring visual elements or structures that can be recognised and analysed within the context of an image. These patterns can be detected and utilised for various purposes, including image recognition, classification, compression, and enhancement.

- **Repeating shapes:** Look for shapes that appear multiple times in the image.
  Are they of the same size and colour?
  Do they form a larger shape?
- **Symmetry:** Does the image have a mirror effect where one side reflects the other?
  Is it symmetrical vertically, horizontally, or both?
- **Colour patterns:** Are there repeating colours or a specific colour scheme used?
  Do the colours alternate or follow a gradient?
- **Progression:** Do the elements in the image change in size, colour, or position in a predictable way?

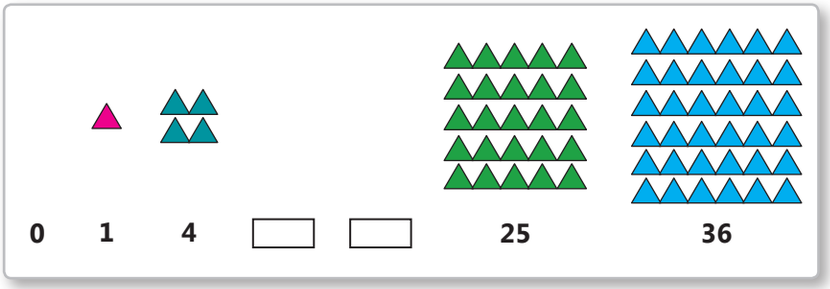
Ai Task

Problem Solving & Logical Reasoning

1. Identify a pattern in the given image?



2. Find the missing numbers in the given series and draw the pattern.



3. Which skill is used to identify the patterns in above questions?

Artificial Intelligence helps computers to understand and recognise patterns, much like how humans do. Just as we learn to recognise patterns in different things we see or hear, AI can do the same with various types of data, whether it's numbers, images, or even speech and text. These patterns help AI to solve puzzles – like identifying dogs and muffins, or predicting floods, earthquakes, etc.

Ai Task

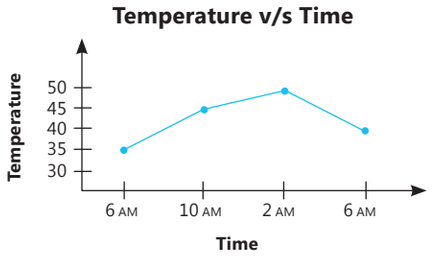
Problem Solving & Logical Reasoning

Let's think and answer a few more!

1. Find out the middle value from the given numbers?
   2, 5, 8, 11, 14

2. Fill in the missing number and find the next number in the given series?
   3, 6, 12, 24, ___, 96, ___

3. Identify the highest temperature in the given graph and mention the time of it.



4. How many faces are there in a dice?

5. How many sides does a coin have?

6. What is the shape of a ball?

Essential Mathematics for AI

Math will help us to better understand AI and its way of working, but what kind of math is needed for AI?
The kind of math needed for AI includes:

1. **Probability and statistics (exploring data):** Probability theory and statistics are key fundamentals for many AI algorithms, particularly those involving machine learning. It is useful in tasks such as natural language processing, computer vision, and decision-making.
2. **Linear algebra (finding out unknown or missing values):** Linear algebra is involved in large-scale data processing playing a vital role in machine learning and AI. It performs operations in neural networks, image processing, and data transformations.
3. **Calculus (training and improving AI model):** Calculus is essential for understanding the best possible solution algorithms used in machine learning. It minimise mistakes and maximise the parameters of machine learning models.
4. **Graph theory:** Graph theory is used in AI representing trends using data visualisation.

5. **Information theory:** Information theory provides mathematical tools for data analysis.
6. **Logic and set theory:** Logic and set theory are used in concepts like expert systems, database systems, and knowledge graphs. Mathematics and AI are deeply integrated fields, with mathematics providing the theoretical foundation for many AI algorithms and techniques.
7. **Algorithm design:** The design of the algorithms often uses mathematical principles and structures such as functions, matrices and graphs.

Statistics

Statistics is used for collecting, exploring, and analysing the data. It also helps in concluding data. It enables AI systems to detect patterns, identify relationships, and infer conclusions from data.

- Data is collected from various sources.
- Data is explored and cleaned to be used.
- The analysis of data is done to understand it better.
- Conclusions and decisions can be made from the data.

Let's consider an example to illustrate these steps:

A school wants to improve the performance of its students and decides to collect data on study habits and grades.

- **Collecting Data:** The school conducts a survey where students report the number of hours they study each week and their grades in various subjects.

- **Exploring and Cleaning Data:** The school first looks at the data to find patterns, like the range of study hours and grades.

  They also clean the data by fixing any missing or incorrect information (e.g., if some students didn't fill in all the fields or gave unrealistic answers).

- **Analysing Data:** The school summarises the average study hours and grades to get an overview. They also check if there is a significant relationship between study hours and grades.

- **Drawing Conclusions:** The analysis might reveal that students who study more hours tend to have higher grades. Based on this conclusion, the school might decide to implement study support programs to encourage students to study more.

Thus, we can say statistics helps in transforming the raw data into meaningful insights, enabling better decisions and strategies in various fields such as business, healthcare, education, and more.

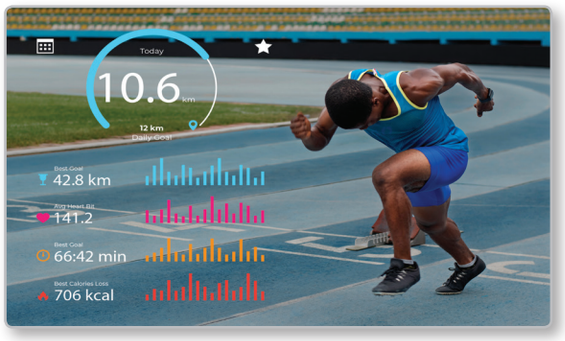
Brainy Fact

During the COVID-19 pandemic, statistical models were used to predict the spread of the virus and allocate medical resources efficiently. The World Health Organization (WHO) reported in 2020 that these models helped reduce the strain on healthcare systems by optimising resource distribution.

Application of Statistics

Statistics can be used in numerous fields because of its ability to extract meaningful insights from data and assist in decision-making. Let us discuss some of the fields where statistics is used.

Sports



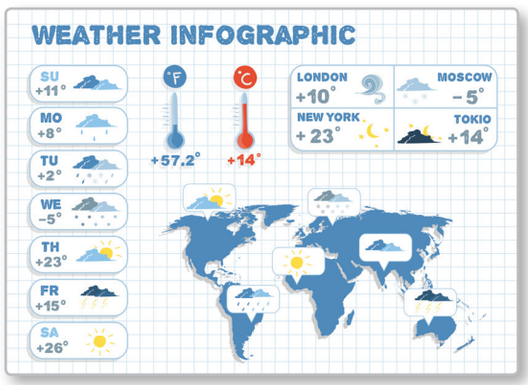
- Statistics helps you to assess how well players perform, how teams plan their games, and the results of matches.
- Statistics gives data that helps you make smart decisions about hiring players, training them, performance monitoring, preventing injury, and planning game strategies.
- Studying opponents' strengths and weaknesses to devise better game plans.
- The Tokyo 2020 Olympics were postponed due to the developing global situation of the Covid-19 pandemic. Statistics revealed that COVID cases sharply increased in Japan during the planned period of the Olympics.

Education

- Analysing test scores and grades to evaluate student learning, identify areas for improvement and allocate resources effectively.
- Using data to identify gaps in the curriculum and areas where students need more support.
- Analysing how students and teachers use educational technology for future implementations.
- Statistics helps in determining the average skills of students in a particular school or grade. This information shows which areas need more focus and help improve education strategies.

Disaster Management

- Authorities use statistics in analysing existing risks and alert the citizens residing in places that might be affected by a natural disaster in the near future.
- The disaster management teams use statistics to know about the population, and about the services and infrastructure present in the affected area.
- Statistics are used to study trends in disaster occurrence and impact to adapt strategies over time.
- Statistics are used to design and evaluate public education campaigns to raise awareness and preparedness levels.



Weather Forecast

- Statistical summaries create easy-to-understand weather forecasts for the public.
- Statistics compare the weather conditions with information about past seasons and conditions.
- Statistical methods are used to analyse long-term climate data and detect trends related to global warming and climate change.
- Statistics are used to predict the future based on data from the past and provide forecasts that express the likelihood of various weather events (e.g., "There is a 70% chance of rain tomorrow").

Disease Prediction

- Statistical models are used to predict the spread and impact of infectious diseases, such as influenza, COVID-19, and Ebola.
- Data analysis creates epidemic curves that show the progression of disease outbreaks over time.
- The US government uses statistics to understand which disease is affecting the population the most. This helps them in curing these diseases more effectively. To understand the extent of any of the infectious diseases in a given population.

For example, the government can analyse the areas where COVID cases are increasing, or where the vaccination drive needs to be improved.

Options of Voters

- Statistics are used in elections to enhance transparency for voters and encourage voters to participate in elections.
- By surveying a group of voters and analysing the results, statistics can reveal what the majority of people think about political candidates, policies, or upcoming elections. This helps politicians and decision-makers to understand public opinion.
- Statistics play an important role in Election Forecasts, making Campaign Strategies and micro-target individuals based on available data.

AI Reboot

Explain any two applications of statistics in education.

What is Probability in Statistics?

Probability is a branch of statistics that deals with the likelihood or chance of an event to happen or different outcomes occurring in a given situation.

- It measures the amount of certainty of an event.
- It helps us make predictions about future events based on the data we have.
- It helps us understand how likely something is to happen.

The formula for probability is given as the ratio of the Number of Favourable Outcomes to the Total Number of Possible Outcomes.

$$P(A) = \frac{\text{Number of Favourable Outcomes}}{\text{Total Number of Possible Outcomes}}$$

Let us understand it using the example of a coin.

When a coin is tossed, there are two possible outcomes: head or tail.

What is the probability of getting a head?

$$P(\text{Head}) = \frac{1}{2}$$

So, we see the chances of getting head is $\frac{1}{2}$ and the chances of getting tail is $\frac{1}{2}$.

Therefore, we conclude that when we toss a coin, there are equal chances of getting a head or a tail.

Let us take another example, consider the probability of rolling a fair six-sided die.

A fair six-sided die has six faces numbered from 1 to 6.

The probability of rolling any specific number (say, 3) on a fair six-sided die is calculated as,

Number of favourable outcomes (rolling a 3) = 1

Total number of possible outcomes = 6 (since there are 6 faces on the die)

So, the probability of rolling a 3 on a fair six-sided die is:

$$P(6) = \frac{\text{number of favourable outcomes}}{\text{total number of possible outcomes}} = \frac{1}{6}$$

Can you find the probability of getting even numbers in a Dice?

Ai Task

Digital Literacy

Let's understand probability with the following activity.

Purpose: To understand the possibility of occurrence of an event.

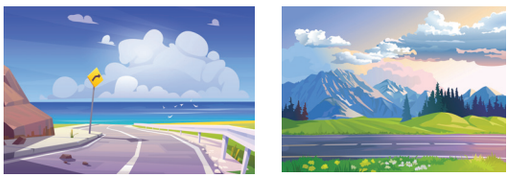
Case 1: Varun went to the park to play football. Which pet is he more likely to see, a cat or a dog?

(A) Cat (B) Dog

Case 2: It's Ankita's birthday; which sweet is she most likely to distribute toffee or chocolate to her friends in class?

(A) Chocolate (B) Toffee

Case 3: In the summer, Veer starts to drive and reaches a diversion, where one road takes him to Sunny Beach and the other takes him to the mountains. What is the probability that he would select the mountain road?



(A) Beach Road (B) Mountain Road

Ai Task

Experiential Learning

About the Game

A deck containing 52 cards is grouped into four suits of clubs, diamonds, hearts, and spades. Each of the clubs, diamonds, hearts, and spades have 13 cards each, which sum up to 52.

Game Structure

Find the probability of drawing a heart from a standard deck of 52 cards?

What is the probability of drawing a spade or diamond from a standard deck of 52 cards?

When discussing probability, we often rely on specific terms to describe the likelihood of events occurring. Let us discuss the specific terms in detail.

Certain Events

These events are guaranteed to happen; there is no doubt about their occurrence. It will have a probability of 1. For example:

- ✶ If you flip a fair coin, the probability of it landing heads up or tails up is certain, as one of these outcomes is guaranteed.
- ✶ The occurrence of sunrise and sunset each day is certain.
- ✶ When you flip a light switch, the light bulb will either turn on or off.
- ✶ The act of inhaling and exhaling is certain as long as a person is alive.
- ✶ Time consistently moves forward, and each passing moment is certain.
- ✶ The beating of the heart is a certain event, as long as a person is alive.
- ✶ If it is Sunday today, it is certain, tomorrow is going to be a Monday.

Likely Events

These events have a higher probability of occurring as compared to other events. For example:

- ✶ If you roll a fair six-sided die, the likelihood of rolling a number greater than 2 (3, 4, 5, or 6) is higher than rolling a number less than or equal to 2 (1, or 2).
- ✶ If you visit a store known for carrying a wide variety of brands, it's likely that you'll find your favourite brand among their products.
- ✶ If you study well for an exam, you're more likely to pass with good grades.
- ✶ If you buy many raffle tickets compared to others, you're more likely to win a prize.

Unlikely Events

These events have a lower probability of occurring compared to other events. For example,

- ✶ If you randomly select a card from a standard deck, the probability of drawing an ace of spades is lower compared to drawing a card of a different suit.
- ✶ Observing a shooting star in the night sky is an unlikely event compared to seeing regular stars.

- ✶ Getting into a really good university, especially one that's really hard to get into, is not very likely because there are so many people applying.
- ✶ If a board result is declared, the chances to be the topper of the school is very unlikely as there is only one topper in the school.

Impossible Events

These events have no chance of occurring; they are not feasible outcomes. The event will never happen or is impossible, it will have a probability of 0. For example,

- ✶ Rolling a fair six-sided die to get a number greater than 6 or less than 1.
- ✶ Drawing a card from a standard deck and it turns out to be both a red card and a black card at the same time.
- ✶ The probability that you can pick a red ball from a bag containing only blue balls is 0.
- ✶ Picking a random day of the week, and it turns out to be both Monday and Friday at the same time.
- ✶ Picking a guava in an orchard of apples.
- ✶ The probability of a fish to climb a tree is an impossible event.

Equal Probability Events

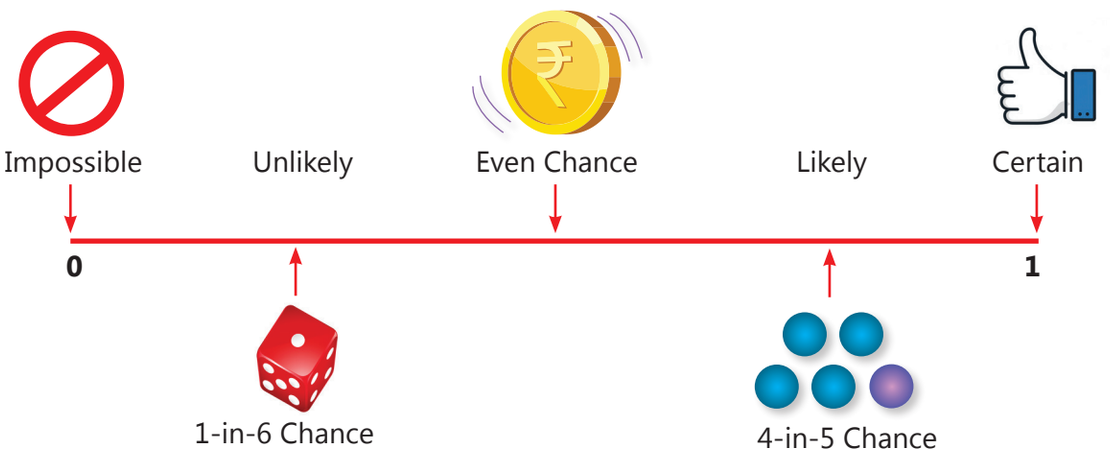
In this scenario, each event has an equal chance of occurring. For example,

- ✶ If you roll a fair six-sided die, the probability of rolling any specific number (1, 2, 3, 4, 5, or 6) is the same because each face of the die has an equal chance of landing face up.
- ✶ To guess a number between 1 and 10, each number has an equal probability of being selected.
- ✶ The chances of selecting a particular coloured marble from a bag of marbles containing the equal number of red, blue, and green marbles are equal.
- ✶ When students enter in a classroom and select their seats, each desk or chair has an equal chance of being chosen.
- ✶ Selecting a boy or a girl from a class having equal number of girls and number of boys has an equal probability.

The probability of an event occurring is somewhere between impossible and certain. If an event is certain or sure to happen, it will have a probability of 1.

For example, the probability that it will rain in the State of Florida at least once in a specific year is 1. If an event will never happen or is impossible, it will have a probability of 0.

So, the probability of an event occurring lies between 0 and 1.



Ai Task

Experiential Learning

It's Arshia's birthday and she has a bag full of 7 toffees and 3 chocolates. She asked you to pick one from the bag.

Try to fill in the blanks with – likely, unlikely, certainly, impossible, equal probability

1. If you pick from the bag without looking, it is _____ that you will pick toffee.
2. If you pick from the bag without looking, it is _____ that you will pick a chocolate.
3. If you pick from the bag without looking, it is _____ that you will pick a cake.
4. If you remove 4 toffees from the bag, and pick without looking, there is an _____ that you pick a chocolate or a toffee.
5. If you pick from the bag with all toffees removed and pick without looking, you will _____ pick a chocolate.

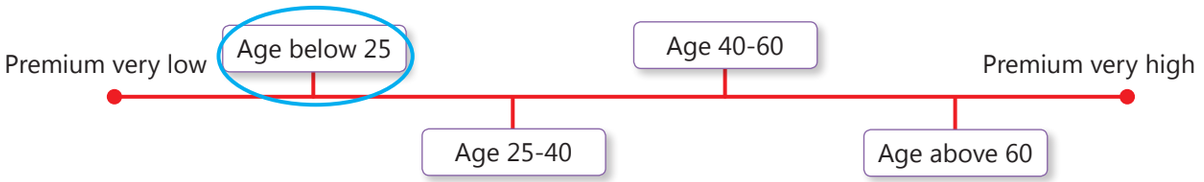
Case 1

Health insurance companies often use probability to determine how likely it is that certain individuals will spend a certain amount on healthcare each year and determine risk factors for contracting a disease and for being cured.

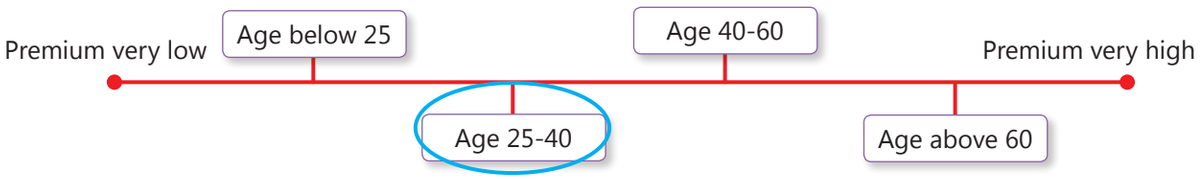
For example, a company might use factors like age, existing medical conditions, current health status, etc. to determine that there's a 90% probability that a certain individual will spend ₹10,000 or more on healthcare in a given year.

Individuals who are likely to spend more on healthcare will be charged higher premiums because the insurance company knows that they'll be more expensive to insure.

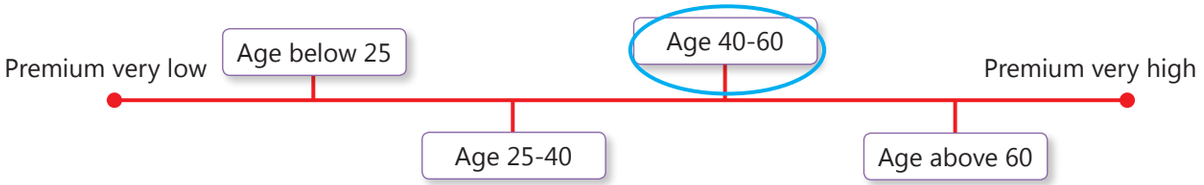
- **Scenario 1**: If the person is young, below 25 years of age, then the insurance premium will be very low.



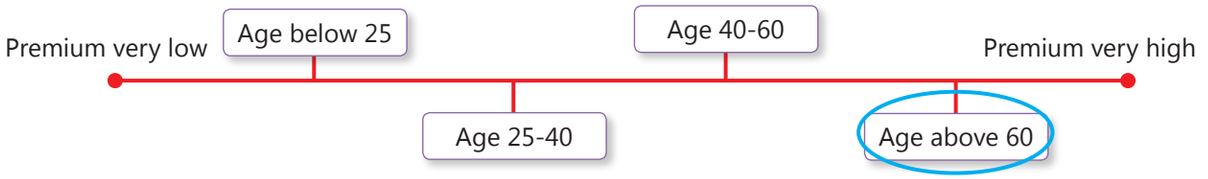
- **Scenario 2**: If the person is middle-aged (25–40), then the insurance premium will be of higher value than that in the first category.



- **Scenario 3**: If the age of the person is 40-60, then the premium charged will be quite high.



- **Scenario 4**: If the person is above 60 years of age then the premium charged is of the highest bracket.



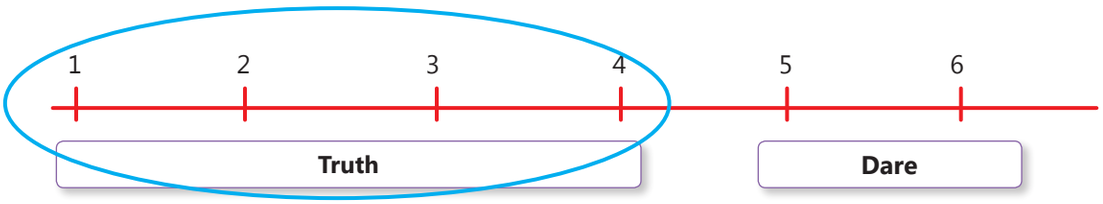
Brainy Fact

GPT-3 (Generative Pre-trained Transformer 3) is a large language model, released by OpenAI in 2020, has 175 billion parameters and uses probabilistic sampling techniques for text generation.

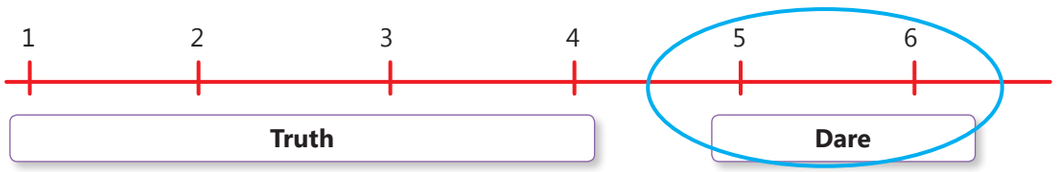
Case 2

Let there be a bet among friends that a person will have to perform a dare if the die rolls out 5 or 6 and otherwise the person has to speak the truth.

- **Scenario 1:** When die rolls into any number between 1-4. The person replies to a question with the truth.



- **Scenario 2:** When the dice rolls out 5 or 6 then a person gets to do a dare.



Applications of Probability

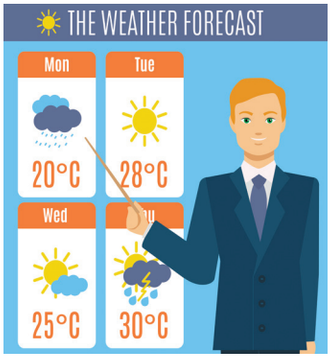
Probability has a wide range of applications across various fields, making it an essential concept in many areas of study and professional practice. Here are some key applications of probability.

Sports

- In cricket, the batting average represents how many runs a batsman would score before getting out. Probability can help in estimating the batting average. For instance, if a batsman had scored 20 runs out of 80 from only boundaries in the last match. Then, there is a chance that he will score 25% of his runs in the next match from boundaries.

- Probability helps in predicting the outcomes of sports events by analysing historical data and other relevant factors. For example, in football (soccer), the probability of a team winning can be calculated based on home advantage, current form, and historical head-to-head results.
- Coaches and analysts use probability to predict individual player performance. For example, in basketball, the probability of a player making a free throw can be estimated based on their past success rate.
- During games, coaches use probability models to make strategic decisions. For example, in football, when a team scores a touchdown, they have a choice. They can either kick the ball for an easy point or try for a more challenging two-point conversion by running or passing the ball into the end zone again. Coaches use probability to decide which option gives them the best chance of scoring more points.
- Teams with players suited to the ground conditions may have a higher probability of winning. For example, if the nature of the pitch favours batsmen or bowlers, can have a substantial impact.

Weather Forecasting



- Probability is used by weather forecasters to assess how likely it is that there will be rain, wind, snow, clouds, etc., on a given day in a certain area.

  For example, forecasters may say things like "there is a 70% chance of rain today between 4 PM and 6 PM" to indicate a medium to high likelihood of rain during certain hours.
- Probability helps in understanding long-term climate patterns and changes.

  For example, estimating the probability of extreme weather events under different climate scenarios aids in planning and mitigation efforts.

Traffic Estimation

- People often use probability when they decide to drive to someplace. Vehicular traffic flow is examined and an estimated vehicle waiting time in each direction is estimated through probability.
- Based on the time of day, location in the city, weather conditions, etc. people tend to make probability predictions about how bad traffic will be during a certain time. For example, if you think there's a 90% chance that traffic will be heavy from 6 PM to 7:30 PM in your vicinity then you may decide to wait during that time.

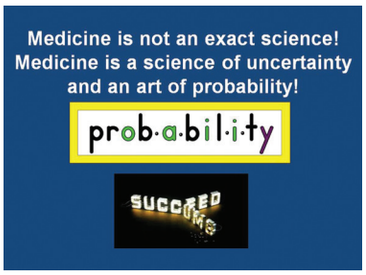
Finance

- Probability is used to assess the risk of investments and financial decisions.

  For example, calculating the probability of a stock's return falling within a certain range helps investors make informed decisions.
- Probability models help investors spread their investments across different assets to reduce risk.

  For example, by looking at the chances of returns for various investments, investors can build a balanced portfolio.
- Probability is crucial in figuring out insurance costs and how policies are structured by various insurance agencies.

  For example, actuaries use probability to predict events like accidents or natural disasters and set insurance premiums based on those chances.



Medicine

- Probability helps in understanding how diseases spread and how well treatments work.

  For example, by estimating the chance of disease transmission in different situations, we can better plan public health measures.
- Probability is used to determine the effectiveness of new treatments or drugs.

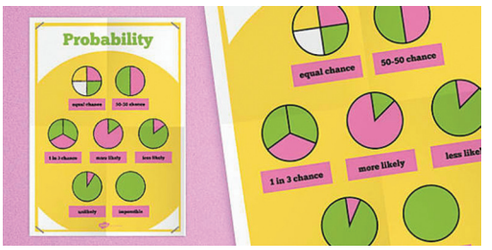
  For example, researchers calculate the probability that observed treatment effects are due to the drug rather than random chance.

Education

- Probability is used to estimate how well students will do in exams based on different factors.

  For example, considering study habits and attendance can predict the chances of students getting certain grades.
- Probability helps create fair and accurate standardised tests.

  For example, figuring out the chances of different scores in tests, it can better reflect how well students understand the subject taught.



At a Glance

- Mathematics is crucial to artificial intelligence (AI) because it provides the theoretical foundation and practical tools for developing, assessing, and optimising AI systems.
- Mathematics and AI are interrelated fields, where mathematics provides theoretical concepts to many of the AI algorithms.
- Artificial Intelligence assists computers to understand and recognise patterns, much like how humans do.
- Number patterns, often referred to as patterns, are sets of numbers that fit into a particular order.
- Patterns in images refer to recurring visual elements or structures that can be recognised and analysed within the context of an image.
- Probability theory and statistics are one of the key fundamentals to many AI algorithms, particularly those involving machine learning.
- Linear algebra is involved in large-scale data processing, playing a vital role in machine learning and AI.
- Calculus is essential for understanding the best possible solution algorithms used in machine learning.
- Statistics is used for collecting, exploring, and analysing the data. It also helps in concluding data.
- Statistics can be used in numerous fields because of its ability to extract meaningful insights from data and assist in decision-making.
- Probability is a branch of statistics that deals with the likelihood or chance of different outcomes occurring in a given situation.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. _____ are repeating designs or sequences that can be observed in numbers, shapes, images, languages, or objects in our surroundings.
   a. Patterns ◯ b. Iterations ◯
   c. Sequences ◯ d. Statistics ◯
2. Which of the following statements is not true?
   a. Mathematics helps in the study of patterns. ◯
   b. With the use of mathematics, you can solve puzzles. ◯
   c. Mathematics helps to identify an order/ arrangement in the list of images or numbers. ◯
   d. The patterns only exist in mathematics. ◯
3. _____ measures the amount of certainty of an event.
   a. Probability ◯ b. Calculus ◯
   c. Series ◯ d. Pattern ◯
4. _____ also helps in concluding data.
   a. Statistics ◯ b. Information ◯
   c. Probability ◯ d. Sequence ◯
5. The _____ of an event occurring is somewhere between impossible and certain.
   a. possibility ◯ b. existence ◯
   c. probability ◯ d. outcome ◯
6. If an event is certain or sure to happen, it will have a probability of _____ .
   a. 0 ◯ b. 1 ◯
   c. True ◯ d. None ◯
7. Picking a random day of the week, and it turns out to be both Monday and Friday at the same time. Identify events.
   a. Likely events ◯ b. Unlikely events ◯
   c. Impossible events ◯ d. Certain events ◯
8. If you toss a coin, the chance of getting a head or tail is an example of _____ .
   a. equal probability ◯ b. unlikely events ◯
   c. impossible events ◯ d. certain events ◯

B. Fill in the blanks.

1. _____ assist computers to understand and recognise patterns, much like how humans do.
2. _____ is used in AI representing trends using data visualisation.
3. _____ is used for collecting, exploring, and analysing the data.

4. _____ is a branch of statistics that deals with the likelihood or chance of different outcomes occurring in a given situation.
5. _____ events have a lower probability of occurring compared to other events.
6. Certain events are guaranteed to happen and have a probability _____ .
7. _____ helps in understanding long-term climate patterns and changes.
8. _____ the famous sequence starts with 0 and 1, and each subsequent number is the sum of the two preceding numbers.

C. State whether the following statement is true or false.

1. Graph theory is essential for understanding the best possible solution algorithms used in machine learning. _____
2. Data is collected only from a single source. _____
3. Conclusions and decisions can be made from the data. _____
4. Probability helps create fair and accurate standardised tests. _____
5. Mean is used to determine the effectiveness of new treatments or drugs. _____
6. AI can see patterns only in numeric data types. _____
7. Information theory provides mathematical tools for data analysis. _____
8. Statistical methods are used to analyse long-term climate data and detect trends related to global warming and climate change. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. What is probability in statistics?

Ans. Probability is a branch of statistics that deals with the likelihood or chance of different outcomes occurring in a given situation.

2. How do you calculate probability?

Ans. The formula for probability is given as follows:

$$P(A) = \frac{\text{Number of Favourable Outcomes}}{\text{Total Number of Possible Outcomes}}$$

3. Define the term "Statistics".

Ans. Statistics is used for collecting, exploring, and analysing the data. It also helps in drawing conclusions from data.

4. Explain the use of statistics in disease prediction.

Ans. Statistics models can predict the spread and impact of diseases, helping in planning public health measures.

5. What is an impossible event? Provide an example.

Ans. An impossible event has no chance of occurring, such as rolling a number greater than 6 on a fair six-sided die.

6. Why is calculus important for AI models?

Ans. Calculus is crucial for training and improving AI models by understanding the best possible solution algorithms.

7. What is an example of a certain event in probability?

Ans. The occurrence of sunrise and sunset each day is a certain event with a probability of 1.

8. Find the probability of getting two heads when five coins are tossed?

Ans. Number of outcomes of getting two heads = 5 × 2 = 10

Total number of Outcomes = 2 × 2 × 2 × 2 × 2 = 32

$$P(\text{two heads}) = \frac{10}{32} = \frac{5}{16}$$

B. Long answer type questions.

1. What are the possible ways to express probability?

Ans Probability can be expressed in the following ways:

- **Certain events:** An event will happen without a doubt.
- **Likely events:** The probability of one event is higher than the probability of another event.
- **Unlikely events:** One event is less likely to happen than another event.
- **Impossible events:** There's no chance of an event happening.
- **Equal Probability events:** Chances of each event happening is the same.

2. List down the areas of applications of statistics.

Ans. The following are the areas of applications of statistics:

- Sports
- Education
- Disaster management
- Weather forecast
- Disease prediction
- Opinion of voters

3. In the upcoming elections, the election commissioner wants to know whether it will be a hung parliament or a party will have a clear majority. Can this be achieved using statistics? How?

Ans. Yes, it can be achieved. By surveying a group of voters and analysing the results, statistics can reveal what the majority of people think about political candidates, policies, or upcoming elections. This helps politicians and decision-makers understand public opinion.

4. How is probability helpful for a student in studying a specific topic?

Ans.
- Students can use probability to estimate the likelihood of achieving certain grades based on their past performance and study habits.
- By analysing the probability of different outcomes, students can prioritise their study time more effectively.
- If a topic has a high probability of appearing on an exam, students can prioritise their study efforts accordingly.

5. How can probability help in the field of medicine?

Ans.
- Probability helps in understanding how diseases spread and how well treatments work. For example, by estimating the chance of disease transmission in different situations, we can better plan public health measures.
- Probability is used to determine the effectiveness of new treatments or drugs. For example, researchers calculate the probability that observed treatment effects are due to the drug rather than random chance.

C. Competency-based/Application-based questions:

1. Nikhil wants to understand customer satisfaction and collects data through feedback forms where customers rate their satisfaction on a scale from 1 to 10. What steps should Nikhil take to explore and clean the customer satisfaction data?

Ans. Nikhil should look first at the data to find patterns, like the range of study hours and grades. He also cleans the data by fixing any missing or incorrect information (e.g., if some students didn't fill in all the fields or gave unrealistic answers).

2. A school is organizing a raffle where there are 10 tickets, and only one ticket is a winner. Each student gets one ticket. What is the probability that a specific student wins the raffle?

Ans. Number of Favorable Outcomes = 1 (the winning ticket)

Total Number of Possible Outcomes = 10 (total tickets)

Probability P(A) $= \frac{1}{10} = 0.1$ or 10%.

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

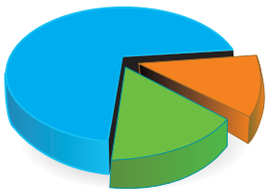
A. Tick (✓) the correct option.

1. _____ are used in concepts like expert systems, database systems and knowledge graphs.
   a. Math ◯ b. Logic and set theory ◯
   c. Graph theory ◯ d. Calculus ◯

2. In an _____ event, there's no chance of an event happening.
   a. unlikely ◯ b. certain ◯
   c. impossible ◯ d. likely ◯

3. If an event will never happen or is impossible, it will have a probability of _____ .
   a. 1 ◯ b. 0 ◯
   c. >1 ◯ d. <1 ◯

4. If you throw an arrow to this pie chart, in which colour is the arrow more likely to fall on? [CBSE Handbook]



   a. Red ◯ b. Blue ◯
   c. Orange ◯ d. Green ◯

5. If you select a balloon from a bag having equal number of red, green and yellow balloons, how likely is it that you pick up a blue balloon? [CBSE Handbook]
   a. Probable ◯ b. Certain ◯
   c. Unlikely ◯ d. Impossible ◯

6. With one throw of a 6-sided die, what's the probability of getting an even number? [CBSE Handbook]
   a. 1/5 ◯ b. 2/5 ◯
   c. 5/6 ◯ d. 1/2 ◯

7. Time consistently moves forward, and each passing moment is _____ .
   a. uncertain ◯ b. likely ◯
   c. certain ◯ d. unlikely ◯

8. GPT-3, released by OpenAI in 2020, has 175 billion parameters and uses _____ sampling techniques for text generation.

a. data ◯ b. probabilistic ◯

c. statistical ◯ d. pattern ◯

B. Fill in the blanks.

1. _____ can see patterns in different types of data.
2. _____ compare the weather conditions with the information about past seasons and conditions.
3. In _____ events chances of each event happening is same.
4. The probability that you can pick a red ball from a bag containing only blue balls is _____ .
5. During games, coaches use _____ models to make strategic decisions.
6. If someone asks you to guess a number between 1 and 10, each number has an _____ probability of being selected.
7. _____ helps us make predictions about future events based on the data we have.
8. _____ can be used to analyse data to create epidemic curves that show the progression of disease outbreaks over time.

C. State whether the following statement is true or false.

1. Data is explored and cleaned to be used. _____
2. Getting seven in die thrown is a possible event. _____
3. Probability can be used in estimating batting average in cricket. _____
4. Statistics can only be used in the mathematical field because of its ability to extract meaningful insights from data and assist in decision making. _____
5. Analysis of data is done to understand it better. _____
6. Linear algebra is involved in small data processing. _____
7. Understanding math will help us to better understand AI and its way of working. _____
8. The design of the algorithms often uses scientific derivations. _____

SECTION B (Subjective Type Questions)

A. Short answer type questions.

1. How is mathematics and AI related? Explain in short.
2. What are patterns? Explain with examples.
3. What is meant by equal probability events?
4. What is the role of "Collecting Data" in statistics?
5. Give any two applications of statistics in real life.
6. If you have 10 red dresses and 3 white dresses. What is the probability of wearing white dress?
7. Give one use of statistics in disaster management.
8. Give one use of probability in finance.

B. Long answer type questions.

1. "Statistics is used for collecting, exploring, and analysing the data." Elaborate with the help of an example.
2. List any three uses of statistics in education.

3. Explain the concept of probability with the help of an example of a deck of 52 cards.
4. When discussing probability, we often rely on specific terms to describe the likelihood of events occurring. Explain any one likelihood of an event with examples.
5. What role does probability play in estimating the traffic on the road? List any three with examples.
6. Identify the likely, unlikely, impossible and equal probability events from the following with proper explanation:

   a. Tossing a coin
   b. Rolling an 8 on a standard die
   c. Throwing ten 5's in a row
   d. Drawing a card of any suite
7. Write any two examples of impossible and equal probability events.
8. Define certain events and likely events with examples.

C. Competency-based/Application-based questions:

Problem Solving & Logical Reasoning

1. Imagine a student preparing for a math exam that covers five topics. Based on past exams and the instructor's hints, the student estimates the probability of each topic appearing on the exam as follows:

   Topic A: 0.8

   Topic B: 0.6

   Topic C: 0.4

   Topic D: 0.7

   Topic E: 0.5

   How will this help him in preparing for the exam to score good marks?
2. Let's say a company wants to launch a new product—a smartphone—into the market. Before launching the product, the company conducts market research to understand consumer preferences and potential demand for the new smartphone. What role will the statistics play in smoothening this process?
3. Predicting earthquakes with precise accuracy is incredibly challenging due to the complex nature of seismic activity. However, probability can still play an important role in this. Can you find a few of the important applications of probability in predicting earthquakes?
4. Aman is confused, how probability theory is utilised in artificial intelligence, help Aman by providing two examples to illustrate its importance.

Ai In Life

Students are introduced to the concept of probability at a very early stage of life, it begins with a real life situation. For example, flip a coin and ask what the chances are that it will come out heads. Or, place the coin in one hand, and put both hands behind your back. Ask one student to guess which hand it is in.

Discuss your findings with your classmates, and illustrate some more examples from real life.

Ai Deep Thinking

The adoption of AI in healthcare could save between 5% to 10%. AI in healthcare statistics shows that 90% of nursing tasks will still be performed by humans in 2030.

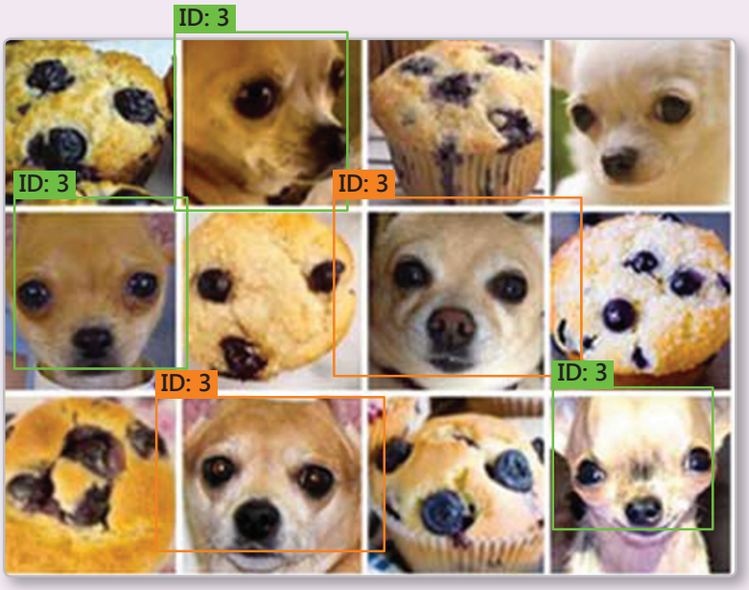
Considering the above facts, create a report on which all tasks of nurse shall be taken over by AI.

Class Activity

Problem Solving & Logical Reasoning

Communication

1. Identify Dogs [CBSE Handbook]



ID: 0

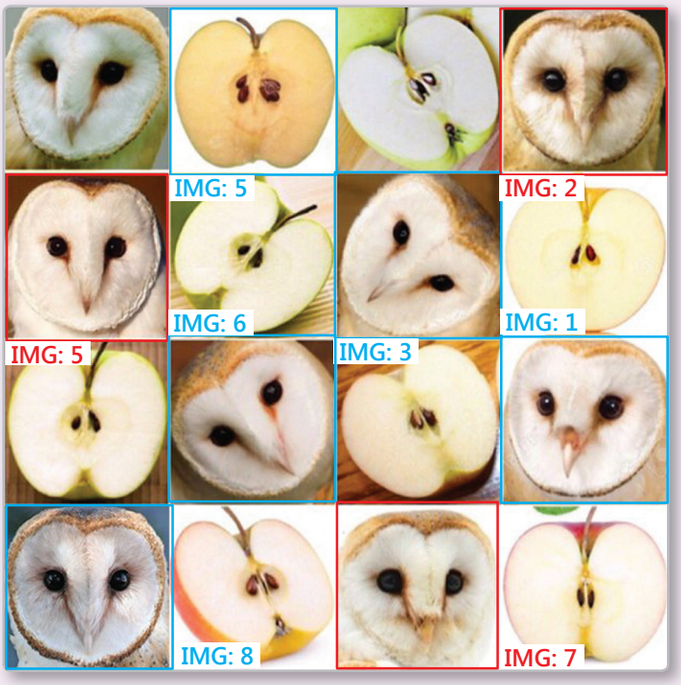
Type: Dog

Breed: Chihuahua (41.0%)

Emotion: Scared (98.0%)

Scared (98.0%), Angry (2.0%), Happy (0.0%), Neutral (0.0%), Sad (0.0%)

2. Identify Owl from Apples



Experiential Learning

Type: Owl

Breed: Barn

Active: 100% (Awake) 0% (Sleeping)

Takeaways from the above activity

- Just like we can recognise patterns in numbers, words, pictures, etc. AI can also recognise similar patterns.
- AI is a way to recognise patterns in order to make decisions.
- AI needs Math to study and recognise patterns in order to take decisions.

3. Can you identify any pattern in the image given below? [CBSE Handbook]

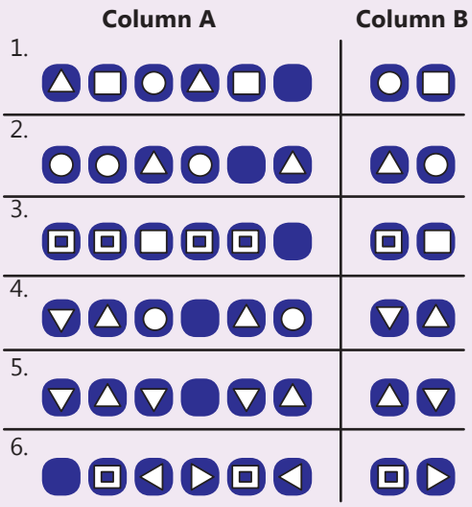
$$1 \times 9 + 2 = 11$$
$$12 \times 9 + 3 = 111$$
$$123 \times 9 + 4 = 1111$$
$$1234 \times 9 + 5 = 11111$$
$$12345 \times 9 + 6 = 111111$$
$$123456 \times 9 + 7 = 1111111$$
$$123457 \times 9 + 8 = 11111111$$
$$12345678 \times 9 + 9 = 11111111$$
$$123456789 \times 9 + 10 = 111111111$$

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1

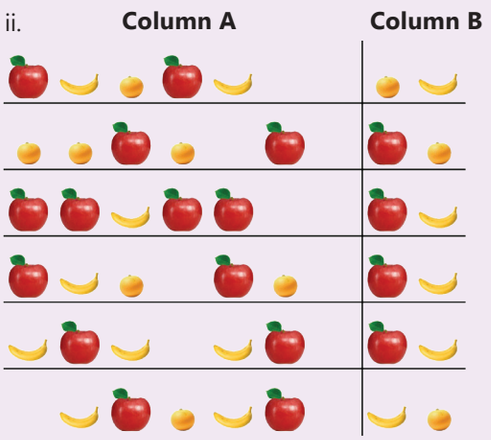
4. Find connections between sets of images and use that to solve problems, think smartly, and grasp tricky ideas.

Complete the sequence in the left column by identifying the correct missing piece in the right column out of the given options. [CBSE Handbook]

i.



ii.



Data Collection

- Visit the following link:
  https://www.youtube.com/watch?v=4A5L3x3TVuc&ab_channel=CarvingCanyons
- Fill the table while watching the video using tally.

Car Colour	Number of Cars spotted
Red	
Black	
White	

Use the given Reference Tally to maintain the count of each

1	I	6	~~𝍸~~ I
2	II	7	~~𝍸~~ II
3	III	8	~~𝍸~~ III
4	IIII	9	~~𝍸~~ IIII
5	~~𝍸~~	10	~~𝍸~~ ~~𝍸~~

Answer the following questions based on the table filled above:

Data Analysis

How many cars are spotted in total?

Which colour has been spotted the maximum amount of time?

Data Interpretation

What is the most common colour choice for the residents of this area?

Answers

AI Quiz Section A (Objective Type Questions)

A. 1. a 2. d 3. a 4. a 5. c 6. b 7. c 8. a

B. 1. Artificial Intelligence 2. Graph theory 3. Statistics 4. Probability 5. Unlikely 6. one 7. Probability 8. Fibonacci series

C. 1. False 2. False 3. True 4. True 5. False 6. False 7. True 8. True

Answer the following questions:

1. Explain one example in your day to day life where you think that AI and Maths play a vital role.

2. When you order food online, you receive a link for the feedback. Based on the questions asked by the chatbot, you are able to give your feedback. What is the probability of the feedback shared by you is not altered by chatbot?

3. How is probability affecting the weather forecast system?

UNIT-4

INTRODUCTION TO GENERATIVE AI

Learning Outcomes

- Real Images vs AI-Generated Images
- Supervised Learning and Discriminative Modelling
- What is Generative AI?
- Examples of Generative AI
- Generative AI Tools
- The Potential Negative Impact on Society
- AI or Real Image....How to Identify?
- Unsupervised Learning and Generative Modelling
- Types of Generative AI
- Generative AI Boon or Bane
- Ethical Considerations of using Generative AI
- Responsible Use of Generative AI

The simple it is to understand the context through visualisation (images), more difficult it is to identify, whether the image is real or artificial. Today 50% of the images we see online are a result of Generative Artificial Intelligence.

Let us learn more about it.

Real Images vs AI-Generated Images

Real images are captured by cameras, they are visual representation depicting scenes, objects or people in the same way as they exist in the real world. They are created on the same side of the lens or mirror as the viewer. These images are either created by or clicked by humans.

AI-generated images are created using AI algorithms. These algorithms use large amount of data and learn patterns to create new images that look like real ones. Sometimes AI incorporates small details that don't exist in the original picture to enhance the look of the scene. AI can create images that can be modified and enhanced. It can also create entirely new, imaginative images.

Real Image

AI Generated Image

Ai Task

Problem Solving & Logical Reasoning

Examine the images marked as Image 1 and Image 2 and determine whether the image is a real image or an AI-generated image. Also, give reasons for your answer.

Image 1	Image 2
	
1. _____	a. _____
	
2. _____	b. _____
	
3. _____	c. _____

Image 1	Image 2
	
4. _____	d. _____
	
5. _____	e. _____

Brainy Fact

Companies like OpenAI, Google, Facebook, and NVIDIA have invested heavily in research and development of generative AI technologies.

AI or Real Image....How to Identify?

Distinguishing between a real image and one generated by AI can be challenging as AI-generated images continue to become more sophisticated. However, there are a few indicators you can look for:

- Artificial intelligence shows inconsistencies if observed closely, although it tries to piece together its creations from the original work. The artifacts can include unnatural blurriness, inconsistent lighting/shadowing, or repeating patterns specially in the backgrounds.
- AI-generated images may include elements that seem unrealistic or improbable, such as impossible perspectives, mismatch colours, or objects that defy physics making the image appear unnatural or inconsistent with the scene.
- Odd outlines to sharpen or smoothen the edges, stray pixels to cover inconsistency, and abnormal shapes can be easily seen, if an image is zoomed to the maximum, on each of its parts.

Let's look at the concepts behind the generation of these images.

Supervised Learning and Discriminative Modelling

Supervised Learning is a type of Machine Learning where we teach models using examples that have labels. It uses labelled datasets to train algorithms to predict outcomes and recognise patterns.

These labels tell the model the correct answer for each example. Discriminative Modelling is a special kind of supervised learning that focuses on learning how to distinguish different classes. It looks at the features of the data to figure out which class it belongs to.

Supervised Learning

Supervised Learning is a Machine Learning where a model is trained on a labelled dataset, implying that each input data point is associated with a corresponding output label. The goal of supervised learning is to learn the mapping between input data and output labels, enabling the model to make predictions on new, unseen data.

Supervised learning is when we train the machine using labelled data. The machine is provided with a new set of labelled data so that the supervised learning algorithm analyses the training data and generates the most suitable and related outcome from the trained-labelled data. Labeled data contains data with the correct output or classification. In simple words, input data is paired with the desired output thus making the machine learn to predict the output for new input data.

For example, in the given images, first is the input image and characteristics of this image are marked as boy and ball which can be seen in center image. Now according to supervised learning it has to learn the mapping between input labels and output labels, which is shown in last image and highlights "ball" as red, "boy" as purple and "boy playing with a ball" in a rectangle.

Input	**Features of given image**	**Output** Label for the item
		
	Boy Ball	Boy playing with a ball

In a supervised learning model, a labelled dataset is given to the machine. A labelled dataset is the information which is tagged with identifiers of data. For example, clothes in a store are marked under various categories of clothing like Shirts, Trousers, Coats, etc. They are further labelled as per gender and size.

Discriminative Modelling

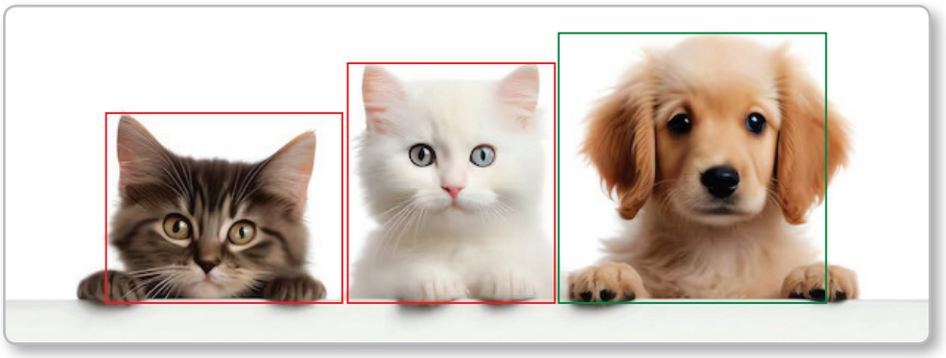
Discriminative Modelling is an approach in Machine Learning where the focus is on learning the boundary or decision boundary that separates different classes or categories directly from the data. So, if an image contains a combination of Dogs and Cats, the model is able to tell which is a Dog and which is a Cat.

Input Image

Label: Dogs

Label: Cats

Output



Cats = 2 Dog = 1

In Supervised Learning, Discriminative Modelling contrasts with Generative Modelling, where the goal is to model the joint probability distribution of both the input features and the output labels. Generative models can be used to generate new data points that resemble the training data, whereas discriminative models are primarily focused on classification or regression tasks.

Let us consider an example.

A father has two kids, Kid A and Kid B. Kid A has a special character whereas he can learn everything in depth. Kid B have a special character whereas he can only learn the differences between what he saw.

One fine day, father shows his kids (Kid A and Kid B) two kinds of animals say a dog and a cat. After a few days, father showed them an animal and asked both of them "Is this animal a dog or a cat?"

Kid A drew the image of dog and cat on a piece of paper based on what he saw earlier. He compared both the images with the animal standing before him and answered based on the closest match of image & animal, he answered: "The animal is Dog." Kid B knows only the differences, based on different properties learned, he answered: "The animal is a Dog".

Here, we can see both of them is finding the kind of animal, but the way of learning and the way of finding answer is entirely different. In Machine Learning, we generally call Kid A as a generative model & Kid B as a discriminative model.

A discriminative model models the decision boundary between the classes. A generative model explicitly models the actual distribution of each class. In final both of them is predicting the conditional probability P (Animal Features). But both models learn different probabilities.

Unsupervised Learning and Generative Modelling

Unsupervised Learning is a type of Machine Learning where models are trained using data that does not have labels. This means the model has to find patterns and relationships in the data on its own. Generative Modelling is a specific approach within Unsupervised Learning that focuses on understanding and modelling how the data is generated. Generative models try to learn the underlying rules that produce the data, so they can create new examples that look similar to the original data. In summary, Unsupervised Learning is about finding patterns in unlabelled data, and Generative Modelling is a method within this type of learning that aims to understand and replicate how the data is made.

Unsupervised Learning

Input Unstructured/Unlabelled dataset	**Output** Emergent pattern/inherent structure	Example that's similar to what's in the dataset
		

The goal of Unsupervised Learning is to find patterns, structure, or representations in the data without human intervention. An Unsupervised Learning approach works on an unlabelled dataset. This means that the data which is fed to the machine is random and there is no know-how available about it to the trainer.

Ai Reboot

1. Differentiate between Supervised Learning and Unsupervised Learning approach.

   _____

   _____

2. List any two points to differentiate between real and AI-generated images.

   _____

   _____

Generative Modelling

Generative Modelling do not necessarily require labelled datasets. It can work with unlabelled data to learn the underlying distribution of the data and can **generate structured data from the random noise dataset**. So, if random images are fed as training data for the model it can create relevant output based on the features of the input data. If there are random images which depict streets, cars, buildings, sky, etc. In a given dataset of street images, a Generative Modelling can learn to generate new street scenes that look like the ones in the dataset. In another example, if given a dataset of news articles, a generative model can learn to generate new articles that resemble the style and content of the training data. Let us take an example.

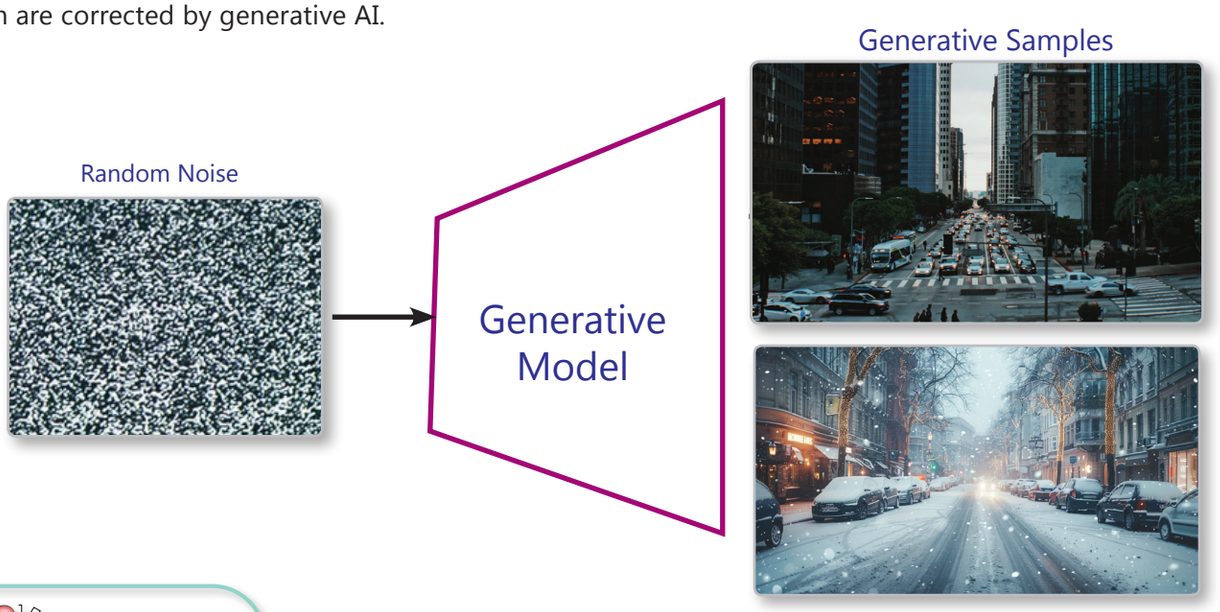
The following images are given as input to the Generative AI model:

The output produced based on input images by generative AI are as follows:

The Generative AI model takes the input and can recreate its own set of objects in the image which may or may not exist in reality.

Random Noise Dataset

A "random noise dataset" typically refers to a collection of data points or samples where each data point is generated randomly. There are unpredictable fluctuations and disarranged data which makes it impossible to identify target patterns or relationships in it. This may result in decreased accuracy or reliability of the output, which are corrected by generative AI.



Ai Task

Problem Solving & Logical Reasoning

Based on your learning so far, identify the real and AI generated image among the given set of images. Also state any three reasons for the same.

AI Reboot

What is “Random Noise” dataset? Give an example of it from life around you.

What is Generative AI?

Generative Artificial Intelligence (AI), also known as Gen AI, encompasses algorithms designed to produce new data closely resembling human-generated content across various forms, including audio, code, images, text, simulations, and videos. Leveraging existing data and content for training, Generative AI holds promise for applications spanning natural language processing, computer vision, the metaverse, and speech synthesis. As it undergoes continuous training with more data, this technology evolves, progressively refining its capabilities. Its growing popularity stems from the intuitive interaction it enables, allowing users to prompt the AI using natural language. Notably, the output generated by generative AI often achieves a level of quality indistinguishable from human creation. Within Generative AI, a "prompt" serves as the initial input or instruction guiding the AI model in generating desired content. For instance, in a text generation model, a prompt may consist of a sentence or keywords, while in an image generation model, it could entail a description of the desired image.

Experiential Learning

Video Session

Scan the QR code or visit the following link to watch the video:

https://www.youtube.com/watch?v=26fJ_ADteHo

This video gives you a clear picture of generative AI. Now answer the following questions:

1. What is your understanding of generative AI?

2. Give two examples of generative AI.

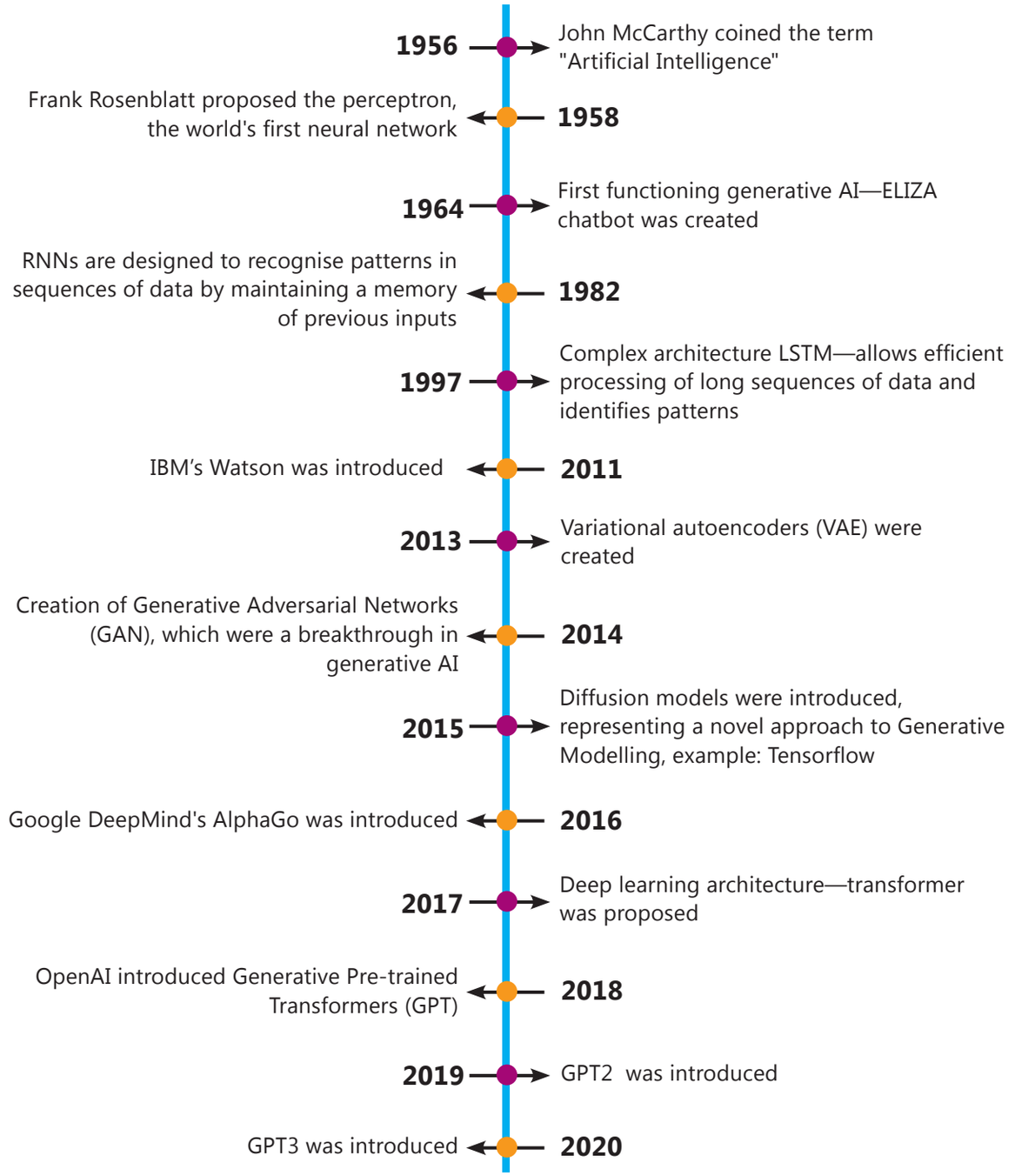
Brainy Fact

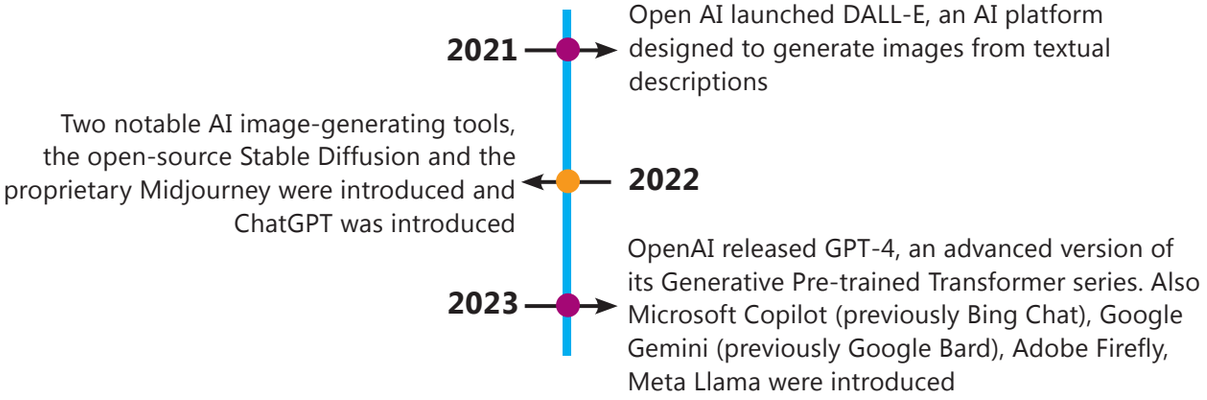
Generative AI is now part of the workforce in the US, UK, Australia, and many other parts of the world.

Timeline of Generative AI

Generative AI has been getting better and better over the years. Scientists have been working hard to improve the technology using things like neural networks and deep learning. They've been trying out different ideas and making big discoveries in how to make it work better. Now, Generative AI can do lots of cool stuff like writing text, making pictures, and creating new things. It's been a long process of learning and making things better, but now we can see all the amasing things it can do!

Timeline of Generative AI:





Generative AI vs Conventional AI

Generative AI and Conventional AI represent two different approaches in the field of artificial intelligence. The difference between them is given in the following table:

	Generative AI	Conventional AI
Goal	Generative AI creates new content which mimics the original content. This content includes images, text, music, or other forms of media.	Conventional AI analyses, processes, and classifies data. It works to improve the accuracy, precision, recall, and speed within the scope of the defined task.
Training	Generative AI models are often trained using techniques such as generative adversarial networks (GANs), variational autoencoders (VAEs), or autoregressive models.	Conventional AI models are typically trained using supervised, unsupervised, or reinforcement learning techniques.
Dataset	Generative AI models typically require large amounts of diverse and representative data to learn effectively. These datasets often contain thousands or even millions of examples across various categories or classes.	Conventional AI models rely on smaller, more curated datasets that are tailored to the task at hand.
Output	Generative AI output is fresh, innovative, and often unexpected.	Conventional AI produces more predictable output based on existing data.
Applications	Generative AI is used in the fields of art, music, literature, gaming, and design.	Conventional AI is used in banking, healthcare, image recognition, and language processing.

Types of Generative AI

Generative AI comes in a variety of forms, each with unique advantages and uses. Some of the most typical varieties are as follows:

- Generative Adversarial Networks (GANs)
- Variational Autoencoders (VAEs)
- Recurrent Neural Networks (RNNs)
- Autoencoders (AEs)

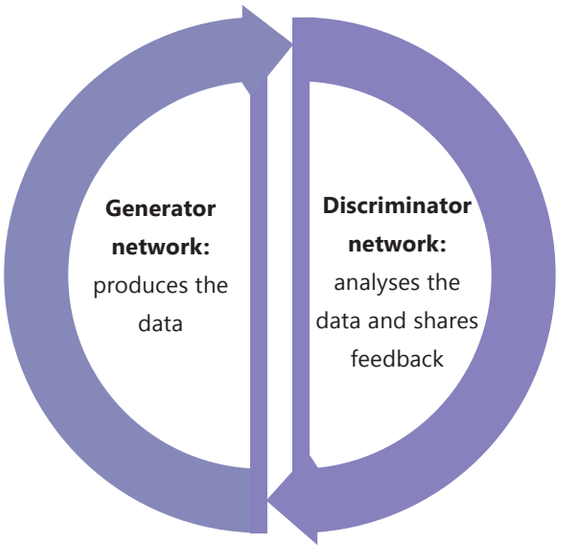
Let's study about them in detail.

Generative Adversarial Networks (GANs)

GANs are neural networks that work to produce fresh data. It is made up of two neural networks which work together in a unique adversarial process to create realistic synthetic data. These two neural networks are as follows:

- **Generator Network:** It produces the data that is as close as possible to real data.
- **Discriminator Network:** It analyses the data and provides feedback, i.e. it takes real data and the data generated by the generator as input and attempts to distinguish between the two.

These two networks work together in a cycle where the generator tries to create realistic fake data, and the discriminator tries to identify whether the data is real or fake. This back-and-forth process helps the generator improve and produce more convincing data over time.

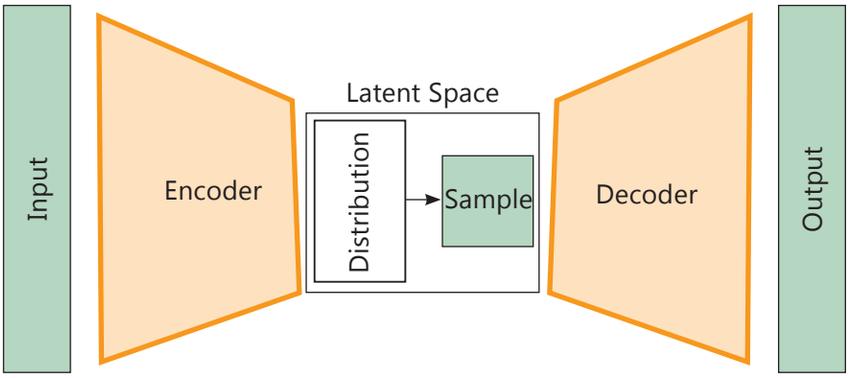


Some of the examples of GANs are as follows:

- It can create portraits of non-existing people.
- It can convert images from day to night.
- It can generate images based on textual description, for example, if we give description of a bird then it will create an image that is similar to the description.
- It can generate realistic video which can be used in film production, video games, and generating synthetic data for training other AI models, etc.

Variational Autoencoders (VAEs)

A variational autoencoder (VAE) is a Generative AI algorithm that uses deep learning to generate new content, detect anomalies and remove noise. This is another class of generative models. To produce fresh data, VAEs learn the distribution of the data and then sample from it.

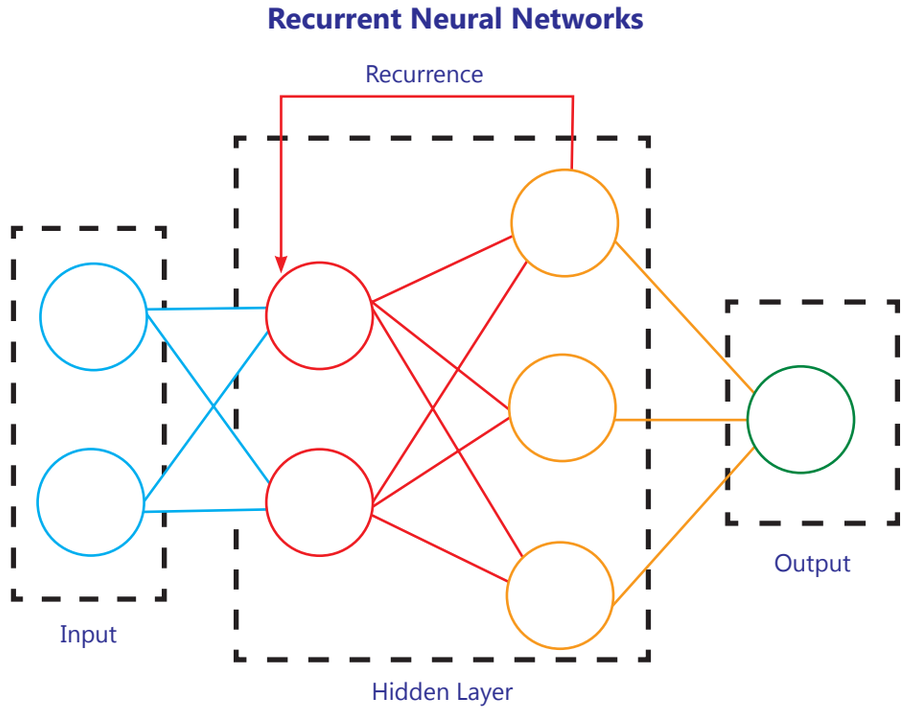


Some of the examples of VAEs are as follows:

- It can generate new images like the given training set. For instance, a VAE trained on images of faces can generate new, realistic-looking faces.
- VAE's can produce new text that follows the same style and structure as the training data, assisting writers with drafts and ideas.
- It can be used for composing new music pieces or creating sound effects, music composition etc.

Recurrent Neural Networks (RNNs)

RNNs are a special class of neural networks that excel at handling sequential data, like music or text. They excel at tasks where the order of the data points is important, as they can remember previous inputs and use this information to influence current outputs.

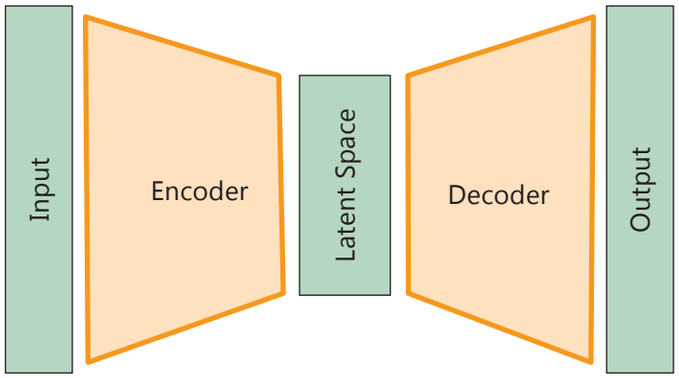


Some of the examples of RNNs are as follows:

- It can generate novel text in the style of a specific author or genre, like creating new sentences that mimic the style of Shakespeare or generating dialogue for a chatbot.
- It can predict the next character or word in a sequence, like autocomplete features in text editors and predictive text input on smartphones.
- It can be used to predict future values in a time series, such as stock prices or weather data, by learning patterns from historical data.

Autoencoders (AEs)

These are neural networks that have been trained to learn a compressed representation of data. They work by compressing the data into a lower-dimensional form (encoding) and then decompressing it back to its original form (decoding). This process helps the network learn the most important features of the data.



Some of the examples of AEs are as follows:

- It can help in cleaning up noisy images to produce clear and highly realistic samples.
- It can help in compressing high-resolution images for efficient storage and transmission.
- It can create artistic images based on learned features from famous paintings.
- It can help in drug discovery by learning and generating molecular structures that have desirable properties.

Autoencoders (AEs) and Variational Autoencoders (VAEs)

The similarities and differences between Autoencoders (AEs) and Variational Autoencoders (VAEs) are as follows:

Similarities

- Both AE and VAE are neural network architectures that are used for Unsupervised Learning
- Both AE and VAE consist of an encoder and a decoder network. The encoder maps the input data to a latent representation, and decoder maps the latent representation back to the original data.
- Both AE and VAE can be used for tasks such as dimensionality reduction, data generation, and anomaly detection.

Differences

	AE	VAE
Basic Function	Neural network model that learns to encode input data into a compressed representation and then decode it back to the original data.	Similar to AE but incorporates probabilistic elements to learn a latent space representation of input data.

	AE	VAE
Latent Space Representation	Deterministic, fixed-dimensional encoding of input data.	Probabilistic, continuous latent space representation, allowing for sampling of data points.
Reconstruction Loss	Minimises the difference between the input data and its reconstructed output.	Same as AE but also includes a regulariser to enforce a Gaussian-like distribution in the latent space.
Handling Overfitting	Can suffer from overfitting due to the fixed encoding structure.	Less prone to overfitting due to the probabilistic nature of the latent space, which allows for smoother generalisation.
Applications	Image compression, denoising, feature extraction.	Data generation, Unsupervised Learning, anomaly detection.
Training Complexity	Relatively simpler training process.	More complex training process due to the inclusion of regularisation terms and sampling from the latent space.

Examples of Generative AI

Generative AI has many applications, from art and music to language and natural language processing.

Let's study about some examples of how Generative AI is being used in various fields.

Art

Generative AI can create new artworks by learning styles from famous painters and generating novel pieces in similar styles. For example:

- AI artists like "AI Portraits" and "DeepArt" have gained popularity for their ability to create visually stunning images.
- The Next Rembrandt project used data analysis and 3D printing to create a new painting in the style of Rembrandt.

Gen AI can be used for Style Transferring & Portrait Creation

Object	Style
	
Gen AI created an image with given style	
	
Style Transferring	

Portrait Creation

Experiential Learning

Video Session

Scan the QR code or visit the following link to watch the video:

https://www.youtube.com/watch?v=IuygOYZ1Ngo

This video is on The Next Rembrandt. Now, answer the following question:

What are the points to be noted while using generative AI in recreating traditional art?

Music

Generative AI is transforming the music industry by enabling the creation of new music, either through composing original pieces or remixing existing ones.

One prominent example of this innovation is AIVA, an AI composer capable of creating original music in various genres.

Music composition

Music Genre

Experiential Learning

Video Session

Scan the QR code or visit the following link to watch the video:

https://www.youtube.com/watch?v=wYb3Wimn01s

This video is on How AI could compose a personalised soundtrack to your life | Pierre Barreau. Now, answer the following question:

What is the name of AI composer mentioned in the video?

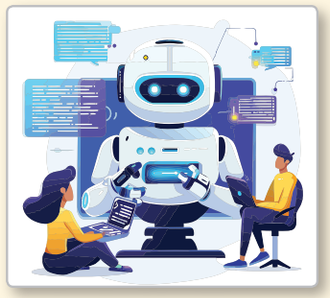
Language

Generative AI is being used to generate new languages, such as chatbots that can hold conversations with users or natural language generation systems that can produce written content.

GenAI in Language

Language translation

Text to speech



Data creation

Experiential Learning

Video Session

Scan the QR code or visit the following link to watch the video:

https://www.youtube.com/watch?v=BWCCPy7Rg-s

This video is on What is ChatGPT, the AI software taking the internet by storm? – BBC News. Now, answer the following question:

ChatGPT is being used by masses, what are the concerns arising because of it?

Generative AI Boon or Bane

Generative AI, a powerful new technology, has both advantages and disadvantages. On one hand, it helps people be more creative, makes complex tasks easier, and sparks new ideas in various fields. Artists and writers can use it to come up with fresh concepts, and ideas, and businesses can work more efficiently. On the other hand, it can cause problems like ethical issues, job loss, and being used for harmful purposes. For example, it can create fake images or spread false information, which can be dangerous. To make the most of Generative AI while avoiding its risks, we need smart rules and careful use. It can be a great tool or a big problem, depending on how we manage it.

Benefits of using Generative AI

In general, Generative AI brings many good things to the table. It boosts creativity, helping to come up with new and interesting ideas. It also makes things run smoother and faster, saving time and effort. Plus, it tailors things to each person's preferences, making experiences more personal. With Generative AI, businesses and groups can explore new possibilities and reach more people. It's also easier for everyone to use, making it accessible to a

wider audience. And as needs grow, Generative AI can easily adapt and handle bigger tasks. By tapping into these advantages, businesses and organisations can make their content creation processes better and give their users even better experiences.

- **Creativity:** Generative AI is a real game-changer when it comes to creativity. It helps artists, designers, and musicians explore new horizons and make their creative work more efficient and personal. Whether it's coming up with fresh and innovative ideas, refining designs, or composing music, Generative AI can lend a hand in making the creative process smoother and more tailored to individual tastes. This is super valuable in fields like art, design, advertising, and music, where pushing the boundaries and expressing unique visions are key.

- **Efficiency:** Generative AI is all about making things easier and quicker. By automating content creation, it helps save a ton of time and money compared to doing things manually. Instead of spending hours or even days creating content, Generative AI can whip up something in a fraction of the time. This efficiency boost is a huge advantage, especially for businesses and organisations looking to streamline their operations and get things done faster.

- **Personalisation:** Generative AI takes personalisation to a whole new level. It can tailor content specifically for each person, taking into account their likes, dislikes, and behaviors. This means that instead of getting generic recommendations or articles, users get content that's made just for them. Whether it's suggesting products they might love or delivering news articles on topics they're interested in, Generative AI ensures that each user's experience is unique and relevant to their tastes and preferences. It's like having a personal assistant that knows exactly what you want, making the online experience more enjoyable and engaging.

- **Exploration:** Generative AI helps us explore new things and make existing stuff better. For example, it can help scientists design new drugs or make industrial processes work smoother. It's like having a super-smart assistant that helps us discover new ideas and improve how things work in different fields.

- **Accessibility:** Generative AI makes it easier for everyone to create top-notch content, even if they don't have fancy tools or tons of know-how. It's like leveling the playing field, giving everyone a chance to make something awesome without needing special skills or expensive equipment. Whether it's designing graphics, writing stories, or making music, Generative AI opens up a world of creative possibilities for people from all walks of life.

- **Scalability:** Generative AI is a powerhouse when it comes to creating lots of content in a short time. It's like having a super-speedy content creator that can churn out stuff at large scale without breaking a sweat. This makes it perfect for businesses and organisations that need to produce heaps of content, whether it's articles, images, videos, or anything else. With Generative AI, scaling up content production is a breeze, helping businesses keep up with demand and reach more people without sacrificing quality.

Limitations of Using Generative AI

Data bias: When Generative AI learns from biased or incomplete data, it tends to reflect those biases in its output. This means the results may be skewed or flawed, especially in critical areas like facial recognition or natural language processing, where accuracy is paramount.

Uncertainty: Generative AI has a knack for surprising us with unexpected results. While this can sometimes lead to exciting discoveries, it can also be a bit of a mixed bag, as the outcomes can be unpredictable.

Computational demands: Generative AI isn't shy about its appetite for computational power. It demands hefty resources for training and generating output, which can be both expensive and time-consuming. So, while it's undeniably powerful, it can also drain the resources.

Ai Task

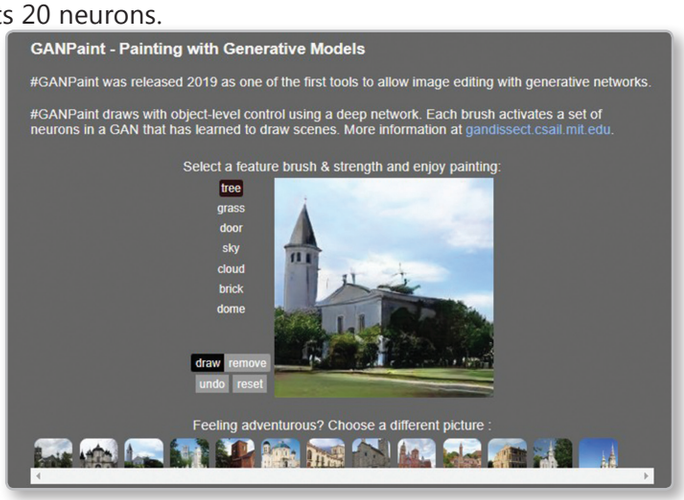
Experiential Learning

GAN Paint

CBSE Handbook

Link: https://ganpaint-v2.vizhub.ai/

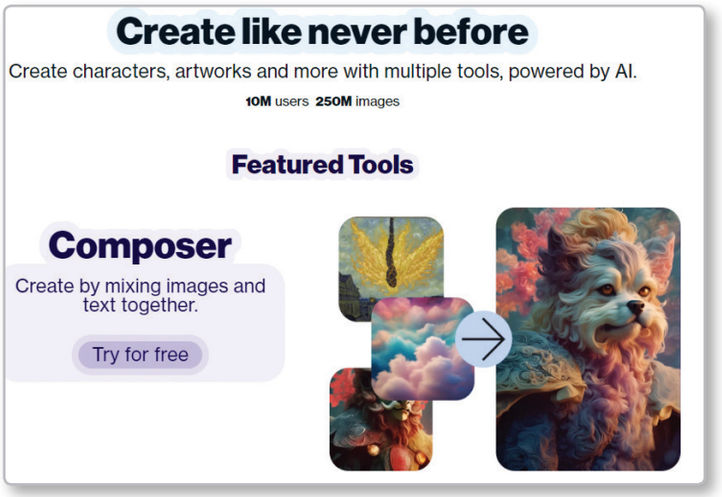
- GAN Paint directly activates and deactivates neurons in a deep network trained to create pictures.
- Each left button ("door", "brick", etc.) represents 20 neurons.
- The software shows that the network learns about trees, doorways, and roofs by drawing.
- Switching neurons directly shows the network's visual world model.
- To use GAN Paint, you will first need to select a base image from the website's library. You can then use the brush tool to add objects and textures to the image. As you paint, the GAN network will learn to generate more realistic images.
- You are encouraged to experiment with GAN Paint and see what you can create. Have fun!



Generative AI Tools

There are many generative AI tools available today that enable users to create and experiment with generative models. These generative AI tools offer strong abilities for both creative and practical uses in many areas. By using these tools, people can improve their projects, create new content, and discover the potential of AI-driven creativity. Let's study about some of the popular tools.

Artbreeder



Artbreeder is a web-based tool where you can make and change pictures using advanced AI technology called generative adversarial networks (GANs). With Artbreeder, you can mix different images and text together and adjust specific features to create completely new and unique artworks.

You may work on Artbreeder using the given link https://www.artbreeder.com/

You can use it for free with some limitations or choose a paid plan for more features and options. It helps you:

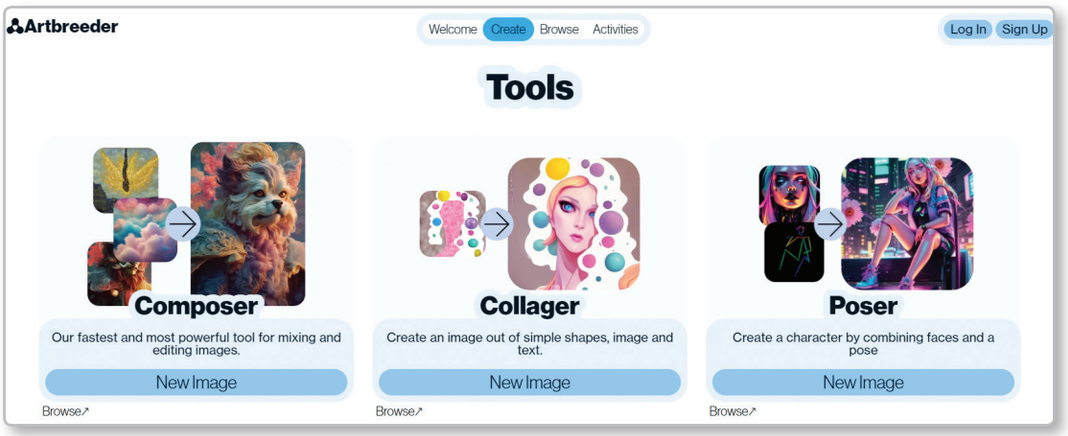
- create new characters by blending and modifying existing images.
- generate imaginative and unique artworks to use in stories, games, or movies.
- experiment with different image combinations and features to discover new artistic possibilities.

Ai Task

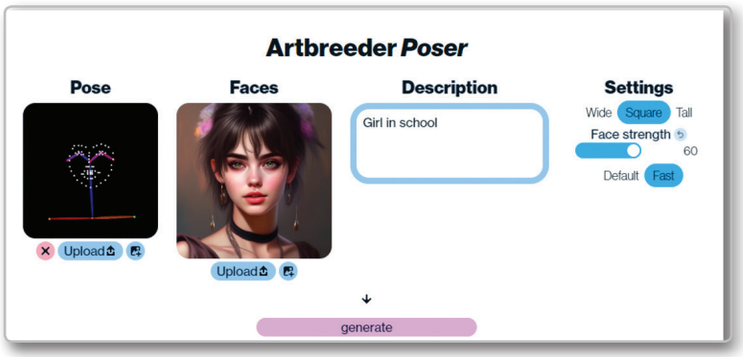
Experiential Learning

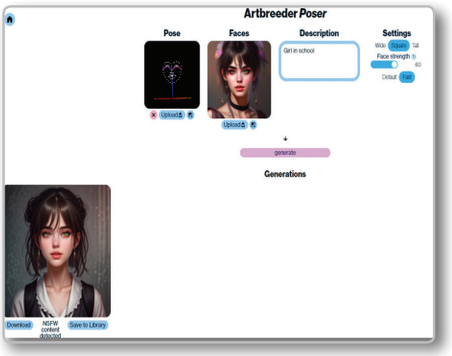
Generate Images with Text Prompt

1. Go to https://www.artbreeder.com/
2. Select **Create** from the menu bar and click on **New Image** under the **Poser** category.



3. Enter text in the **Description** box and click on the **generate** button. See how AI generates a picture based on the description.





Now, create a new image of a boy standing in rain.

You may try creating new images.

ChatGPT

It is an AI tool created by OpenAI. It generates responses like humans in real-time, based on the user's input. It can give natural answers to questions in a conversational tone and can generate stories, essays, and poems.

You may work on ChatGPT using the given link https://chat.openai.com.

It can:

- answer any type of questions
- solve maths or scientific problems
- translate between languages
- debug and fix code
- write a story/poem
- differentiate the things given as input
- rephrase text input

Runway ML

Runway ML is a platform for creating, training, and deploying generative models. It provides a user-friendly interface for building and training various types of generative models, including GANs, VAEs, and image classifiers.

Experiential Learning

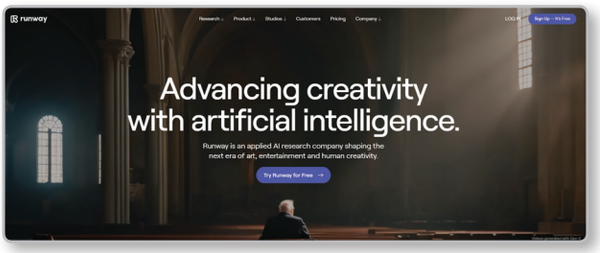
Video Session

Scan the QR code or visit the following link to watch the video:

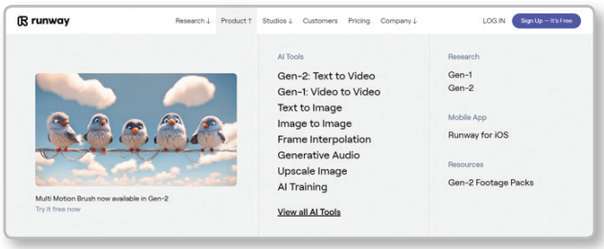
https://www.youtube.com/watch?v=trXPfpV5iRQ

Try the following activities:

1. Go to https://runwayml.com/.



2. Select a tool of your choice and generate new content with it



Create a video by giving text commands.

Brainy Fact

ChatGPT, the world's most popular genAI platform, had an average of 1.5 billion monthly visits in 2023.

Gemini

Gemini is a generative multimodal AI model created by Google. Just like ChatGPT, Google Gemini is designed to understand text, images, audio, video, computer code, and more.

It can:

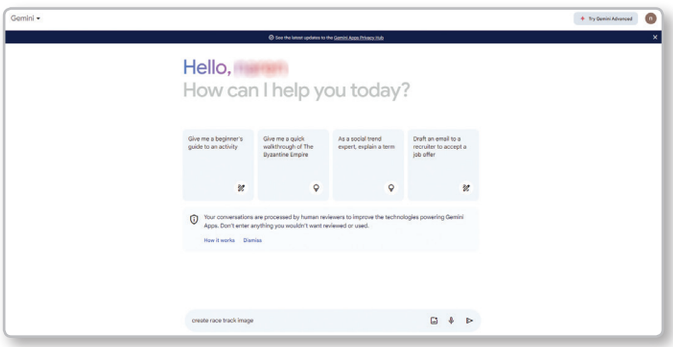
- answer any type of questions
- solve maths or scientific problems
- translate between languages
- debug and fix code
- write a story/poem
- differentiate the things given as input

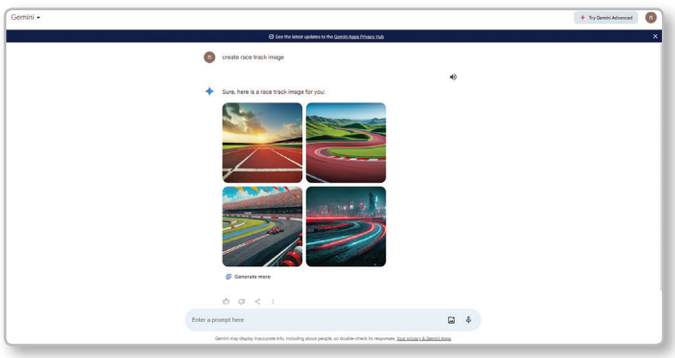
Ai Task

Experiential Learning

1. Go to https://gemini.google.com/.

2. Click on the **Chat with Gemini** button and follow the screens.
3. On the Welcome page, give the prompts listed below and see the answer that you get.





Now, explore some new AI apps which can help ease your work.

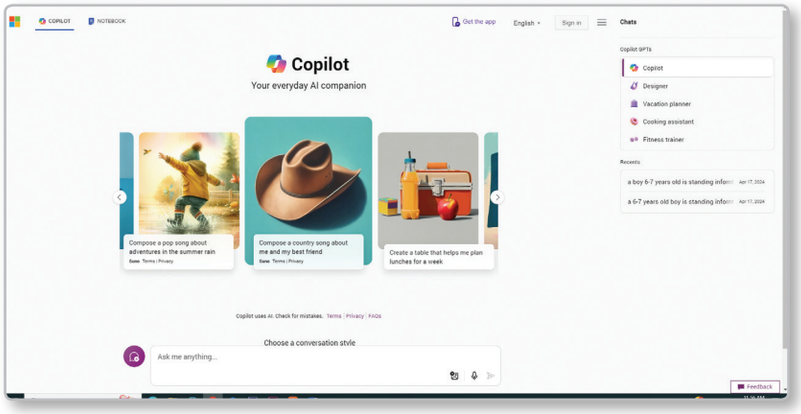
Copilot

Copilot is an AI tool designed by Microsoft. It can do all the jobs just like ChatGPT and Gemini but it focuses more on software development assistance to streamline the coding process, increase productivity, and assist developers in writing high-quality code faster.

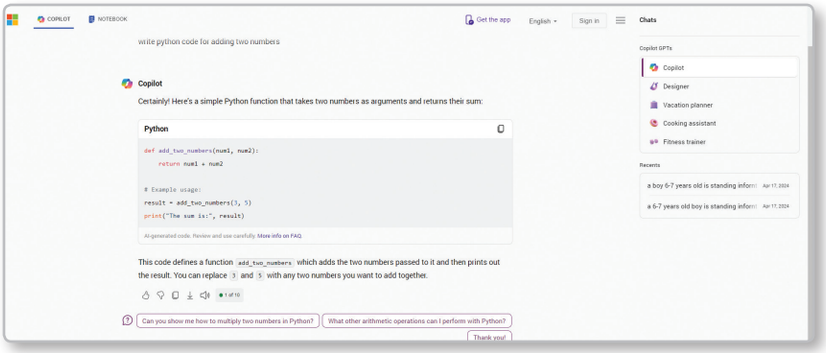
AI Task

Creativity & Innovativeness

1. Go to https://copilot.microsoft.com/.



2. Give the prompts listed below and see the answer that you get.



Now, create a new image of a man buying groceries.

Some More Generative AI Tools

Some of the popular generative AI tools that can be used in various fields are shown in the given table.

Video	Muse AI	Vista AI	Topaz AI
Text	Quillbot	Notion AI	Compose
Images	Midjourney	Magic Studio	Pebblely
Design	Viesus	Piggy AI	Galileo

Coding	Bugasura	CodeGPT	Replit Ghostwriter
Audio	FineShare	Boomy AI	Playlist AI
Productivity	Briefly AI	Socra AI	Leexi AI

Ethical Considerations of using Generative AI

While Generative AI offers many benefits, there are also several ethical considerations that should be considered when using this technology.

- **Ownership:** There's a gray area concerning who owns content created by Generative AI. This is especially significant in creative fields like music, literature, and art, where AI can produce original works that blur the lines between human and machine authorship.
- **Human Agency:** Generative AI prompts questions about human control and autonomy. As technology advances, it may become harder to distinguish between content made by humans and that made by machines. This blurring of lines could result in a loss of human agency and control.
- **Bias:** Generative AI learns from the data it's fed, meaning biased data can result in biased AI-generated content. This bias can have harmful effects, particularly in high-stakes areas like hiring, loan approvals, or criminal justice.
- **Misinformation:** Generative AI can be used to create fake news or deep fakes, which can spread misinformation and sway public opinion. This poses a threat to democracy and undermines trust in authorities.
- **Privacy:** There's a risk that Generative AI could be used to generate sensitive personal information, such as credit card numbers or medical records, which could then be exploited for malicious purposes. Protecting privacy is crucial in preventing such misuse.

The Potential Negative Impact on Society

1. **Spread of Misinformation and Fake News**
   a. **Fake News:** AI can generate convincing fake news articles and social media posts, spreading misinformation quickly and widely.
   b. **Deepfakes:** The term "deepfake" combines "deep learning" and "fake," referring to AI techniques that create realistic but fake videos and audios. These AI-generated videos can mislead people by making it seem like someone said or did something which they didn't, undermining trust in public figures and institutions.
2. **Lead to Job Displacement**
   a. **Automation:** Generative AI can perform tasks traditionally done by humans, such as writing, graphic design, and customer service, potentially leading to significant job displacement.

b. **Economic Disruption:** Workers who lose their jobs may struggle financially, face higher unemployment, and find it harder to get new jobs without the required skills.

3. **Privacy and Data Security Risks**

   a. **Sensitive Information:** AI can inadvertently or maliciously generate sensitive personal information, like social security numbers or medical records, which can be exploited for identity theft or fraud.

   b. **Data Breaches:** The misuse of AI to access or generate personal data poses significant risks to individual's privacy and security.

4. **Ethical and Moral Concerns**

   a. **Bias and Discrimination:** AI models trained on biased data can produce biased outputs, reinforcing stereotypes and discrimination.

   b. **Dehumanisation:** Relying too much on AI for tasks that need human understanding and compassion can make services and interactions feel less personal and caring.

5. **Security Threats**

   a. **Cyber Attacks:** AI can be used to develop sophisticated cyber-attacks, including generating malicious code or automated phishing schemes.

   b. **Weaponisation:** Generative AI could be used to create harmful technologies or weapons, posing significant national security risks.

6. **Environmental Impact**

   a. Training and running large AI models require significant computational resources, contributing to high energy consumption and environmental impact.

   b. Many devices are being exchanged due to outdated hardware leading to increase in e-waste.

Ai Reboot

1. Provide examples illustrating instances where biases in Generative AI are evident.

2. What is deepfake?

Responsible Use of Generative AI

The responsible use of Generative AI involves:

- Ensuring that the training data used are diverse and representative. Use datasets that reflect a wide range of demographics, cultures, and contexts to avoid biases in AI outputs. Regularly audit and adjust datasets to address and reduce biases.

- The outputs are scrutinised for bias and misinformation. Implement tools and processes to detect biases in AI-generated content.
- Prioritising user privacy and informed consent. Apply strong encryption and data anonymisation techniques to protect user data. Ensure users are aware of and consent to how their data is being used by AI systems.
- Educating Stakeholders on ethical use and risks. Provide ongoing education for developers on ethical AI practices and the potential risks of AI misuse. Educate users about the capabilities, limitations, and ethical considerations of AI technology.
- Establishing clear guidelines on ownership and attribution. Define and enforce guidelines regarding the ownership of AI-generated content. Clearly attribute AI-generated content to its sources, distinguishing between human and AI contributions.
- Engaging in public discussions around the social and ethical implications. Foster open discussions with the public about the benefits and risks of Generative AI. Collaborate with policymakers, ethicists, and other stakeholders to develop guidelines and regulations that ensure AI is used in socially beneficial ways.

All these points ensure the responsible use of Generative AI. By emphasising ethics, creating trust, limiting negative repercussions, defining legislation, and encouraging innovation, we maximise generative AI's potential and use it in ways that benefit the society.

At a Glance

- AI-generated images are created using AI algorithms.
- Distinguishing between a real image and one generated by AI can be challenging as AI-generated images continue to become more sophisticated.
- Artificial intelligence shows inconsistencies if observed closely, although it tries to piece together its creations from the original work.
- AI-generated images may include elements that seem unrealistic or improbable, such as impossible perspectives, mismatched colors, or objects that defy physics.
- In a supervised learning model, a labelled dataset is given to the machine.
- A labelled dataset is the information which is tagged with identifiers of data.
- Discriminative Modelling is an approach in Machine Learning where the focus is on learning the boundary or decision boundary that separates different classes or categories directly from the data.
- Unsupervised Learning is a type of Machine Learning where the model is trained on input data without any corresponding output labels.
- In Generative Modelling there is no labelled dataset, and the model can generate structured data from the Random Noise dataset.
- A "random noise dataset" typically refers to a collection of data points or samples where each data point is generated randomly.
- Generative Artificial Intelligence also called Gen AI, refers to the algorithms that generate new data that resembles human-generated content, such as audio, code, images, text, simulations, and videos.
- Generative AI is trained with existing data and content, creating the potential for applications such as natural language processing, computer vision, the meta-verse, and speech synthesis.
- GANs are neural networks that work to produce fresh data.
- Variational Autoencoders (VAEs) produces fresh data, learn the distribution of the data and then sample from it.
- RNNs are a special class of neural networks that excel at handling sequential data, like music or text.
- Autoencoders are neural networks that have been trained to learn a compressed representation of data.

Exercise

Solved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. ..... are captured by cameras, they are visual representations depicting scenes, objects or people in exactly the same way as they exist in the real world.
   a. Real images ◯ b. AI generated images ◯
   c. Imaginative images ◯ d. Drawings ◯

2. ..... generated images can add modifications, enhancements and even entirely new imaginative details.
   a. Computer ◯ b. AI ◯
   c. Algorithm ◯ d. Real ◯

3. In Generative Modelling, the model can generate structured data from the ..... dataset.
   a. Random Noise ◯ b. labelled ◯
   c. unlabelled ◯ d. Temporary ◯

4. ..... refers to AI techniques that create realistic but fake videos and audios.
   a. AI ◯ b. Deepfakes ◯
   c. Copyright ◯ d. Plagiarism ◯

5. Creating portraits of non-existing people can be done using .....
   a. Paint ◯ b. GANs ◯
   c. VAEs ◯ d. RNNs ◯

6. In a supervised learning model, a ..... dataset is given to the machine.
   a. Unlabelled ◯ b. labelled ◯
   c. New ◯ d. Stored ◯

7. ..... can be used to generate large volumes of content quickly and efficiently.
   a. Supervised learning ◯ b. Unsupervised ◯
   c. Generative AI ◯ d. Deepfake ◯

8. Name the type of Generative AI used for creating new sentences that mimic the style of Shakespeare or generating dialogue for a chatbot.
   a. Real Data ◯ b. GANs ◯
   c. VAEs ◯ d. RNNs ◯

9. In generative AI, a ..... is the initial input or instruction given by the user to the AI model to guide it in generating the desired content.
   a. dataset ◯ b. prompt ◯
   c. data ◯ d. information ◯

10. Name the type of generative AI used for artistic image creation based on learned features from famous paintings.

a. Autoencoders ◯ b. GANs ◯

c. VAEs ◯ d. RNNs ◯

B. Fill in the blanks.

1. An AI generated image is created using AI .....
2. The misuse of AI to access or generate personal data poses significant risks to individual's .....
3. A ..... dataset is the information which is tagged with identifiers of data.
4. In ..... modelling there is no labelled dataset.
5. Gen AI is gaining popularity due to the fact that people can use ..... to prompt AI.
6. GAN stands for ..... .
7. If generative AI is trained on biased or incomplete data, the output may be similarly .....
8. ..... is a platform for creating, training, and deploying generative models.
9. ..... is a web-based tool where you can make and change pictures using advanced AI technology called Generative Adversatial Networks (GANs).
10. ..... is a generative multimodal AI model created by Google.

C. State whether these statements are true or false.

1. Generative AI is trained with new data and content. .....
2. In a text generation model, a prompt could be a sentence or a few keywords. .....
3. In a supervised learning model, an unlabelled dataset is given to the machine. .....
4. Generative AI can produce content even better than humans. .....
5. GNNs work by compressing the data into a lower-dimensional form (encoding) and then decompressing it back to its original form (decoding). .....
6. "The Next Rembrandt Project used in data analysis" is an example of Generative AI in Music. .....
7 The Discriminator in GANs helps to distinguish between real and generated data. .....
8. Misinformation created by generative AI does not have any serious social impacts. .....
9. Generative AI tools are not capable of personalising content for individual users. .....
10. Privacy concerns in generative AI include the risk of generating sensitive personal information. .....

D. Match the following:

1. Labelled Data	a. Cameras
2. Real images	b. AI algorithm
3. AI generated images	c. Generating human-like text responses
4. Conventional AI	d. Discriminative Modelling
5. ChatGPT	e. Generating code as response
	f. Banking

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. How can you tell if an image is real or AI?

Ans. Artificial intelligence shows inconsistencies if observed closely, although it tries to piece together its creations from the original work. Odd outlines to sharpen or smoothen the edges, stray pixels to cover inconsistency, and abnormal shapes can be easily seen, if an image is zoomed to the maximum, on each of its parts.

2. What are environmental impacts of generative AI?

Ans. a. Training and running large AI models require significant computational resources, contributing to high energy consumption and environmental impact.

b. Many devices are being exchanged due to outdated hardware leading to increase in e-waste.

3. List the important features of real images captured by camera.

Ans. Real images captured by cameras are visual representations depicting scenes, objects or people in exactly the same way as they exist in the real world. These images are created by humans or nature and are unaltered.

4. What is a supervised learning model?

Ans. In a supervised learning model a labelled dataset is given to the machine. A labelled dataset is the information which is tagged with identifiers of data. For example, clothes in a store are marked under various categories of clothing like Shirts, Trousers, Coats, etc. They are further labelled as per gender and size.

5. Explain the term deepfake.

Ans. The term "deepfake" combines "deep learning" and "fake," referring to AI techniques that create realistic but fake videos and audio. These AI-generated videos can mislead people by making it seem like someone said or did something they didn't, undermining trust in public figures and institutions.

6. What is the use of Generative Modelling?

Ans. Generative Modelling is a specific approach within Unsupervised Learning that focuses on understanding and modelling how the data is generated.

7. What is a random noise dataset?

Ans. A "random noise dataset" typically refers to a collection of data points or samples where each data point is generated randomly. There is unpredictable fluctuations and is disarranged data which makes it impossible to identify target patterns or relationships in it. This may result in decreased accuracy or reliability of the output, which the generative AI model takes care of.

8. What is Generative AI?

Ans. Generative Artificial Intelligence (AI) also called Gen AI, refers to the algorithms that generate new data that resembles human-generated content, such as audio, code, images, text, simulations, and videos.

9. Differentiate between Generative AI and Conventional AI in terms of goals.

Ans. Generative AI creates new content which mimics the original content. This content includes images, text, music, or other forms of media. Whereas conventional AI analyses, processes, and classifies data.

It basically works to improve the accuracy, precision, recall, and speed within the scope of the defined task.

10. Explain VAEs with example.

Ans. This is another class of generative models. In order to produce fresh data, VAEs learn the distribution of the data and then sample from it.

Example: Generation of new images similar to the given training set. For instance, a VAE trained on images of faces can generate new, realistic-looking faces.

B. Long answer type questions:

1. How does Discriminative Modelling contrast with Generative Modelling?

Ans. In Supervised Learning, Discriminative Modelling contrasts with Generative Modelling, where the goal is to model the joint probability distribution of both the input features and the output labels. Generative models can be used to generate new data points that resemble the training data, whereas discriminative models are primarily focused on classification or regression tasks.

2. What are the limitations of Gen AI?

Ans. Limitations of Using Generative AI

- **Data bias:** If generative AI is trained on biased or incomplete data, the output may be similarly biased or flawed. This can lead to inaccurate or problematic results in certain applications, such as in facial recognition or natural language processing.
- **Uncertainty:** Generative AI can produce unexpected and often unpredictable results, which can be both a benefit and a drawback.
- **Computational demands:** Generative AI requires significant computational resources to train and generate its output, which can be expensive and time-consuming.

3. What are Recurrent Neural Networks? List its important features with examples.

Ans.
- RNNs are a special class of neural networks that excel at handling sequential data, like music or text.
- They excel at tasks where the order of the data points is important, as they can remember previous inputs and use this information to influence current outputs.
- Example:

  It can generate novel text in the style of a specific author or genre, like creating new sentences that mimic the style of Shakespeare or generating dialogue for a chatbot.

4. Explain how generative AI is being used in Art.

Ans.
- Generative AI can create new artworks by learning styles from famous painters and generating novel pieces in similar styles.
- AI artists like "AI Portraits" and "DeepArt" have gained popularity for their ability to create visually stunning images.
- The Next Rembrandt project used data analysis and 3D printing to create a new painting in the style of Rembrandt.

5. Give three benefits of using generative AI.

Ans.
- **Creativity:** Generative AI can assist creatives in pushing the boundaries in making creative processes more efficient and personalised. This can be particularly valuable in fields such as art, design, and music.
- **Efficiency:** Generative AI can automate content creation processes, which can save time and reduce costs compared to traditional manual processes.
- **Personalisation:** Generative AI can be used to create personalised content for individual users based on their preferences and behaviours, such as customised product recommendations or personalised news articles.

C. Competency-based/Application-based questions:

Environmental Awareness

1. Hema is a content creator using a generative AI tool to produce images for a marketing campaign. How would she ensure the images generated by the AI looks original and not infringing on any existing copyrights? What steps would she take to verify the originality of the AI-generated content?

Ans. Hema can ensure the images generated by AI look original by utilizing AI tools designed to create unique content and by cross-referencing the outputs with existing image databases to check for similarities. She should employ copyright verification services to detect any potential infringements. Additionally, using licensed AI tools that guarantee original outputs can help. Finally, Hema should keep detailed records of the AI's generation process for accountability.

2. "OrStim" organization is using generative AI to create synthetic voices for virtual assistants. How would the company ensure the synthetic voices are natural and engaging while avoiding any potential misuse, such as generating misleading or harmful audio content? What guidelines should company establish for ethical use?

Ans. "OrStim" can ensure the synthetic voices are natural and engaging by training the AI on diverse and high-quality voice datasets, focusing on clarity and emotional expression. To avoid misuse, they should implement filters to detect and block harmful or misleading content generation.

Unsolved Questions

SECTION A (Objective Type Questions)

AI Quiz

A. Tick (✓) the correct option.

1. Based on the ..... data in Discriminative Modelling, the model is able to identify individual datasets.
   a. Unlabelled
   b. New
   c. labelled
   d. Named

2. ..... is unpredictable fluctuations and disarranged data which makes it impossible to identify target patterns or relationships in it.
   a. Deepfake
   b. Random Noise Dataset
   c. Gen AI
   d. Discriminative Modelling

3. Which of the following is not a generative AI tool?
   a. Galileo
   b. CodeGPT
   c. Magic AI
   d. Vista AI

4. ..... are neural networks that work to produce fresh data.
   a. Deepfake
   b. Gen AI
   c. GANs
   d. VAE

5. What is a primary advantage of using Autoencoders?
   a. Generating realistic videos
   b. Creating noise-free images
   c. Handling sequential data
   d. Generating new text

6. Which generative AI tool is designed to understand text, images, audio, video, and code?
   a. ChatGPT
   b. Runway ML
   c. Gemini
   d. Copilot

7. What potential negative impact on society is associated with generative AI?
   a. Enhanced creativity
   b. Job displacement
   c. Increased efficiency
   d. Personalisation of content

8. What does a "prompt" in generative AI refer to?
   a. The initial input or instruction given by the user to the AI model
   b. The output generated by the AI model
   c. The training dataset used by the AI model
   d. The algorithm used to train the AI model

9. Which technology introduced the concept of the world's first neural network?

a. Transformers ◯ b. Perceptron ◯

c. RNNs ◯ d. GAN ◯

B. Fill in the blanks.

1. An ..... learning approach works on an unlabelled dataset.
2. Generative AI could be used to create harmful technologies or weapons, posing significant ..... risks.
3. The full form of VAE is ..... .
4. The ethical consideration of ..... in generative AI involves the risk of generating biased or discriminatory content.
5. ..... can help in drug discovery by learning and generating molecular structures that have desirable properties.
6. A major risk associated with generative AI is the potential to create misinformation, such as ..... and deepfakes.
7. Recurrent Neural Networks (RNNs) are especially good at handling ..... data like text and music.
8. Generative AI refers to algorithms that create new data resembling ..... generated content.
9. The AI tool Artbreeder is used for blending and modifying ..... using GANs.
10. ..... provides a user-friendly interface for building and training various types of generative models, including GANs, VAEs, and image classifiers.

C. State whether these statements are true or false.

1. Generative AI can explore new design spaces and optimise systems. .....
2. Generative AI is a powerhouse when it comes to creating lots of content in a short time. .....
3. The generator in GANs evaluates the generated data to ensure it is realistic. .....
4. An important application of RNNs is predictive text input. .....
5. Both AE and VAE consist of an encoder and a decoder network. .....
6. Runway ML is a platform for creating and deploying Art work and is part of chatGPT. .....
7. Generative AI is not capable of composing new music or remixing existing pieces. .....
8. Variational Autoencoders (VAEs) are used to sample new data from learned data distributions. .....
9. Generative AI algorithms are primarily used to analyse data rather than create new data. .....
10. We have no setbacks using Generative AI. .....

D. Match the following:

1. Supervised Learning model	a. Arts
2. GAN	b. Initial input given to the AI model
3. Generative AI	c. Generator Network & Discriminator Network
4. Prompt	d. Creates noise-free images
5. VAEs	e. Labelled

SECTION B (Subjective Type Questions)

A. Short answer type questions:

1. Write one difference between Autoencoders (AEs) and Variational Autoencoders (VAEs).
2. Give two examples of VAEs.
3. When was Google DeepMind's AlphaGo introduced?
4. Write the names of any four generative AI tools.
5. Differentiate between generative AI and conventional AI in terms of applications.
6. Generative AI models need large training data. Elaborate.
7. Explain the limitations of using generative AI in terms of data bias.
8. Give two important feature of the popular tool - Artbreeder.
9. Give two ethical considerations of using generative AI.
10. "Generative AI is a threat to privacy and has data security risks." Explain in short.

B. Long answer type questions:

1. Explain in detail the use of generative AI in the field of music.
2. What are Autocoders? List its important features with examples.
3. Draw a diagram to represent Recurrent Neural Networks.
4. Write any two AI tools each for the following:
   a. Generative AI image generation tools
   b. Generative AI text generation tools
   c. Generative AI audio generation tools
5. Give one point to support how generative AI can be helpful in following fields:
   a. Architecture
   b. Coding
   c. Music
   d. Content Creation

C. Competency-based/Application-based questions:

Environmental Awareness

1. Sakshi has been assigned a homework essay on the topic, "The Impact of Climate Change on Coral Reefs." The essay requires Sakshi to research and explain various aspects of climate change, such as ocean acidification and rising sea temperatures, and their effects on coral reef ecosystems. His friend suggested using some text generation tool. List some guidelines for Sakshi to prevent misuse of generative AI and use it constructively.
2. How do you think generative AI can revolutionise the creative industry, such as art and fashion, by enabling the generation of unique and innovative designs?
3. Considering the ethical challenges associated with generative AI, what are your thoughts on establishing guidelines or regulations to ensure responsible use of these technologies? How can we balance the potential benefits and risks?
4. Find out ChatGPT vs Gemini vs Copilot on the basis of:
   - Parameter 1: Human-Like Response.
   - Parameter 2: Training Dataset and Underlying Technology.
   - Parameter 3: Authenticity of Response.
   - Parameter 4: Access to the Internet.

- Parameter 5: User Friendliness and Interface.
- Parameter 6: Text Processing: Summarisation, Paragraph Writing, Etc.
- Parameter 7: Charges and Price

5. Shinjini is a university student majoring in Architecture. She is working on a project for her digital art class where she need to create a series of concept sketches for designing a fictional bridge. However, she is struggling to come up with creative ideas and wants to explore different concepts quickly where generative AI can be used.

   List how she can responsibly use generative AI to design her project.

Ai In Life

Generative AI enables media and entertainment companies to streamline content creation processes. These AI models can generate original scripts, articles, and even music compositions, freeing up human creators to focus on more complex and creative tasks. Using one of these tools, prepare a skit for your morning class assembly related to the topic chosen.

Deep Thinking

How might the widespread adoption of generative AI impact job markets, particularly in creative industries such as art, writing, and music?

Ai Lab

Experiential Learning

1. Visit the shared link and determine which one is a real image or an AI generated image.

   https://britannicaeducation.com/blog/quiz-real-or-ai/

   State 5 points you have noted for easy identification of AI generated image.

   _____

   _____

   _____

   _____

   _____

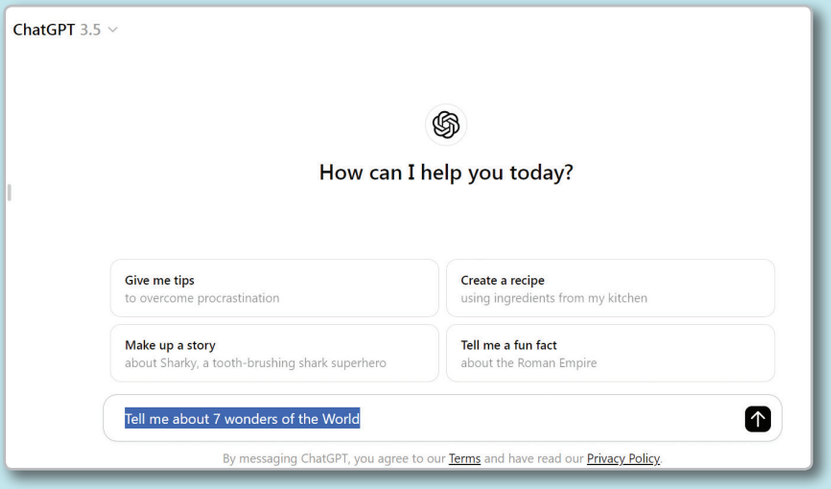
   _____

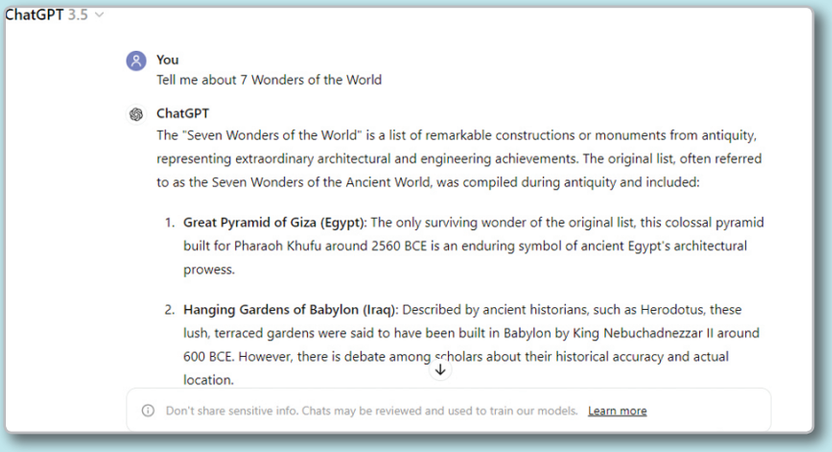
   _____

2. Explore the ChatGPT tool

   a. Go to https://chat.openai.com/

   b. Give the prompts listed below and see the answer that you get.





3. Enjoy Exploring!

Answers

AI Quiz Section A (Objective Type Questions)

A. 1. a 2. b 3. a 4. b 5. b 6. b 7. c 8. d 9. b 10. a

B. 1. algorithms 2. privacy and security 3. labelled 4. Generative 5. Natural Language
6. GANs (Generative Adversarial Networks) 7. biased 8. Runway ML
9. Artbreeder 10. Gemini

C. 1. False 2. True 3. False 4. True 5. False
6. False 7. True 8. False 9. False 10. True

D. 1. d 2. a 3. b 4. f 5. c

Ai Ready 4

1. Can we consider the data acquired from ChatGPT to be authentic? Justify your answer.

2. What sets Gemini apart from ChatGPT?

3. List any two jobs in the field of Art that will be soon replaced by AI.

4. Expand the term RNN. Give two applications of RNN.