Supplement
2024-2025

Code 417
Under NSQF

artificial intelligence

A Textbook of Class IX

Part A: Employability Skills
Part B: Vocational Skills

SUMITA ARORA

DHANPAT RAI & Co.

Code : 417

Under NSQF

ARTIFICIAL INTELLIGENCE

[Class IX]

Supplement 2024-25

PART A

EMPLOYABILITY SKILLS

PART B

VOCATIONAL SKILLS

SUMITA ARORA

DHANPAT RAI & Co. (Pvt.) Ltd.

EDUCATIONAL AND TECHNICAL PUBLISHERS

Published by : GAGAN KAPUR
Dhanpat Rai & Co. (P) Ltd., Delhi

Regd. Office : 4576/15, Agarwal Road
Darya Ganj, New Delhi-110002
Phone : 2324 7736, 37, 38, dhanpatrai@gmail.com

EDITION : 2024

PRICE : ₹75/-

Typesetting by : North Delhi Computer Convention, Delhi–110009,
ndcc.in@gmail.com
Printed at : Natraj Offset, Delhi

Syllabus

The unit-wise distribution of periods and marks for Class IX is as follows :

Total Marks : 100 (Theory 50 + Practical 50)

Units		No. of Periods for Theory and Practical	Max. Marks for Theory and Practical 100
Part A	Employability Skills		
	Unit 1 : Communication Skills-I	10	2
	Unit 2 : Self-management Skills-I	10	2
	Unit 3 : Basic 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 = 55	10
	Unit 2 : Data Literacy	22 + 28 = 50	10
	Unit 3 : Math for AI (Statistics & Probability)	12 + 13 = 25	07
	Unit 4 : Introduction to Generative AI	08 + 12 = 20	05
	Unit 4 : Introduction to Python	01 + 09 = 10	08
	Total	160	**40**
Part C	Practical Work – **PYTHON** 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 (Any one has to be done)		15
	Total		**15**
	Grand Total	**210**	**100**

Unit Contents

Part A : Employability Skills (Common for all Vocational Courses)

S.No.	Units	Duration in Periods
1.	Unit 1 : Communication Skills – I	10
2.	Unit 2 : Self-management Skills – I	10
3.	Unit 3 : Basic Information and Communication Technology Skills – I	10
4.	Unit 4 : Entrepreneurial Skills – I	15
5.	Unit 5 : Green Skills – I	05
	Total	50

Unit 1 : Communication Skills – I

S. No.	Learning Outcome	Theory	Practical
1.	Demonstrate knowledge of various methods of communication.	❖ Methods of communication. ❍ Verbal. ❍ Non-verbal. ❍ Visual.	❖ Writing pros and cons of written, verbal and non-verbal communication. ❖ Listing do's and don't for avoiding common body language mistakes.
2.	Identify elements of communication cycle.	❖ Meaning of communication ❖ Importance of communication skills ❖ Elements of communication cycle : (*i*) sender, (*ii*) ideas, (*iii*) encoding, (*iv*) communication channel, (*v*) receiver, (*vi*) decoding, and (*vii*) feedback	❖ Draw a diagram of communication cycle ❖ Role plays on communication process related to the sector/job role.
3.	Identify the factors affecting our perspectives in communication	❖ Perspectives in communication ❖ Factors affecting perspectives in communication ❍ Visual perception. ❍ Language. ❍ Past experience. ❍ Prejudices. ❍ Feelings. ❍ Environment.	❖ Group discussion on factors affecting perspectives in communication. ❖ Sharing of experiences on factors affecting perspectives. ❖ Sharing experiences on factors affecting communication at workplace.
4.	Demonstrate the knowledge of basic writing skills	❖ Writing skills related to the following : ❍ Phrases ❍ Kinds of sentences ❍ Parts of sentence ❍ Parts of speech ❍ Use of articles ❍ Construction of paragraph	❖ Demonstrate and practice of writing sentences and paragraphs on topics related to the subject.

Unit 2 : Self-Management Skills – I

S. No.	Learning Outcome	Theory	Practical
1.	Describe the meaning and importance of self-management.	❖ Meaning of self-management. ❖ Positive results of self-management. ❖ Self-management skills.	❖ Identification of self-management skills ❖ Strength and weakness analysis
2.	Identify the factors that helps in building self-confidence.	❖ Factors that help in building self-confidence – social, cultural, and physical factors. ❖ 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.	❖ Role play exercises on building self-confidence. ❖ Use of positive metaphors/words. ❖ Positive stroking on wakeup and before going bed. ❖ Helping others and working for community.

Unit 3 : Basic ICT Skills – I

S. No.	Learning Outcome	Theory	Practical
1.	Describe the role of ICT in day-to-day life and workplace.	❖ Introduction to ICT ❖ Role and importance of ICT in personal life and at workplace ❖ ICT in our daily life (examples) ❖ ICT tools – Mobile, tab, radio, TV, email, etc.	❖ Discussion on the role and importance of ICT in personal life and at workplace. ❖ Preparing posters/collages for showing the role of ICT at workplace.
2.	Identify components of basic computer system and their functions.	❖ Computer system – Central Processing Unit (CPU), memory, motherboard, storage devices. ❖ Hardware and software of a computer system. ❖ Role and functions of Random Access Memory (RAM) and Read Only Memory (ROM). ❖ Role and functions of Central Processing Unit. ❖ Procedure for starting and shutting down a computer.	❖ Connecting the cables and peripherals to the Central Processing Unit. ❖ Starting and shutting down a computer. ❖ Group discussion on the various aspects of hardware and software.
3.	Demonstrate use of various components and peripherals of computer system	❖ Peripheral devices and their uses – mouse, keyboard, scanner, webcam, etc. of a computer system.	❖ Identification of various parts and peripherals of a computer. ❖ Demonstration and practice on the use of mouse. ❖ Demonstration and practice on the use of keyboard. ❖ Demonstration of the uses of printers, webcams, scanner and other peripheral devices. ❖ Drawing diagram of computer system and labelling it.
4.	Demonstrate basic computer skills.	❖ Primary operations on a computer system – input, process, storage, output, communication networking, etc.	❖ Identification of the various input and output units and explanation of their purposes.

Unit 4 : Entrepreneurial Skills – I

S. No.	Learning Outcome	Theory	Practical
1.	Identify various types of business activities	❖ Types of businesses– service, manufacturing, hybrid. ❖ Types of businesses found in our community, Business activity around us.	❖ Prepare posters of business activities found in cities/ villages, using pictures. ❖ Discuss the various types of activities, generally adopted by small businesses in a local community. ❖ Best out of waste. ❖ Costing of the product made out of waste. ❖ Selling of items made from waste materials. ❖ Prepare list of businesses that provides goods and services in exchange for money.
2.	Demonstrate the knowledge of distinguishing characteristics of entrepreneurship	❖ Meaning of entrepre-neurship development. ❖ Distinguishing characteristics of entrepreneurship. ❖ Role and rewards of entrepreneurship.	❖ Prepare charts showing advantages of entrepreneurship over wages. ❖ Group discussions on role and features of entrepreneurship. ❖ Lectures/presentations by entrepreneurs on their experiences and success stories. ❖ Identify core skills of successful entrepreneur.

Unit 5 : Green Skills – I

S. No.	Learning Outcome	Theory	Practical
1.	Demonstrate the knowledge of the factors influencing natural resource conservation.	❖ Introduction to environment. ❖ Relationship between society and environment, ecosystem and factors causing imbalance. ❖ Natural resource conservation. ❖ Environment protection and conservation.	❖ Group discussion on hazards of deteriorating environment. ❖ Prepare posters showing environment conservation. ❖ Discussion on various factors that influence our environment.
2.	Describe the importance of green economy and green skills.	❖ Definition of green economy ❖ Importance of green economy	❖ Discussion on the benefits of green skills and importance of green economy. ❖ Prepare a poster showing the impor-tance of green economy with the help of newspaper/magazine cuttings.

PART B : VOCATIONAL SKILLS

S.No.	Units
1.	Unit 1 : AI Reflection, Project Cycle and Ethics
2.	Unit 2 : Data Literacy
3.	Unit 3 : Math for AI (Statistics & Probability)
4.	Unit 4 : Introduction to Generative AI
5.	Unit 5 : Introduction to Python

Unit 1 : AI Reflection, Project Cycle and Ethics

Sub-unit	*Learning Outcome*	*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 ❖ Modeling ❖ 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/soceity ❖ 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.

Sub-unit	Learning Outcome	Session/Activity/Practical
	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 modeling (Rule-based & Learning-based)	**Session** : Modeling ❖ Introduction to modeling 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.

Sub-unit	Learning Outcome	Session/Activity/Practical
	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 Outcome	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 Literac y ❖ 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

Sub-unit	Learning Outcome	Session/Activity/Practical
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 Outcome	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 and 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,
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 Outcome	*Session/Activity/Practical*
Students will be able to define Generative AI and classify different kinds. ❖ Students will be able to explain how Generative AI works and recognize how it learns. ❖ Applying Generative AI tools to create content. ❖ Understanding the ethical considerations of using Generative AI.	**Recommended Activity :** **Activity :** ❖ Guess the Real Image vs. the AI-generated image **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
	Recommended Activities : ❖ Hands-on Activity : GAN Paint ❖ Generative AI tools
	Session : ❖ Ethical considerations of using Generative AI

Unit 5 : Introduction to Python

Learning Outcome	*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 ❖ Students go through lessons on conditional and iterative statements (if, for and while) ❖ Students will try some basic problem-solving exercises using conditional and iterative statements on Python Compiler. **Practical :** Python Lists ❖ Students go through lessons on Python Lists (Simple operations using list) ❖ Students will try some basic problem-solving exercises using lists on Python Compiler.

PART C : PRACTICAL WORK

UNIT 5 : Introduction to Python : Suggested Program List

PRINT	❖ To print personal information like Name, Father's Name, Class, School Name. ❖ To print the following patterns using multiple print commands  ❖ To find square of number 7 ❖ To find the sum of two numbers 15 and 20. ❖ To convert length given in kilometers into meters. ❖ To print the table of 5 up to five terms. ❖ To calculate Simple Interest if the principle_amount = 2000 rate_of_interest = 4.5 time = 10
INPUT	❖ To calculate Area and Perimeter of a rectangle ❖ To calculate Area of a triangle with Base and Height ❖ To calculating average marks of 3 subjects ❖ To calculate discounted amount with discount % ❖ To calculate Surface Area and Volume of a Cuboid
LIST	❖ Create a list in Python of children selected for science quiz with following names- Arjun, Sonakshi, Vikram, Sandhya, Sonal, Isha, Kartik ❖ Perform the following tasks on the list in sequence- ○ Print the whole list ○ Delete the name "Vikram" from the list ○ Add the name "Jay" at the end ○ Remove the item which is at the second position. ❖ Create a list num = [23, 12, 5, 9, 65, 44] ○ Print the length of the list ○ Print the elements from second to fourth position using positive indexing ○ Print the elements from position third to fifth using negative indexing ❖ Create a list of first 10 even numbers, add 1 to each list item and print the final list. ❖ Create a list List_1 = [10, 20, 30, 40]. Add the elements [14, 15, 12] using extend function. Now sort the final list in ascending order and print it.
IF, FOR, WHILE	❖ Program to check if a person can vote ❖ To check the grade of a student ❖ Input a number and check if the number is positive, negative or zero and display an appropriate message ❖ To print first 10 natural numbers ❖ To print first 10 even numbers ❖ To print odd numbers from 1 to n ❖ To print sum of first 10 natural numbers ❖ Program to find the sum of all numbers stored in a list
Important Links	❖ https://cbseacademic.nic.in/web_material/Curriculum21/publication/secondary/Python_Content_Manual.pdf ❖ https://drive.google.com/drive/folders/1qRAckDculA5i164OUFDlilxb8mT65MMb

PART D : Project Work / Field Visit / Student Portfolio (relate it to Sustainable Development Goals)

Suggested Projects/ Field Visit / Portfolio (Any one has to be done)

Suggested Projects	1. Create an AI Model using tools like- ❖ Teachable Machine (https://teachablemachine.withgoogle.com/) ❖ Machine Learning For Kids (https://machinelearningforkids.co.uk/) 2. Choose an issue that pertains to the objectives of sustainable development and carry out the actions listed below. ❖ To understand more about the problem identified, create a 4Ws problem canvas. ❖ Identify the data features and create a system map to understand relationship between them ❖ Visualize the data collected graphically (Spreadsheet software to be used store and visualize the data) ❖ Suggest an AI enabled solution to it (Prototype/Research Work)
Suggested Field Visit	Visit to an industry or IT company or any other place that is creating or using AI applications and present the report for the same. Visit can be in physical or virtual mode.
Suggested Student Portfolio	Maintaining a record of all AI activities and projects (For Example, Letter to Futureself, Smart Home Floor Plan, Future Job Advertisement, Research Work on AI for SDGs and AI in Different Sectors, 4Ws canvas, System Map). (Minimum 5 Activities)

CONTENTS

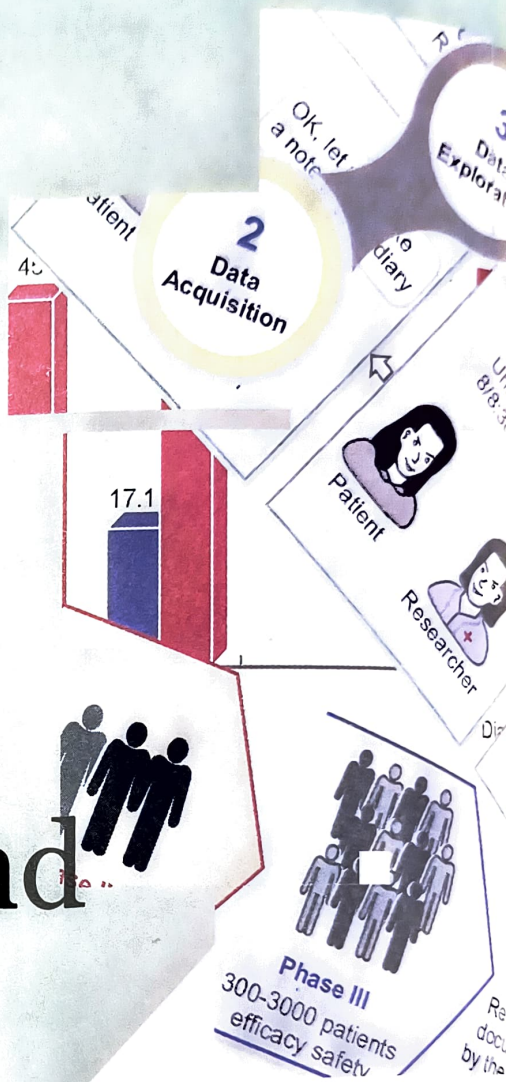
PART B : Vocational Skills

VOCATIONAL SKILLS

PART B

I

AI Reflection, Project Cycle and Ethics



In This Unit

Session 1	Introduction to AI (Old Unit 1)
Session 2	Introduction to AI Project Cycle (Old Unit 2)
Session 3	Problem Scoping – Set Goals (Old Unit 2)
Session 4	Problem Scoping – Identify Stakeholders, Actions and Ethical Issues (Old Unit 2)
Session 5	Problem Scoping – Explore and Understand Data (Old Unit 2)
Session 6	Data Acquisition (Old Unit 2)
Session 7	System Maps for the Problem Depiction (Old Unit 2)
Session 8	Data Visualisation (Old Unit 2)
Session 9	Modelling, Evaluation and Deployment
Session 10	AI Ethics (Old Unit 1)

ROADMAP AS PER CHANGES IN UNIT I 2024 SYLLABUS

SESSION 9

Modelling, Evaluation and Deployment

- Types of AI Technologies
- AI Modelling
- Evaluation Techniques
- Deployment of AI Projects

9.1 INTRODUCTION

'Artificial Intelligence' mimics the real intelligence. The real intelligence uses everything from data to models to rules to decisions and so on. In this session, you will learn how Artificial Intelligence uses *modelling* for making decisions.

Artificial Intelligence

Artificial Intelligence (AI) refers to the ability of modern machines and computers to mimic human intelligence. AI machines work through data, algorithms and models.

9.2 TYPES OF AI TECHNOLOGIES

There are many AI technologies and systems. But most used of these are **Machine Learning** and **Deep Learning.**

- **Machine Learning (ML)** is a branch of AI that enables machines to automatically learn and improve at tasks with experience and by the use of data. ML based machines undergo lots of repetitions of taking data and testing it; these then keep track of when things went wrong or right, and keep improving their results.

Machine Learning

Machine Learning (ML) is a branch of AI that enables machines to automatically learn and improve at tasks with experience and by the use of data.

The ML systems can automatically learn and improve without explicitly being programmed. The recommendation systems on music and video streaming services are examples of ML. Machine learning finds patterns in data and uses them to make predictions.

- **Deep Learning (DL)** is a subset of machine learning where learning takes place through examples. Deep Learning computer-models filter the input data using layers and rules-based algorithms to predict and classify information. Tasks like speech and image recognition are performed through deep learning systems. Driver-less cars are being developed using deep learning technologies.

Deep Learning

Deep Learning (DL) is a subset of machine learning where learning takes place through examples by filtering the input data using layers and rules-based algorithms to predict and classify information.

Most of the deep learning methods use neural network architectures, so they are often also referred to as deep neural networks.

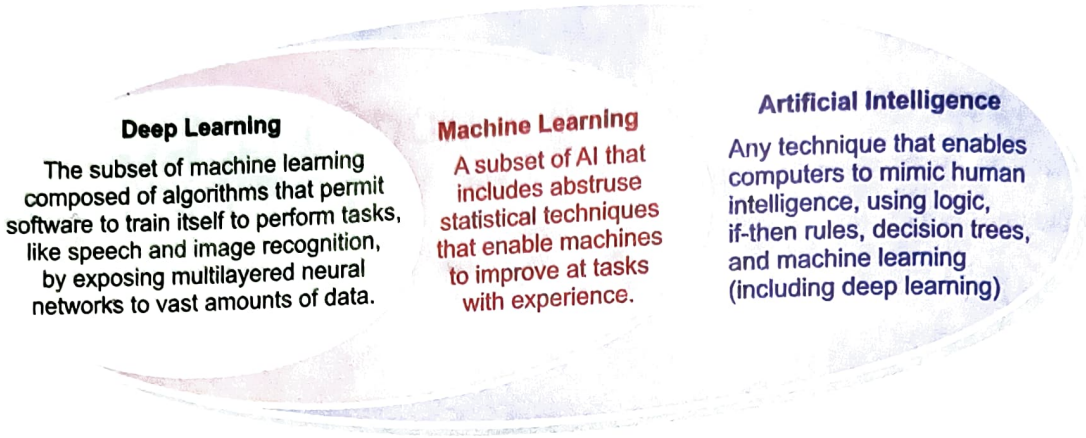


Figure 9.1 AI and its Subsets

Note

Deep Learning is the key technology behind driver-less cars, that enables them to recognize a stop sign and to distinguish between a pedestrian and lamp post.

PART B

9.2.1 Categories of AI

Artificial intelligence is developed through various ways. It can either be *Data driven (also called learning-based AI) or model driven (also called Rule based AI)*. Let us quickly discuss these two types of developing AI models.

1. Model-driven AI (Rule based AI)

It is a system which derives decisions about the right answer through explicit representation and rules. For example, in such a system, a cat would be explicitly represented as a four-legged animal, with two eyes, a nose and a mouth that is furry (except when not) and that is relatively small (except when not), etc.

Note

The rule based AI is used when we have known or labelled dataset.

In this, an AI system is developed using the predefined *labels, rules, patterns or relationships* as given by the developer in the algorithm. Thus, the machine would follow the algorithm's rules or instructions and perform the given tasks or take the decisions accordingly.

Model Driven AI

Rule Based AI (Model driven AI) refers to the branch of AI where models are developed using the algorithms having pre-defined labels, rules, patterns and relationships.

2. Data driven AI (Learning based AI/Machine Learning Approach)

It is a system in which lots of data is shown and questions/answers are asked in order to **train** the system about the right answer. For example, to train a system about recognising

cats, it would be shown lots and lots of images of cats, and other animals and letting it know when it "guessed" it correctly or not. After many (millions) of training cycles it will "learn" to get it increasingly right.

Data Driven AI

The **Learning based AI (Data Driven AI)** refers to that branch of AI where models are trained to learn by inputting them tons of data. Here there are no patterns, rules and relationships predefined by the developer, rather machine learns with each new input and comes up with own algorithm.

Using this style when an AI system is developed then the relationship or patterns in data are not defined by the developer. Rather, random data is repeatedly input to the machine. The machine analyses each input and tries to figure out patterns and trends out of the input collectively. This style is appropriate when the data to be processed is unlabelled and too random to fit into a frame of common rules and patterns.

Note

The *learning based approach* is used when data is unknown or random or unlabelled.

Both **Machine Learning (ML)** and **Deep Learning (DL)** fall under the category of learning-based AI.

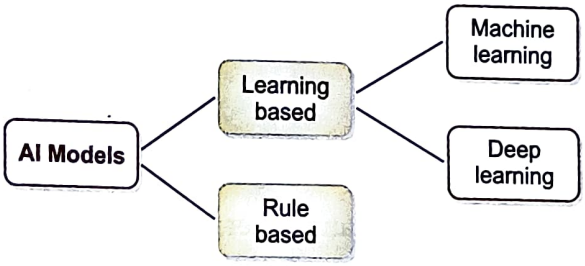


Figure 9.2 Types of AI models

9.3 AI MODELLING

The process of developing AI based algorithms (AI model), is called *AI modelling*. An ***AI model*** *is a program or algorithm that utilizes a set of data that enables it to recognize certain patterns*. This allows it to reach a conclusion or make a prediction when provided with sufficient information, *i.e.*, give intelligent outcomes. In short, AI modelling refers to the process of writing codes to make intelligent AI algorithms to produce intelligent outcomes. An AI model can be developed using any of the AI styles we discussed above.

AI Modelling

The process of developing an AI algorithm (AI model) is called **AI modelling** where code is developed to produce intelligent outcomes using a set of data.

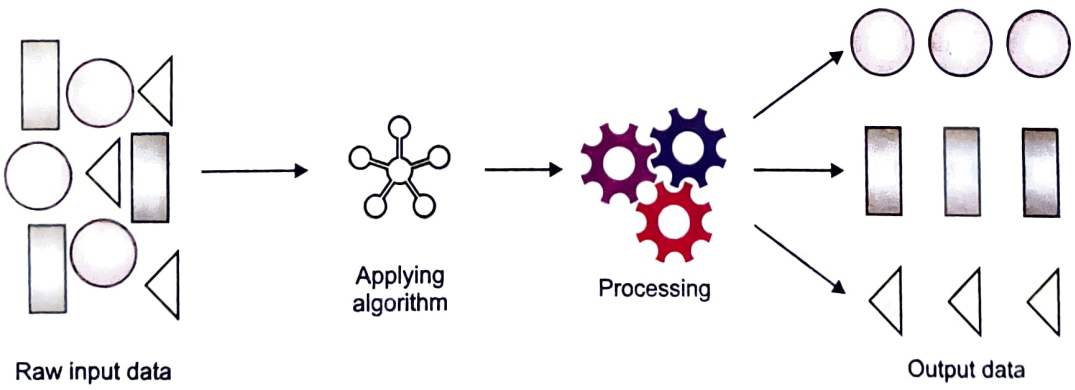


Figure 9.3 AI modelling

There are many types of AI models which are popular today :

(*i*) ***Linear Regression.*** Linear regression models use statistics and supervised learning. These are widely used in banking, healthcare, insurance, etc.

(*ii*) ***Deep Neural Networks (DNN).*** DNNs are AI models based on interconnected units known as *artificial neurons* (inspired by the neural network of the human brain). DNN models take decisions mimicking human brains. The DNNs find applications in *speech recognition, image recognition, natural language processing* and so on.

(*iii*) ***Logistic Regression.*** This is similar to the Linear regression model, except that it is only used in solving classification-based problems.

(*iv*) ***Decision Trees (DT).*** The Decision Tree (DT) model is used to arrive at a conclusion based on the data from past decisions. In this model, the data is divided into smaller portions resembling the structure of a tree (hence the name). This model can be applied for both regression and classification problems.

Note

The decision tree is the most common and basic model of AI.

There are many other AI models too, but talking about them is not possible here. Most AI models broadly belong to *three* categories of AI models : ***Regression***, ***Classification*** and ***Clustering***.

PART B

9.4 EVALUATION TECHNIQUES

You have learnt about AI modelling. AI models also need to be evaluated to see if they're making the right decisions.

There are multiple techniques used to evaluate AI models. Let us talk about some simple evaluation techniques below.

(*i*) True Positive (TP). *True positive* refers to an instance for which both *predicted value* of the AI model and *actual value* are positive. For example, while testing a patient for Covid, if the test also produced the result (**predicted value**) as *positive* and the actual result (**actual value**) is also *positive*, it is ***True positive.***

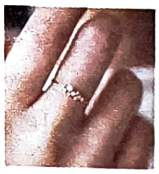
(*ii*) True Negative (TN). *True negative* refers to an instance for which both *predicted value* of the AI model and *actual value* are negative. For example, while testing a patient for Covid, if the test also produced the result (**predicted value**) as *negative* and the actual result (**actual value**) is also *negative*, it is ***True negative.***

(*iii*) False Positive (FP) **(also called Type I Error).** *False positive* refers to an instance for which *predicted value* of an AI model is positive but *actual value* is negative. For example, while testing a patient for Covid, if the test produced the result (**predicted value**) as *positive* and the actual result (**actual value**) is *negative*, it is *False positive.*

(iv) False Negative (FN) **(also called Type II Error).** False negative refers to an instance for which *predicted value* of an AI model is negative but *actual value* is positive. For example, while testing a patient for Covid, if the test produced the result (**predicted value**) as *negative* and the actual result (**actual value**) is *positive*, it is ***False negative.***

The first two terms signify when both the actual and predicted values match. In the 3rd and 4th terms listed above, the latter part (*Positive/Negative*) represents the predicted value and the *False* means the actual value is opposite of the predicted value.

For example, if looking at picture, the AI model has to identify if it is the picture of vegetable Lady finger, then *True Positive/Negative* and *False Positive/Negative* will be identified as :

True Positive  **Result.** Lady finger Vegetable	False Negative  **Result.** NOT Lady finger Vegetable
False Positive  **Result.** Lady finger Vegetable	True Negative  **Result.** NOT Lady finger Vegetable

9.5 DEPLOYMENT OF AI PROJECTS

Deployment is the mechanism through which the developed and trained AI model is made available to users for actual use. AI models can be deployed in multiple ways. For each deployment method, one should consider the following factors :

- **Accessibility.** How easily can the AI system be accessed, especially in remote or underserved areas ?
- **Infrastructure Requirements.** What are the hardware, software, and connectivity needs for each method ?
- **Data Privacy and Security.** How well does each method protect sensitive patient data and ensure compliance with regulations ?
- **Cost and Scalability.** What are the upfront and ongoing costs, and how easily can the solution scale to meet demand ?

- **Performance and Reliability.** How accurate and consistent is the AI model's performance across different deployment methods ?

Based on these factors, some common deployment techniques[1] are :

Deployment

Deployment is the mechanism through which the developed and trained AI model is made available to users for actual use.

(i) On Device/Premise Deployment

This is the case when the actual device at the point of use (*e.g.*, your phone or smartwatch) directly houses or stores the AI model. It makes it very easy to use it, because it can work even when you're not connected to the Internet.

(ii) Cloud Deployment

AI models can occasionally be so large and complex that they require a lot of processing power to execute. In such case, deploying them on the device is not feasible. So, these are deployed on the cloud. As you must be aware of that a **cloud** means the storage, databases and processing systems available on the Internet servers. There are many clouds, some of these are by companies like *Microsoft*, *Google*, and *Amazon*.

Note

A **cloud** means the storage, databases and processing systems available on the Internet servers.

(iii) Edge Deployment

In this deployment, the deployment happens on edge devices. Edge devices, are computers or intelligent devices that enable real-time raw data collection and processing for quicker, more accurate analysis, e.g., smart sensors of IoT.

You must have seen smart cameras that can spot and differentiate a specific entity from their surroundings? (*e.g.*, AI cameras on traffic lights can differentiate vehicles and people from surroundings, or you must have seen or heard about the robots that move around without crashing ?)

Edge devices

Edge Devices, are computers or intelligent devices that enable real-time raw data collection and processing for quicker, more accurate analysis.

They're like little brains in the camera or robot itself. They are able to efficiently work as they have functionality embedded in them and they need not refer to connect to a server or a faraway computer. This is example of edge deployment where the AI model is deployed at the point of action in away so that it can function independently.

Hybrid deployment involves the use of more than one deployment technique.

1. There are some advance deployment techniques like *Containerization*, *Serverless Deployment*, *Continuous Deployment* etc. But we are not going in details of these as it beyond the scope of discussion here.

Practical Case Study **Preventable Blindness**

Many eye diseases, if detected early, can be treated to prevent vision loss or blindness. However, access to skilled eye care professionals is limited, in remote or underserved areas.

In this case study, you will explore how AI can be deployed to address the challenge of preventable blindness, considering different deployment methods and their implications.

Scenario : *You are part of a team developing an AI-powered solution to detect and diagnose vision-related conditions in remote or underserved communities.*

Task : *Discuss various deployment methods for this AI model and decide upon the best suited one.*

1. Cloud Deployment

In this method, the AI model would be hosted on a cloud platform, and healthcare providers or patients could upload retinal images securely for analysis.

Pros	Cons
Easy to update and maintain the AI model centrally.	Requires a stable internet connection, which may be challenging in remote areas.
Accessible from anywhere with an internet connection.	Data privacy and security concerns with transmitting sensitive medical data.
Cost-effective, as infrastructure costs are shared.	

2. On-premise Deployment

With this approach, the AI model would be installed locally at healthcare facilities or clinics, running on their own servers or computers.

Pros	Cons
No need for an internet connection, suitable for remote areas.	Higher upfront costs for each location to acquire the necessary hardware.
Data remains on-site, addressing privacy and security concerns.	Maintenance and updates become more complex across multiple sites.
Faster analysis, as data doesn't need to be transmitted.	Potential for inconsistent performance if hardware specifications vary.

3. Edge Device Deployment

In this method, the AI model would be deployed on edge devices like *smartphones, tablets, or specialized medical devices* with built-in cameras.

Pros	Cons
Highly portable and accessible, even in remote areas.	Limited processing power on edge devices may impact model performance.
No need for an internet connection or specialized hardware.	Battery life and storage constraints on mobile devices.
Immediate analysis and feedback for patients.	Securing and updating models on numerous devices can be challenging.

Factors to Consider

When evaluating these deployment methods, consider factors like: *Accessibility and reach in target communities; Cost and resource requirements; Data privacy and security concerns; Integration with existing systems or infrastructure;* and *Scalability and potential for future expansion.*

Considering these factors, **hybrid deployment** would be best suited that may combine the benefits of these while keeping a tab on costs.

Check Point

1. Most used AI technology categories are _____ .
   (a) Machine learning (b) Deep learning
   (c) Flat learning (d) Device learning
2. _____ AI enables machines to automatically learn and improve at tasks with experience, by the use of data.
   (a) Machine learning (b) Deep learning
   (c) Flat learning (d) Device learning
3. _____ AI enables learning through examples and filtering of input data using layers and rule-based algorithms.
   (a) Machine learning (b) Deep learning
   (c) Flat learning (d) Device learning
4. Deep Learning is a subset of Machine learning.
   (a) True (b) False
5. Name the key technology behind driver-less cars, that enables them to recognize and distinguish between signs and people.
   (a) Machine learning (b) Deep learning
   (c) Flat learning (d) Device learning

PART B

6. Models developed using the algorithms having pre-defined labels, rules, patterns and relationships are termed as _____ _____ AI.
   (a) Rule based (b) Data driven
   (c) Learning based (d) Model driven
7. Models developed by inputting them tons of data without any pre-set rules, patterns or relationships are termed as _____ _____ AI.
   (a) Rule based (b) Data driven
   (c) Learning based (d) Model driven
8. A predictive AI model that reaches to an outcome based on a branching series of decision-making questions is _____ model.
   (a) Training model (b) Testing model
   (c) Decision tree model (d) None of these
9. _____ is a technique that categorises predicted values of an AI model and the actual result values.
   (a) Accuracy (b) Precision
   (c) Recall (d) Confusion Matrix
10. When for an AI model, both the predicted and actual values are positive, it is known as _____ .
    (a) True Positive (b) False Positive
    (c) True Negative (d) False Negative
11. When for an AI model, both the predicted and actual values are negative, it is known as _____ .
    (a) True Positive (b) False Positive
    (c) True Negative (d) False Negative

12. When for an AI model, the predicted value is positive and actual value is negative, it is known as _____ .
    (*a*) True Positive
    (*b*) False Positive
    (*c*) True Negative
    (*d*) False Negative
13. When for an AI model, the predicted value is negative and actual value is positive, it is known as _____ .
    (*a*) True Positive
    (*b*) False Positive
    (*c*) True Negative
    (*d*) False Negative
14. False Positive is also known as _____ .
    (*a*) Error
    (*b*) Type I error
    (*c*) Type II error
    (*d*) Mismatch error
15. False Negative is also known as _____ .
    (*a*) Error
    (*b*) Type I error
    (*c*) Type II error
    (*d*) Mismatch error
16. A set of data that is used to develop a model to train a computer to perform a task is known as :
    (*a*) Training data
    (*b*) Data set
    (*c*) Testing data
    (*d*) Data model
17. A set of data that is used to test the outcome of developed AI model is known as :
    (*a*) Training data
    (*b*) Data set
    (*c*) Testing data
    (*d*) Data model
18. Match the column A with column B

Column A	Column B
(*i*) Problem Scoping	(*A*) implements a suitable model that matches the requirements
(*ii*) Data Acquisition	(*B*) interprets some useful information out of data
(*iii*) Data Exploration	(*C*) collects data from reliable sources
(*iv*) Modelling	(*D*) finalises the aim of the AI project

    (*a*) (*i*) – (*B*) ; (*ii*) – (*A*) ; (*iii*) – (*D*) ; (*iv*) – (*C*)
    (*b*) (*i*) – (*D*) ; (*ii*) – (*B*); (*iii*) – (*C*) ; (*iv*) – (*A*)
    (*c*) (*i*) – (*D*) ; (*ii*) – (*C*) ; (*iii*) – (*B*) ; (*iv*) – (*A*)
    (*d*) (*i*) – (*C*) ; (*ii*) – (*D*); (*iii*) – (*A*) ; (*iv*) – (*B*) [CBSE Sample Paper Term I, 2021-22]
19. ***Statement 1.*** The data with which the machine can be trained is the testing data
    Statement 2. The data with which the model is evaluated is the training data
    (*a*) Statement 1 is correct and Statement 2 is incorrect
    (*b*) Statement 2 is correct and Statement 1 is incorrect
    (*c*) Both Statement 1 and Statement 2 are correct
    (*d*) Both Statement 1 and Statement 2 are incorrect [CBSE Sample Paper Term I, 2021-22]

Competency Based Questions

20. An event planner company, '***Wow Events***', is planning for the party for which they have been hired. The 'Wow Events' company is known for its meticulous planning. For this party, they have a

venue which has bigger outdoor area and a smaller indoor area. Keeping in mind the weather conditions, the food stalls and other things, they have to tell the organisers whether the party can be organised for smaller number of guests (indoor) or larger number of guests (outdoor). Which of the following tools would be useful in deciding this?

(*a*) Tabular data
(*b*) Decision Trees
(*c*) Training data
(*d*) Testing data

21. **Aani** is a brilliant student of an engineering college. For an assignment project, she is developing a robot namely ***Ginko***, which can remove some objects if found on a way. However, only some and not all objects are to be removed, *e.g.*, the robot has to remove the round shaped objects from the way and turn to right when square shaped objects are found. So, Aani has gathered a lot of images of all types of objects, showing different views and taken from various angles. She has collected about 32 thousand images so far. She wants to store these images in a database. What type of data is Aani collecting ?

(*a*) Tabular data
(*b*) Robotic database
(*c*) Training data
(*d*) Testing data

22. After successfully working with the robot Ginko and its training, now Aani wants to test how much has the Ginko learnt so far. So she has asked for some images of objects, which her support team plans to keep in the way of Ginko. What type of data is Aani collecting now ?

(*a*) Tabular data
(*b*) Robotic database
(*c*) Training data
(*d*) Testing data

PART B

LET US REVISE

- *Artificial Intelligence (AI) refers to ability of modern machines and computers to mimic human intelligence. AI machines work through data, algorithms and models.*
- *Most used of AI technologies are Machine Learning and Deep Learning.*
- *Machine Learning (ML) is a branch of AI that enables machines to automatically learn and improve at tasks with experience and by the use of data.*
- *Deep Learning (DL) is a subset of machine learning where learning takes place through examples by filtering the input data using layers and rules-based algorithms to predict and classify information.*
- *AI can either be Data driven (also called learning-based AI) or model driven (also called Rule based AI).*
- *Rule Based AI (Model driven AI) refers to the branch of AI where models are developed using the algorithms having pre-defined labels, rules, patterns and relationships.*
- *The Learning based AI (data driven AI) refers to that branch of AI where models are trained to learn by inputting them tons of data. Here there are no patterns, rules and relationships predefined by the developer, rather machine learns with each new input and comes up with own algorithm.*
- *The process of developing an AI algorithm (AI model) is called AI modelling where code is developed to produce intelligent outcomes using a set of data.*
- *True positive refers to an instance for which both predicted values by the AI model and actual values are positive.*
- *True negative refers to an instance for which both predicted by the AI model and actual values are negative.*
- *False positive refers to an instance for which predicted value by the AI model is positive but actual value is negative.*

- *False negative refers to an instance for which predicted value by the AI model is negative but actual value is positive.*
- *Deployment is the mechanism through which the developed and trained AI model is made available to users for actual use.*
- *Common deployment techniques are: on device/premise deployment, cloud deployment, edge device deployment.*
- *A cloud means the storage, databases and processing systems available on the Internet servers.*
- *Edge devices, are computers or intelligent devices that enable real-time raw data collection and processing for quicker, more accurate analysis.*

Solution Time

1. *How do Machine Learning systems work ?*

   Ans. Machine Learning (ML) is an AI technology that takes lots of input data and finds patterns in data and uses them to make predictions. ML algorithms use computational methods to "learn" information directly from data without relying on a predetermined equation as a model.

2. *How do Deep Learning systems work ?*

   Ans. Deep learning is a specialized form of machine learning. Deep Learning (DL) enables software to train itself to perform tasks with vast amounts of data. Since the system has got huge set of data, it is able to train itself with the help of multiple machine learning algorithms working altogether to perform a specific task.

3. *Explain true positives and true negatives with the help of an example.*

   Ans. True Positive and true negatives occur when AI identifies correct instance of positive or negative outcome. For example, imagine a doctor uses an AI model to detect if someone has a certain disease based on their symptoms. A true positive would be when the AI correctly identifies a person who has the disease. So, if the AI says someone has the disease, and they really do, that's a **true positive**.

   On the other hand, a true negative is when the AI correctly says someone doesn't have the disease, and they truly don't. For instance, if the AI says someone is healthy, and they are indeed healthy, that's a **true negative**.

4. *Explain false positives and false negatives with the help of an example.*

   Ans. **False Positive** and **False negative** occur when AI makes a mistake in identifying true or false outcomes. For example, if a doctor uses an AI model to detect if someone has a certain disease based on their symptoms and if this AI model wrongly identifies someone as having the disease when they're actually healthy, that's a **false positive**. It's like when you get a question wrong on a test even though you knew the answer.

   Conversely, a **false negative** happens when the AI incorrectly says someone is healthy when they actually have the disease. So, if the AI misses detecting the disease in someone who really has it, that's a false negative.

5. *What do you understand by deployment? Name some commonly used AI model deployment methods.*

   Ans. Deployment is the mechanism through which the developed and trained AI model is made available to users for actual use.

   Common deployment techniques are: on device/premise deployment, cloud deployment, edge device deployment.

GLOSSARY

Cloud Storage databases and processing systems available on Internet.

Deep Learning A subset of machine learning where learning takes place through examples by filtering the input data using layers and rules-based algorithms.

Edge Device Intelligent devices that enable real-time raw data collection and processing.

Machine Learning A branch of AI that enables machines to automatically learn and improve at tasks with experience and by the use of data.

Testing Data A set of data that is used to test the outcome of developed model.

Training Data A set of data that is used to train a computer to perform a task.

Assignment

1. What is AI ? Name its most used technologies.
2. What is Machine learning ?
3. What is Deep learning ?
4. How is Deep learning related to Machine learning ?
5. How is data driven AI different from model driven AI ?
6. What is Rule based AI ?
7. What is Learning based AI ?
8. What do you understand by AI modelling ?
9. Name some popular AI models.
10. Define the terms : (*a*) True Positive (*b*) True Negative (*c*) False Positive (*d*) False Negative.
11. What is deployment of AI model ?
12. Discuss various techniques for the deployment of AI-models.
13. Discuss the pros and cons of various deployment techniques of AI-models.

PART B

PRACTICAL ASSIGNMENT

Explain how AI can be used to personalize learning experiences by adapting to individual student needs, strengths, and weaknesses.

1. **Divide the class into small groups.** Ask each group to identify a specific learning challenge or area where personalized education could be beneficial. For example, it could be in a subject like math, science, or language arts.
2. **Data Collection.** Have students collect relevant data related to the identified problem. This could include :
   - Student performance data (*e.g.*, test scores, assignments)
   - Learning preferences and styles (*e.g.*, visual, auditory, kinesthetic)
   - Demographic information (*e.g.*, age, gender, socioeconomic status)
   - Feedback from teachers and parents

3. **Data Preprocessing.** Guide students in cleaning and organizing the collected data, ensuring it is in a format suitable for analysis.
4. **Model Training.** Introduce students to machine learning algorithms suitable for personalized education, such learning based or rule based such as decision trees.
5. **Model Evaluation.** Discuss the importance of evaluating the trained model's performance and the meaning and role of True positive, true negative, false positive, false negative and their implications.

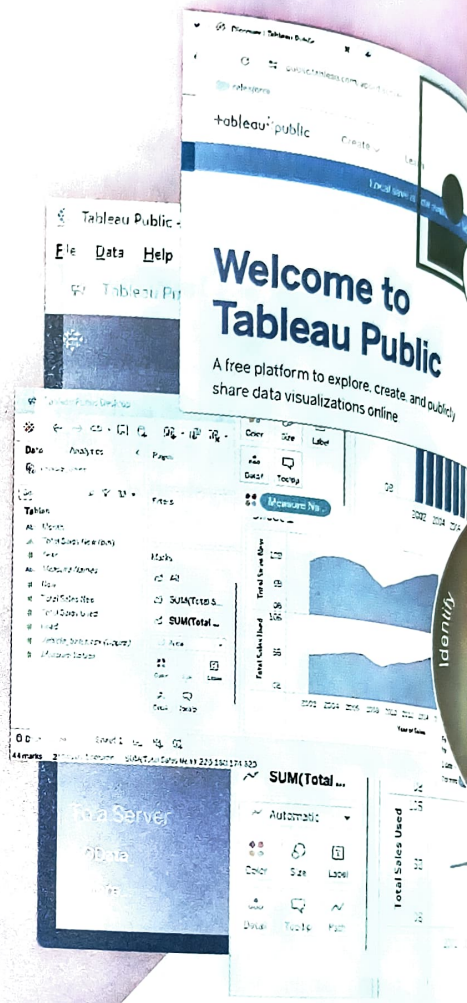
   Have students evaluate their models and interpret the results.
6. **Model Deployment.** Explore ways to deploy the trained model, such as creating a simple web application or integrating it with existing learning management systems or installing the model on a laptop/computer locally.

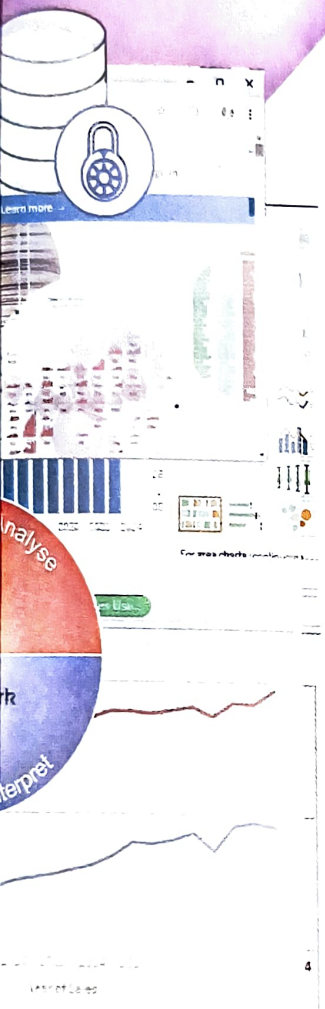
   Discuss the potential challenges and considerations of deploying AI models in real-world educational settings.
7. **Continuous Improvement.** Emphasize the iterative nature of the AI project cycle. Encourage students to gather feedback, refine their models, and explore ways to improve the personalized learning experience.

   Throughout the activity, emphasize the importance of ethical considerations, such as data privacy, algorithmic bias, and the role of human oversight in AI-driven personalized education systems.

II

Data Literacy





In This Unit

Session 1	Basics of Data Literacy
Session 2	Acquiring Data, Processing, and Interpreting Data
Session 3	Project Interactive Data Dashboard and Presentation

SESSION 1

Basics of Data Literacy

- What is Data Literacy ?
- Impact of Data Literacy
- Data Security, Privacy and AI
- Best Practices for Cyber Security

1.1 INTRODUCTION

Modern world is the world of technology AI and Data. In fact, data rules the world. In this data-ruled-world, it becomes important for everyone to understand and befriend data. That is where 'data literacy' comes into picture. In this session, you will learn about *data literacy, its impact, importance* and *role in AI.*

1.2 WHAT IS DATA LITERACY ?

In today's digital world, data is everywhere, and it's essential to know how to work with it. Data literacy is the ability to understand, analyze, and communicate with data, effectively. Data literacy is important because it helps you think critically, make better choices, and understand the world of data.

Data Literacy

Data Literacy is the ability to understand, analyze, and communicate with data, effectively.

Data literacy involves these *three* things (Fig. 1.1) :

1. **Reading Data.** It refers to the ability to read, analyze and understand data in, which may be available in various formats such as numbers, tables, graphs, charts etc.
2. **Working with Data.** It refers to the process of collecting and managing data; checking facts and spotting misleading information in it; storing and transmitting it in appropriate formats.
3. **Communicating with Data.** It refers to the process of understanding and interpreting data ; spotting trends and patterns in the data; tabulating, and reporting and presenting data in different and diverse formats.

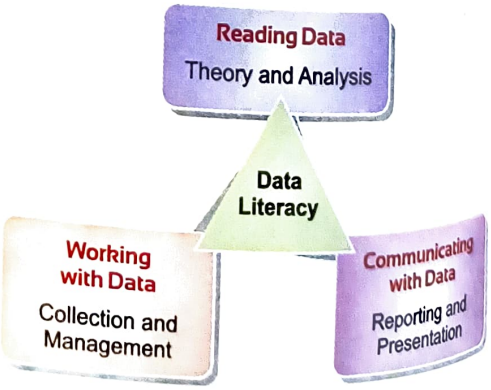


Figure 1.1 Data Literacy Components

1.3 IMPACT OF DATA LITERACY

Data literacy is one of the most important skills needed today. It opens doors to endless possibilities. The impact and importance of data literacy can be summarized through the following benefits it provides.

1. **Spot Data Trends.** Data literacy helps us ***analyze situations***, ***spot trends***, and ***predict outcomes***. For instance, by examining historical data and spotting the data trends, we can understand the causes of significant events like economic recessions.
2. **Foster Critical Thinking.** With data literacy, instead of accepting information at face value, we *learn to question its source, reliability*, and *potential biases*. This saves us from misinformation.
3. **Make Informed Decisions.** With data skills, we can gather and analyse relevant information. This helps us *make choices based on facts*, and not on assumptions or guesses.
4. **Communicate Effectively.** Data literacy empowers us with factual power. We can use data to support our arguments and ideas. With the help of Data visualization, we can *present information more clearly*.
5. **Solve Complex Problems.** Many modern problems involve large amounts of data. Data literacy makes it easier *to break down bulk data, spot patterns in it, identify trends* and *solve challenging problems*.

Overall, data literacy empowers us to navigate our data-driven world confidently.

1.4 HOW TO BECOME DATA LITERATE ?

Becoming proficient in data literacy involves following a structured process called the **Data Literacy Process Framework.** It involves majorly the following *three* steps :

Step 1 : Identify

In this step, all the data & information collection happens, whether it's numbers, words, or pictures, *e.g.,* if we're studying the weather, the data might include *temperature readings, rainfall amounts, wind speeds* and so on.

So, to become data literate, as part of ***Identify*** step, you need to learn to :

- Understand data types (numerical, categorical, etc.).
- Get familiar with common data visualizations. **Data visualisation** is the graphical or visual representation of information/data, such as graphs or charts such as, *bar charts, scatter plots* etc.
- Practice reading and interpreting data from various sources.

Note

Data visualisation is the graphical or visual representation of information and data.

Step 2 : Analyse

In this step, data is carefully studied to uncover ***patterns, connections, correlations, trends,*** and ***outliers***. For example, analysing sales data might reveal which products are performing well and which ones need improvement.

So, to become data literate, as part of ***Analyse*** step, you need to learn to :

- **Gather and organize data** from different sources. Question the data source and collection methods.
- **Tidy data** by removing redundancies and errors. Look for potential biases, errors, or missing information and handle them.
- **Explore & Visualize** examine patterns and create visualizations of the data.

Step 3 : Interpret

In this step, the curated and visualised data goes through the lens of questions like, *"What do these patterns tell us?"*, *"What can we learn from this data?"* For instance, if you are analysing student test scores, you might interpret the data to identify areas where students need extra help.

So, to become data literate, as part of ***Interpret*** step, you need to learn to :

- Determine what insights can be drawn from data.
- Share findings about the data, such as how can this information be applied in real-life scenarios ?
- Consider alternative interpretations or explanations.

Remember, practice makes perfect! The more you engage with data, the more proficient you'll become at interpreting and leveraging its insights.

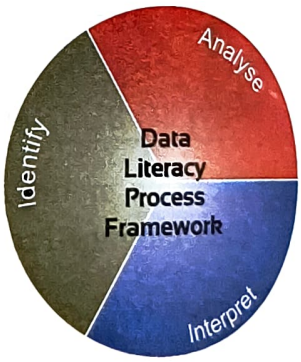


Figure 1.2 Data Literacy Process Framework

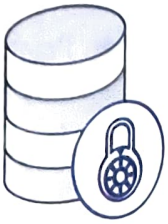
Note

A summarized way of looking at the Data Literacy Process Framework brings it down to two key steps : (*i*) Explore and Visualize, (*ii*) Analyze and Communicate.

1.5 DATA SECURITY, PRIVACY AND AI

With data all around us and data governing the decisions affecting our lives, Data security and privacy become super important. These are twin pillars of safeguarding our digital lives. Let us talk about these two and their relation with AI.

1.5.1 Data Security



Just like we lock our doors to protect our homes, *data security* involves safeguarding data and information from unauthorized access, theft, or alteration, and involves preventives measures to stop data breaches, which can have severe consequences.

For instance, imagine you have a password-protected document containing your personal information. Encrypting the file adds an extra layer of security, making it unreadable to anyone without the decryption key.

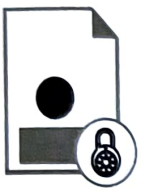
Thus, data security involves :

- protecting data from unauthorized access, theft, or damage
- taking measures like encryption, access controls, firewall and network security to prevent data breaches.

Data Security

Data Security refers to taking steps for safeguarding data and information from unauthorized access, theft, or alteration, and involves preventives measures to stop data breaches.

1.5.2 Data Privacy



Data privacy refers to controlling the access to our personal information and its use. It's like setting boundaries to protect our digital identity from exploitation or misuse. Consider the data you share on social media platforms. By adjusting privacy settings, you can control who sees *your posts, photos, and personal details*, ensuring they're only accessible to trusted individuals.

Data Privacy involves :

- Concerns for safeguarding personal or sensitive information from misuse.
- Practices like data minimization and user consent.
- Control over how one's personal data and digital footprint is collected and used.

Data Privacy

Data Privacy refers to controlling the access to our personal information and its use.

1.5.3 Data Security, Privacy and AI

The AI systems need the vast amounts of data to learn and make decisions. Whether it's training data for machine learning algorithms or personal information used for targeted advertising, maintaining robust data security and privacy measures is crucial to prevent unauthorized access or misuse of data.

So, we can say that the relationship between AI and data security/privacy is complex Fig. 1.3.

AI raises Data Privacy concerns	***AI provides enhanced Data Security***
AI's requirement and use of data raises concerns about data privacy : ★ AI systems require large amounts of data for training and operation; data may include personal or sensitive information. ★ It raises privacy concerns about how this data is collected and used. ★ AI systems may have vulnerabilities that can be exploited. ★ It raises questions about individual rights and AI transparency.	But AI can also enhance security measures for data : ★ Machine learning models can detect cyber threats or anomalies. ★ AI can help identify and protect against data breaches.

As AI becomes more prevalent, it's crucial to balance its capabilities with robust data security and privacy protections.

This requires :

- Developing AI systems with security and privacy in mind right from the beginning (known as ***Security/Privacy by Design***)
- Implementing strong data governance and ethical AI principles
- Ongoing evaluation of AI's impact on security and privacy landscapes

Figure 1.3 Relationship between AI and Data Security/Privacy.

Data security and privacy have a mixed relation with AI. Data used to train AI should be secure and from authentic sources. Comprised data can impact the performance of AI. At the same time, AI can play an important role in detecting and protecting against data breaches.

PART B

1.6 BEST PRACTICES FOR CYBER SECURITY

While AI presents new challenges, there are established best practices for protecting data security and privacy. Here are some best practices for cyber security :

1. Use Strong Passwords

Use unique, complex passwords for each of your accounts. Avoid using easily guessable information like birthdays or pet names. Instead, opt for a combination of letters, numbers, and special characters.

Example : Instead of using "password123," try something like "P@ssw0rd!2#"

2. Keep Software and Systems Updated

Regularly update your operating system, antivirus software, and applications to patch security vulnerabilities and protect against the latest threats.

Example : Your phone prompts you to install a software update that includes security enhancements and bug fixes. You promptly install the update to keep your device secure.

3. Be Cautious with Emails and Links

Be cautious of suspicious links or attachments in emails or messages. These could be phishing attempts designed to steal your personal information or infect your device with malware.

Example : You receive an email claiming to be from your bank asking you to click on a link to update your account information. Instead of clicking, you must first verify the legitimacy of the email directly with your bank.

4. Backup Data Regularly

It is important to take backup of your work and data regularly. Always maintain offline backups of important data. This protects against ransomware or data loss incidents.

5. Use Encryption for Stored Sensitive Data

Always encrypt sensitive data stored on devices or in transit.

6. Visit Secure Sites and Use Secure Connections

While browsing online, always make sure to visit secure sites over secure connections. Secure sites' URLs start with HTTPS and have a padlock sign on the status bar to depict it. Also, make sure not to use public free WiFi's unless you are sure about their security measures. Preferably use VPNs when possible.

7. Limit Access and Sharing of Sensitive Data

Only share personal or confidential information when necessary. Always make sure to restrict access to data on a "need-to-know" basis.

8. Educate Yourself and Others

Always, stay updated on latest cyber threats and best practices. It is equally important to raise awareness about data security and privacy for ourselves and our community.

By following these guidelines and prioritizing data security/privacy, one can protect oneself and one's information in our increasingly digital world.

With this we have come to the end of this session. Let us quickly revise what we have learnt so far.

Check Point

Multiple Choice Questions

1. Data literacy refers to the ability to :
   (*a*) Write computer code
   (*b*) Understand and analyze data
   (*c*) Create databases
   (*d*) Operate software programs
2. Which of the following is NOT an example of how data literacy can be beneficial ?
   (*a*) Fact-checking claims and spotting misinformation
   (*b*) Understanding trends and patterns in your community
   (*c*) Making decisions about your health based on data
   (*d*) Writing computer programs
3. The first step in the Data Literacy Process Framework is :
   (*a*) Explore and Visualize
   (*b*) Analyze and Communicate
   (*c*) Import and Tidy Data
   (*d*) Practice and Repeat

4. Which of the following is NOT a part of the Data Literacy Process Framework ?
   (a) Import and Tidy Data
   (b) Explore and Visualize
   (c) Analyze and Communicate
   (d) Write Code
5. Data security refers to :
   (a) Protecting personal information
   (b) Preventing unauthorized access to data
   (c) Ensuring data accuracy
   (d) Analyzing data for insights
6. Which of the following is a benefit of data literacy ?
   (a) Helps in making informed decisions
   (b) Enhances critical thinking skills
   (c) Enables effective communication with data
   (d) All of these
7. Which of the following is NOT a step in the Data Literacy Process Framework ?
   (a) Import and Tidy Data
   (b) Explore and Visualize
   (c) Analyze and Communicate
   (d) Build Machine Learning Models
8. What does it mean to "tidy" data ?
   (a) Remove any errors or inconsistencies
   (b) Organize data in a structured format
   (c) Visualize data in charts or graphs
   (d) Interpret data and draw insights
9. Which of the following is a measure for ensuring data security ?
   (a) Obtaining user consent
   (b) Data minimization
   (c) Encryption
   (d) Anonymization
10. Data privacy aims to :
   (a) Prevent unauthorized access to data
   (b) Give individuals control over their personal data
   (c) Protect data from being lost or damaged
   (d) Detect and respond to cyber threats
11. Which of the following is NOT a potential security risk associated with AI systems ?
   (a) Vulnerabilities that can be exploited
   (b) Manipulated data
   (c) Challenges to existing privacy regulations
   (d) Improved detection of cyber threats
12. What does "Privacy by Design" refer to ?
   (a) Developing AI systems with privacy in mind from the start
   (b) Implementing strict access controls for sensitive data
   (c) Obtaining explicit consent from users for data collection
   (d) Regularly updating privacy policies and terms of service
13. Which of the following is NOT a best practice for cyber security ?
   (a) Using strong passwords and multi-factor authentication
   (b) Keeping software and systems updated with latest security patches
   (c) Being cautious with emails and links from unknown sources
   (d) Sharing personal information freely over unsecured channels

PART B

14. What is the purpose of data encryption ?
    (*a*) To prevent unauthorized access to sensitive data
    (*b*) To ensure data accuracy and integrity
    (*c*) To backup and restore data in case of loss
    (*d*) To analyze and interpret data effectively
15. Which of the following is an example of a secure connection?
    (*a*) HTTP website
    (*b*) Public Wi-Fi network
    (*c*) HTTPS website
    (*d*) Unencrypted email
16. What is the purpose of limiting access to sensitive data ?
    (*a*) To enhance data security
    (*b*) To comply with privacy regulations
    (*c*) To facilitate data sharing and collaboration
    (*d*) Both (*a*) and (*b*)
17. Which of the following is NOT a way to protect data privacy ?
    (*a*) Using data minimization techniques
    (*b*) Obtaining user consent for data collection
    (*c*) Implementing strong encryption measures
    (*d*) Sharing personal data freely with third parties
18. What does it mean to "fact-check" information using data literacy skills?
    (*a*) Verifying the accuracy of claims or information using data
    (*b*) Creating visualizations to present data effectively
    (*c*) Organizing and cleaning data before analysis
    (*d*) Applying statistical techniques to analyze data
19. Which of the following is NOT a critical question to ask when practicing data literacy ?
    (*a*) What is the source of the data ?
    (*b*) Are there any potential biases or errors in the data ?
    (*c*) How much is the storage cost of data ?
    (*d*) Are there alternative interpretations or explanations for the data ?
20. What does it mean to "explore" data in the Data Literacy Process Framework ?
    (*a*) Gather and organize data from various sources
    (*b*) Examine patterns and trends within the data
    (*c*) Draw insights and communicate findings from the data
    (*d*) Validate the accuracy and completeness of the data
21. Which of the following is NOT a potential risk associated with data breaches ?
    (*a*) Financial losses
    (*b*) Reputational damage
    (*c*) Legal consequences
    (*d*) Improved cybersecurity measures
22. What is the purpose of maintaining offline backups of important data ?
    (*a*) To protect against ransomware or data loss incidents
    (*b*) To enhance data processing speed
    (*c*) To comply with data retention regulations
    (*d*) To facilitate data sharing across multiple devices

23. What is the importance of educating oneself and others about cyber security ?
    (*a*) To increase awareness and promote best practices
    (*b*) To comply with legal and regulatory requirements
    (*c*) To develop new cybersecurity technologies
    (*d*) To earn certifications or degrees in cybersecurity

Fill in the Blanks

1. Data literacy is the ability to _____, _____, and _____ with data effectively.
2. Being data literate helps you _____ claims and spot misleading information.
3. The _____ _____ _____ Framework guides you through working with data effectively.
4. Data privacy concerns safeguarding _____ _____ from misuse.
5. _____ _____ involves measures like encryption, access controls, and network security.
6. The goal of data literacy is to enable _____ decision-making.
7. _____ and _____ are two key steps in the Data Literacy Process Framework.
8. Data security helps prevent _____ _____.
9. "_____ by Design" refers to developing AI systems with security and privacy in mind from the start.
10. Using _____ passwords is one of the best practices for cyber security.
11. Regularly installing _____ is important for maintaining system security.
12. Be cautious with emails and links from _____ sources.
13. Use _____ to protect sensitive data stored on devices or in transit.
14. Connect to _____ websites and use _____ when possible for secure connections.
15. Stay updated on the latest _____ _____ and best practices for cyber security.
16. Raise _____ about data security and privacy in your community.
17. Data literacy empowers you to navigate our _____ _____ world confidently.

PART B

True / False

1. Data literacy is only important for data scientists and analysts.
2. Developing data literacy is a one-time process.
3. Data literacy can help you make better decisions based on facts.
4. Data security and data privacy are the same thing.
5. AI systems do not require any personal or sensitive data.
6. Data literacy only involves reading and interpreting charts and graphs.
7. Being data literate can help you make better decisions based on data.
8. The Data Literacy Process Framework does not include any steps related to data visualization.
9. Data security and data privacy are completely unrelated concepts.
10. AI systems do not require any data for training or operation.
11. AI systems do not pose any challenges to existing data privacy regulations.
12. Using strong passwords is not an important practice for cyber security.
13. It is safe to share sensitive information over unsecured email or websites.
14. Maintaining offline backups of data can protect against ransomware attacks.

15. Encryption is used to ensure data accuracy and integrity.
16. Connecting to HTTPS websites provides a secure connection.
17. It is recommended to share personal data freely with third parties.
18. Fact-checking claims using data is an example of applying data literacy skills.
19. Asking about potential biases or errors in data is not a critical question in data literacy.
20. Exploring data involves examining patterns and trends within the data.
21. Data breaches can only result in financial losses, but not legal or reputational consequences.
22. Educating oneself and others about cyber security is important for increasing awareness.

Competency Based Questions

1. Your local community organization is planning a fundraising event, and they have collected data on past attendance, donation amounts, and demographics of supporters. Apply steps to analyze this data and provide actionable insights to the organization. Your analysis should help them determine the optimal event format, target audience, and marketing strategies to maximize participation and fundraising efforts. How would you accomplish this ?
   (*a*) Ignore the data and plan the event based on personal preferences.
   (*b*) Analyze the data but provide vague or irrelevant insights.
   (*c*) Apply the Data Literacy Process Framework, analyze the data, and provide actionable insights for event planning.
   (*d*) Outsource the data analysis to a third-party without providing any guidance.
2. Investigate the use of AI in various industries (*e.g.*, healthcare, finance, transportation). What would be the best approach to use this report.
   (*a*) Ignore the potential concerns and focus only on the benefits of AI.
   (*b*) Investigate the use of AI but fail to provide any guidelines or recommendations.
   (*c*) Develop a set of guidelines or recommendations without investigating the use of AI or potential concerns.
   (*d*) Investigate the use of AI, identify potential concerns, and develop guidelines or recommendations to address them.
3. Examine a real-world dataset related to a topic of your interest (e.g., environmental issues, social trends, global health). Which of the following approaches would you follow to present your analysis of the same highlighting the importance of critical questioning and data literacy skills.
   (*a*) Examine the dataset without any critical analysis or alternative interpretations.
   (*b*) Propose alternative interpretations without examining or analyzing the dataset.
   (*c*) Critically analyze the data, identify biases/limitations, propose alternative interpretations, and present findings.
   (*d*) Ignore the dataset and present personal opinions on the topic.
4. To investigate the role of data literacy in combating misinformation and fake news, which of the following approaches could have been used to identify and counter the spread of misinformation.
   (*a*) Investigate the role of data literacy without any real-world case study
   (*b*) Analyze a real-world case study general way
   (*c*) Investigate the role of data literacy, analyze a real-world case study, and present findings and recommendations
   (*d*) Compile incidents of misinformation and fake news from real-world data

Assertion & Reasoning Questions

Directions

In the following questions, a statement of assertion (A) is followed by a statement of reason (R).

Mark the correct choice as :

(*a*) Both A and R are true and R is the correct explanation of A.

(*b*) Both A and R are true but R is not the correct explanation of A.

(*c*) A is true but R is false (or partly true).

(*d*) A is false (or partly true) but R is true.

(*e*) Both A and R are false or not fully true.

1. Assertion (A). Data literacy involves understanding and analyzing various types of data, including numbers, graphs, and charts.
   Reason (R). Data literacy is the ability to read and interpret data in only numerical formats.
2. Assertion (A). The Data Literacy Process Framework is a linear process with no interconnected steps.
   Reason (R). The Data Literacy Process Framework includes steps for exploring, visualizing, and communicating data.
3. Assertion (A). Data security measures, such as encryption and access controls, help prevent unauthorized access to data.
   Reason (R). Data security and data privacy are interchangeable terms that refer to the same concept.
4. Assertion (A). AI systems can enhance certain security measures by detecting cyber threats or anomalies.
   Reason (R). AI systems themselves cannot introduce any security risks or vulnerabilities.
5. Assertion (A). Implementing "Privacy by Design" principles involves developing AI systems with privacy considerations from the outset.
   Reason (R). Privacy by Design is a concept that is useful in establishing robust data security and privacy protections controls.
6. Assertion (A). Using strong passwords and enabling multi-factor authentication are best practices for cyber security.
   Reason (R). Regularly updating software and systems with the latest security patches is not an important cyber security practice.
7. Assertion (A). Educating oneself and others about cyber threats and best practices can increase awareness and promote better security practices.
   Reason (R). Raising awareness about data security and privacy is only relevant for IT professionals and not the general public.

PART B

LET US REVISE

- *Data literacy is the ability to understand, analyze, and communicate with data effectively.*
- *Data literacy involves three things: Reading data, Working with Data, and Communicating with data.*
- *Data literacy offers benefits like: spotting data trends, fostering critical thinking, helping in making informed decisions, effective communication, and helping in solving complex problems.*
- *Becoming proficient in data literacy involves following a structured process called the Data Literacy Process Framework.*

- *The Data Literacy Process Framework involves majorly three steps : **Identify** (to understand data), **Analyse** (to uncover patterns in data), and **Interpret** (to get insights from data).*
- *Data visualisation is the graphical or visual representation of information and data, such as graphs or charts like bar charts, scatter plots etc.*
- *Data security refers to taking steps for safeguarding data and information from unauthorized access, theft, or alteration, and involves preventives measures to stop data breaches.*
- *Data privacy refers to controlling the access to our personal information and its use.*
- *Data security and privacy have a mixed relation with AI. Data used to train AI should be secure, from authentic sources. Comprised data can impact the performance of AI. At the same time, AI can play an important role in detecting and protecting against data breaches.*
- *Best Practices for Cyber Security include: using strong passwords, keeping systems and software updated, being cautious with emails and links, backing up data regularly, using encryption for stored sensitive data, visiting secure sites and using secure connections, limiting access and sharing of data, and keep educating oneself about latest threats and protection measures.*

Solution Time

1. *What is data literacy ?*

   Ans. Data literacy is the ability to understand, analyze, and communicate with data effectively.

2. *Why is data literacy important in today's world ?*

   Ans. Data literacy is important because it helps individuals think critically, make better decisions based on facts, communicate effectively using data, and solve complex problems involving large amounts of data.

3. *Describe the three main steps of the Data Literacy Process Framework.*

   Ans. The *three* main steps are : (*i*) Import and Tidy Data (gather and organize data), (*ii*) Explore and Visualize (examine patterns and create visualizations), (*iii*) Analyze and Communicate (draw insights and share findings).

4. *Differentiate between data security and data privacy.*

   Ans. Data security refers to protecting data from unauthorized access, theft, or damage. Data privacy concerns safeguarding personal or sensitive information from misuse and giving individuals control over how their data is collected and used.

5. *How can AI enhance data security measures ?*

   Ans. AI can enhance security measures by using machine learning models to detect cyber threats, anomalies, or potential data breaches.

6. *What are some potential security risks associated with AI systems ?*

   Ans. AI systems may have vulnerabilities that can be exploited, be susceptible to adversarial attacks with manipulated data, and challenge existing data privacy regulations.

7. *What is the purpose of "Privacy by Design" ?*

   Ans. Privacy by Design refers to developing AI systems with privacy considerations and protections in mind from the very beginning, rather than as an afterthought.

8. *Why is it important to use strong passwords and enable multi-factor authentication ?*

   Ans. Using strong, unique passwords and enabling multi-factor authentication are best practices for cyber security as they help prevent unauthorized access to accounts and data.

9. *Explain the importance of keeping software and systems updated with the latest security patches.*

   Ans. Regularly installing software updates and security patches helps address known vulnerabilities and protect against potential cyber threats or attacks.

10. *What is the role of data encryption in ensuring data security ?*

    Ans. Data encryption is used to prevent unauthorized access to sensitive data by converting it into a coded format that can only be read with the correct decryption key.

11. *Why is it recommended to limit access to sensitive data on a "need-to-know" basis ?*

    Ans. Limiting access to sensitive data on a need-to-know basis enhances data security and helps comply with data privacy regulations by restricting access to only those who require it for their work.

12. *How can educating oneself and others about cyber security promote better practices?*

    Ans. Educating oneself and others about cyber threats, best practices, and the importance of data security and privacy can increase awareness and encourage the adoption of safer practices.

GLOSSARY

Data Literacy Ability to understand, analyze, and communicate with data effectively.

Data Privacy Controlling the access to our personal information and its use.

Data Security Taking steps for safeguarding data and information from unauthorized access.

Data Visualisation Graphical or visual representation of information and data.

PART B

Assignment

1. Which of the following is NOT a benefit of data literacy ?
   (*a*) Helps in making evidence-based decisions
   (*b*) Enhances problem-solving skills
   (*c*) Enables effective storytelling with data
   (*d*) Improves artistic abilities
2. What is the first step in the Data Literacy Process Framework ?
   (*a*) Explore and Visualize
   (*b*) Import and Tidy Data
   (*c*) Analyze and Communicate
   (*d*) Interpret and Conclude
3. Data security aims to protect data from :
   (*a*) Being accessed by authorized individuals
   (*b*) Being used for legitimate purposes
   (*c*) Unauthorized access, theft, or damage
   (*d*) Being shared with third parties
4. Data literacy empowers individuals to make _____ based on facts and data. (decisions)
5. The _____ step in the Data Literacy Process Framework involves organizing and structuring data. (Import & Tidy Data)
6. Data security measures help prevent _____ to data. (unauthorized access)

7. AI systems may have _____ that can be exploited by cyber attackers. (vulnerabilities)
8. "_____ by Design" refers to incorporating security and privacy considerations into AI systems from the outset. (Security/Privacy)
9. Use _____ passwords and avoid reusing the same password across multiple accounts. (strong, unique)
10. Regularly update software and install _____ to address known vulnerabilities. (security patches)
11. Be cautious with emails and links from _____ sources, as they may be attempts at phishing or distributing malware. (untrusted, unknown)
12. Use _____ connections (like HTTPS) and _____ when connecting to public networks. (secure, VPNs)
13. Data literacy involves only the ability to read and interpret numerical data. (F)
14. The Data Literacy Process Framework is a linear, step-by-step process with no interconnected stages. (F)
15. Data security and data privacy are interchangeable terms that refer to the same concept. (F)
16. AI systems do not require any data for training or operation. (F)
17. Implementing strong encryption measures is a best practice for protecting data privacy. (T)
18. Privacy by Design principles focus solely on implementing strict access controls for sensitive data. (F)
19. Using unique and complex passwords for all accounts is not an important cyber security practice. (F)
20. It is safe to connect to unsecured public Wi-Fi networks for convenience. (F)
21. Maintaining offline backups of data can help protect against ransomware attacks. (T)
22. Describe the three main steps of the Data Literacy Process Framework.
23. Explain the benefits of being data literate and how it can impact various aspects of life.
24. Describe the process of applying the Data Literacy Process Framework to analyze a dataset.
25. Discuss the relationship between AI and data privacy concerns, providing specific examples.
26. Discuss two potential issues related to AI's relation with data security and privacy concerns.
27. What measures can individuals and organizations take to enhance data security ?
28. Why is it important to be cautious with emails and links from unknown sources ?
29. Describe the purpose and importance of maintaining offline backups of important data.
30. How can data literacy skills help individuals do fact-check claims and spot misinformation ?
31. Discuss the role of critical questioning in developing data literacy.
32. Explain the potential consequences of data breaches for individuals and organizations.
33. How can education and awareness-raising efforts contribute to better cyber security practices ?
34. Describe the steps you would take to protect your personal data privacy online.

SESSION 2

Acquiring Data, Processing, and Interpreting Data

- Types of Data
- Data Acquisition
- Best Practices for Acquiring Data
- Features of Data and Data Preprocessing

2.1 INTRODUCTION

You have learnt in the previous session that today's world is data driven world. In such a world, understanding the types of data is essential for making informed decisions and driving progress. From the moment data is generated, to its interpretation and application, every step in the data journey plays a crucial role.

In this session, we shall talk about various types of data, data acquisition, data processing and the importance of effective data interpretation.

2.2 TYPES OF DATA

Data comes in many forms, each serving a distinct purpose in analysis and decision-making. Although data is available in multiple forms, broadly it can be divided into two primary types of data.

1. Qualitative Data

Qualitative data provides insights into the *characteristics, attributes, and qualities of phenomena.* Qualitative data is often subjective and descriptive.

For instance, following are all examples of qualitative data :

- **Customer feedback**, such as comments, reviews, and testimonials provide qualitative insights into customer satisfaction, preferences, and experiences.
- **Interview transcripts**, such as conversations with individuals or focus groups yield qualitative data, offering perspectives, opinions, and personal stories.
- **Social media sentiment**, such as posts, comments, and discussions on social media platforms reveal qualitative insights into public opinion, trends, and sentiment.

Note

Qualitative data describes qualities or characteristics of some entity or phenomena.

2. Quantitative Data

Quantitative data represents information about something through numerical values, enabling objective analysis and statistical modeling. It is structured, measurable, and helpful for numerical processing and manipulation.

Note

Quantitative data represents information about something through numerical values.

For instance, following are all examples of qualitative data :

- **Sales data**, such as transactional records, revenue figures, and purchase history provide quantitative insights into sales performance, trends, and patterns.
- **Sensor readings**, such as measurements of temperature, pressure, and humidity collected by sensors for monitoring environmental conditions.
- **Financial metrics**, such as stock prices, market indices, and financial statements furnish quantitative insights into economic indicators, investment performance, and financial health.

Qualitative data is also known as ***Categorical data*** and Quantitative data is known as ***Numerical data***. You have already read about these in section 4.2.1 page 119-120.

Note

Qualitative data is also known as Categorical data and Quantitative data is known as numerical data.

Quantitative & Qualitative Data

Quantitative Data is data involving numbers and measuring variables and **Qualitative Data** is data involving descriptive, non-numerical information.

Table 2.1 : *Qualitative Data vs. Quantitative Data*

Qualitative Data	***Quantitative Data***
It describes qualities or characteristics.	It represents quantities or numeric values.
It cannot be measured or counted numerically, *e.g.*, colours, opinions, feelings, descriptions.	It can be measured and expressed numerically, *e.g.*, counts, measurements, scores, ratings.
It is collected through observations, interviews, open-ended questions.	It is collected through structured data collection methods like surveys, sensors.
It is subjective and open to interpretation.	It is objective and statistical analysis is possible.
It provides in-depth insights and context.	It provides precise, measurable, and testable data.
It is analysed using methods like content analysis, coding.	It is analysed using statistical and computational methods.
It answers "why" and "how" types of questions.	It answers "how many" and "how much" types of questions.
Qualitative data types: *Text, audio, video, images*	Quantitative data types: *Numbers, percentages, frequencies.*
It is useful for exploring new phenomena.	It is useful for testing hypotheses, generalizing findings.
Examples : Customer reviews, focus group transcripts, case studies	**Examples** : Sales figures, test scores, survey ratings, sensor readings

2.3 DATA ACQUISITION

Data Acquisition refers to processes, methods or systems that are used to collect information related to a certain theme or objective, to document or analyse some phenomenon. Acquiring data involves collecting information from diverse sources using various methods tailored to specific objectives and contexts.

Data Acquisition

Data Acquisition refers to processes, methods or systems that are used to collect information related to a certain theme or objective, to document or analyse some phenomenon.

You have already read about common approaches to data acquisition, in section 5.3 page 136.

Let us quick recall about some of the common data acquisition methods, here :

1. **Surveys.** Surveys are structured questionnaires administered to individuals or groups to gather data on opinions, preferences, behaviors, and demographics.
2. **Interviews.** Interviews involve direct conversations between researchers and participants to collect qualitative data through open-ended questions and probing inquiries.
3. **Observations.** Observational studies involve systematic observation and recording of behaviors, interactions, and phenomena in natural or controlled settings.

There are many other ways of acquiring data. Please refer to section 5.3 page 136 for the same.

PART B

2.4 BEST PRACTICES FOR ACQUIRING DATA

Acquiring high-quality data requires using the best practices to ensure **validity**, **reliability**, and **ethical integrity**. For effective data acquisition, it is recommended to use following essential guidelines :

1. Define Clear Objectives. Clearly state the purpose, goals, and research questions (as per desired outcomes) that should give clear guidance to the data acquisition process.

2. Select Appropriate Methods. Choose data collection methods and techniques that are suitable for the context of data collection, objectives of data collection, and the target audience from which the data is being collected. For example, to get data from younger audience the questions and languages would be simpler than the case when data is to be collected from older or professional people.

3. Design Robust Instruments. Develop *survey questionnaires, interview guides, observation protocols*, and *experimental procedures* with careful attention to clarity, relevance, and reliability.

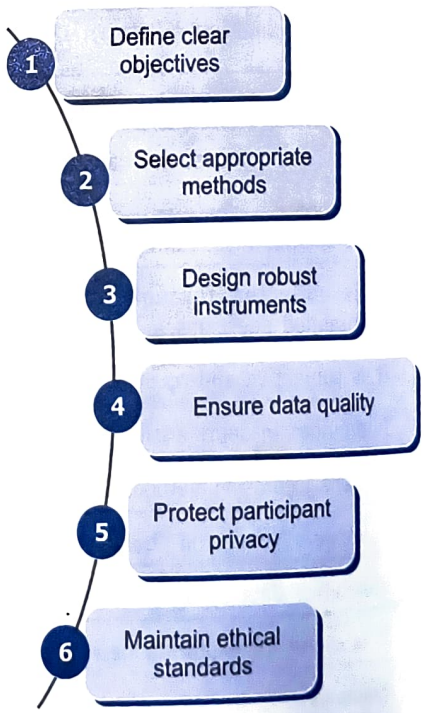


Figure 2.1 Best Practices for Data Acquisition.

4. Ensure Data Quality. Implement measures to validate, verify, and clean the collected data to minimize errors, inconsistencies, and biases.

5. Protect Participant Privacy. Respect confidentiality, anonymity, and informed consent principles to safeguard the rights and privacy of research participants.

6. Maintain Ethical Standards. Adhere to ethical guidelines, codes of conduct, and institutional policies governing research conduct, integrity, and transparency.

Note

High quality data is valid, reliable, and of ethical integrity.

2.5 FEATURES OF DATA AND DATA PREPROCESSING

Raw data often contains errors, inconsistencies, and missing values that require preprocessing to enhance quality, accuracy, and usability. Thus, data is preprocessed to clean it and make it appropriate for use.

Refer to section 4.2, page 119 that has already talked about data features and ways to preprocess data

Some common data preprocessing techniques are :

(*i*) Data Cleaning (*ii*) Data Reduction

(*iii*) Data Transformation (*iv*) Data Integration

Data Preprocessing

Data Preprocessing refers to the process of making data appropriate for use by removing discrepancies in it.

1. Data Cleaning

Data cleaning involves identifying and correcting errors, inconsistencies, and anomalies in raw data so it's easier to understand and work with and become more accurate and reliable, *e.g.*, imagine you have a list of student names and ages, but some ages are missing. You'd fill in the missing ages or remove the incomplete entries to make sure your data is complete with every student entry having an age.

There are multiple data cleaning methods, such as :

Data Cleaning

Data Cleaning is a process of identifying and correcting errors, inconsistencies, and anomalies in raw data.

- **Removing duplicate records.** Identifying and eliminating duplicate entries or observations to prevent redundancy and maintain data integrity.
- **Handling missing values.** Imputing missing values or deleting incomplete records to mitigate the impact of data gaps on analysis and interpretation.
- **Standardizing formats.** Converting data into consistent formats, units, and structures to facilitate comparison, aggregation, and analysis.

2. Data Reduction

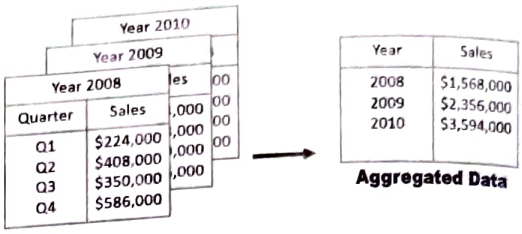
Data reduction is a technique used to reduce the size of a dataset while still preserving the most important information.

Data Reduction

Data Reduction is a process or a set of techniques used to reduce the size of a dataset while still preserving the most important information.

There are multiple data reduction methods, some of these are :

- **Sampling.** Pick samples representing a set of data, *e.g.*, when exit polls after elections are conducted, not everyone is asked. A sample set of people's responses are taken and results are determined.
- **Data aggregation.** Create aggregates or summaries of detailed data, *e.g.*, see the figure.



And many more.

3. Data Transformation

Data transformation involves converting raw data into a suitable format or representation for *analysis*, *visualization*, or *modeling*. There are multiple ways of doing so, some of these are :

- **Normalization**. This refers to scaling of numerical data to a common range or distribution to remove differences in magnitude and facilitate comparison.
- **Encoding categorical variables**. This refers to converting categorical variables into numerical or binary representations for computational analysis and modeling.
- **Feature engineering**. This refers to creating new features or variables from existing data to capture relevant patterns, relationships, or trends for predictive modeling.

Data Transformation

Data Transformation refers to the process of converting raw data into a suitable format or representation for analysis, visualization, or modeling.

PART B

4. Data Integration

Data integration involves combining data from multiple sources or formats into a unified dataset for analysis, reporting, or decision-making. There are multiple ways of integrating data, some of these are:

- **Merging datasets,** which is combining datasets with common identifiers or keys to enrich data with additional attributes or variables, *e.g.*, *if you have sales data in one spreadsheet and customer data in another, you'd merge to create a single dataset with both sales and customer information.*
- **Joining tables,** which is linking relational databases or tables based on shared fields or relationships to consolidate related information for analysis.
- **Concatenating files,** which is appending or concatenating files with similar structures or formats to create a comprehensive dataset for analysis or reporting.

Data Integration

Data Integration refers to the process of combining data from multiple sources or formats into a unified dataset for analysis, reporting, or decision-making.

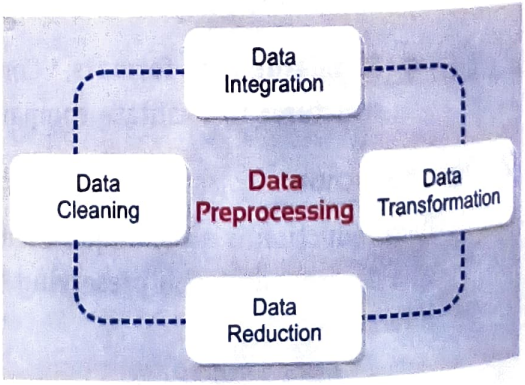


Figure 2.2 Data Preprocessing.

2.6 DATA PROCESSING

Once data is acquired and preprocessed, it undergoes processing and interpretation to extract insights, patterns, and actionable intelligence.

Data processing involves :

(*i*) Data Analysis

(*ii*) Data Interpretation

Data Processing

Data Processing refers to manipulating, analyzing, and Interpreting data to extract meaningful information and derive insights.

2.6.1 Data Analysis

Data analysis is like examining clues to solve a mystery. It involves studying your data to find *patterns*, *trends*, or *answers* to questions. For instance, suppose you have a dataset of student test scores. You might analyze the data **to see if there's a relationship between study time and test performance** by comparing scores of students who studied a lot versus those who studied less.

Data Analysis

Data Analysis is a process of apply many techniques on data to find/extract trends, correlations, outliers, and variations that convey a meaning or point to a specific result.

Data analysis takes place in many forms :

(*i*) Descriptive Analysis. Descriptive analysis shows ***what happened***. It describes data ***using statistics***, *e.g., analysing sales data to get sales numbers for each employee and the average sales.*

- *Descriptive analysis answers* What happened ?

(*ii*) Diagnostic Analysis. Diagnostic analysis finds ***why something happened***, *e.g., hospitals suddenly start having increased number of patients.* Descriptive analysis may find that *many hospital patients had the same virus symptoms, so the virus caused the patient increase.*

- *Diagnostic analysis answers* Why did it happen ?

(*iii*) Predictive Analysis. Predictive analysis predicts ***what might happen in the future*** based on data patterns, *e.g., a product sells best in September and October each year, so high sales are predicted for those months next year.*

- *Predictive analysis answers* What might happen in the future ?

(*iv*) Prescriptive Analysis. Prescriptive analysis recommends ***actions to take*** based on the other analysis types, *e.g., create a marketing plan to boost sales during the slower months after the September/October peak.*

- *Prescriptive analysis answers* What should we do about it ?

2.6.2 Data Interpretation

Interpreting data is like understanding the story behind the numbers. It involves making sense of data analysis results and drawing conclusions from it. For example, after analyzing the dataset of test score of students, you might interpret it as that **students who study more tend to perform better on tests, suggesting a positive relationship between study time and academic achievement.**

Data interpretation is essential for extracting meaningful insights from your data and using them to inform decision-making and problem-solving.

Data Interpretation

Data Interpretation involves making sense of the analyzed data, drawing conclusions, and deriving actionable insights to inform decision-making and problem-solving.

2.6.2A Types of Data Interpretation

Data interpretation takes place in various ways :

- **Exploratory data analysis (EDA).** EDA involves analysing and interpreting data visually to uncover patterns, relationships, and insights without preconceived hypotheses.
- **Confirmatory data analysis (CDA).** CDA focuses on testing hypotheses, confirming theories, or validating existing knowledge using statistical methods and inferential analysis.
- **Predictive data analysis.** Predictive analysis aims to forecast future trends, behaviours, or outcomes based on historical data patterns and predictive modeling techniques.
- **Diagnostic data analysis.** Diagnostic analysis focuses on understanding the underlying causes, drivers, or factors contributing to observe patterns, anomalies, or outcomes in data.

PART B

2.6.2B Importance of Data Interpretation

Effective data interpretation is important for extracting actionable insights, informing decision-making, and driving organizational success. Data interpretation is important and useful as it offers many benefits, such as :

1. It enables data-driven decision-making instead of relying on gut feelings alone.
2. It helps identify patterns, trends, and issues to solve problems and optimize processes.
3. It allows assessing performance of employees, products, strategies etc. to foster improvement.
4. It supports risk management by pinpointing potential risks and developing plans to counter and handle it.
5. It aids cost-control by analyzing past data to find areas to reduce expenses.
6. It predicts likely future outcomes based on historical data trends for better planning ahead.
7. It ensures compliance with regulatory and reporting requirements in many industries.
8. It underpins quality control processes to meet required standards for products/services.

9. It transforms raw data into actionable insights across any field or sector.
10. It delivers factual intelligence for smarter decision-making, forecasting, optimization efforts.

Note

Data interpretation helps in Data Driven Decision Making (DDDM), which involves making decisions based on Facts and data.

After studying about various preprocessing and processing technique, we can summarise it for both data forms as given the table below.

Table 2.2 *Qualitative vs Quantitative Data Handling and Processing*

	Qualitative Data	Quantitative Data
Nature of Data	Non-numerical or categorical information, such as descriptions, opinions, observations etc.	Numerical information that can be measured or counted.
Focus	Subjective or qualitative aspects of a phenomenon.	Objective or quantitative aspects of a phenomenon.
Data Representation	Words, texts, images, or codes ; can be organized into categories, themes, or patterns.	Nnumbers or numerical values ; can be organized into tables, graphs, charts, or statistical summaries.
Data Collection Methods	Interviews, focus groups, observations, or open-ended survey questions.	Surveys, experiments, or structured observations.
Data Analysis Approach	Identifying patterns, themes, or commonalities.	Focuses on numerical relationships, patterns, or trends.
Data Analysis Techniques	Coding, content analysis, or discourse analysis.	Computations, statistical tests, modeling.
Outcome	In-depth explanations, rich descriptions, and contextual insights.	Numerical measurements, statistical relationships, and quantifiable results.
Generalizability	Not easily generalizable to a larger population (findings may be specific to the studied context).	Generalizable for a larger population.

Checkpoint

Multiple Choice Questions

1. Qualitative data deals with :
   (*a*) Descriptions and qualities (*b*) Numeric values and quantities
   (*c*) Mathematical calculations (*d*) Statistical analysis
2. Quantitative data deals with:
   (*a*) Descriptions and qualities (*b*) Numeric values and quantities
   (*c*) Opinions and feelings (*d*) Colors and characteristics
3. Which is an example of qualitative data?
   (*a*) Sales figures (*b*) Test scores
   (*c*) Customer reviews (*d*) Sensor readings
4. Which is an example of quantitative data?
   (*a*) Case studies (*b*) Focus group transcripts
   (*c*) Survey ratings (*d*) Audio recordings

10. Prescriptive analysis uses data to make projections about the future.
11. DDDM involves making decisions based on facts and data rather than intuition.
12. Customer reviews are an example of quantitative data.
13. Survey ratings are an example of qualitative data.
14. Qualitative data is collected through structured surveys.
15. Quantitative data is collected through methods like observations and interviews.
16. Qualitative data helps answer questions like "how many?" and "how much?"
17. Data types for quantitative data include numbers and percentages.
18. Qualitative data is useful for testing hypotheses and generalizing findings.
19. Data preprocessing involves analyzing the data.
20. Data cleaning is a step in data processing.
21. Data aggregation involves performing operations on the data.
22. Data visualization is an example of a data processing operation.
23. Data interpretation is important for data acquisition.

Competency Based Questions

PART B

1. Your school has collected data on the academic performance of students across different subjects and grade levels. Which of the following approaches can help teachers and administrators make informed decisions about curriculum planning and student support.
   (*a*) Create a presentation with text-based explanations of the data.
   (*b*) Create visualizations without any analysis or interpretation.
   (*c*) Analyze the data and create informative visualizations, explaining how they can aid decision-making.
   (*d*) Ignore the data and rely on personal experiences and opinions.
2. A company is analyzing customer feedback to improve its products. They collect data from online surveys, social media comments, and customer support interactions. Which of the following methods would be most effective for acquiring this diverse dataset?
   (*a*) Conducting in-person interviews with a select group of customers
   (*b*) Scraping data from random websites for a broader perspective
   (*c*) Integrating data from various sources including surveys, social media APIs, and CRM systems
   (*d*) Ignoring social media comments as they might not be relevant
3. A research team is studying the effects of climate change on biodiversity in a particular region. They have collected raw data from weather stations, satellite images, and field surveys. What is the first step they should take in preprocessing this data ?
   (*a*) Removing outliers and anomalies (*b*) Converting satellite images into numerical data
   (*c*) Aggregating data from different sources into a single dataset
   (*d*) Conducting statistical analysis on the raw data
4. A marketing team wants to analyze the purchasing behaviour of customers. They have a dataset containing customer IDs, purchase amounts, and timestamps of transactions. What preprocessing step is essential before conducting any analysis?
   (*a*) Scaling the purchase amounts to a standardized range
   (*b*) Removing duplicate entries from the dataset
   (*c*) Converting timestamps into a more readable format
   (*d*) Adding additional features such as customer demographics

Assertion & Reasoning Questions

Directions

In the following questions, a statement of assertion (A) is followed by a statement of reason (R). Mark the correct choice as :

(*a*) Both A and R are true and R is the correct explanation of A.
(*b*) Both A and R are true but R is not the correct explanation of A.
(*c*) A is true but R is false (or partly true).
(*d*) A is false (or partly true) but R is true.
(*e*) Both A and R are false or not fully true.

1. Assertion (A). Descriptive analysis summarizes data using statistics.
   Reason (R). Descriptive analysis aims to determine the conclusion about the distribution of data.
2. Assertion (A). Qualitative data deals with numeric values and quantities.
   Reason (R). Quantitative data deals with descriptions and qualities.
3. Assertion (A). Structured surveys are a method of collecting qualitative data.
   Reason (R). Qualitative data is subjective in nature.
4. Assertion (A). Qualitative data helps answer questions like "how many?" and "how much?"
   Reason (R). Quantitative data helps answer questions like "why?" and "how?"
5. Assertion (A). Quantitative data is useful for testing hypotheses and generalizing findings.
   Reason (R). Qualitative data is useful for exploring new phenomena.
6. Assertion (A). Data cleaning is a step in data preprocessing.
   Reason (R). Data preprocessing involves analyzing the data.
7. Assertion (A). Data aggregation is an example of a data preprocessing operation.
   Reason (R). Data processing involves performing operations on the data.
8. Assertion (A). Data interpretation is important for deriving outcomes.
   Reason (R). Data visualization is useful for communicating insights.

LET US REVISE

- *Data can be divided into two primary types: Qualitative data and Quantitative data.*
- *Qualitative data describes qualities or characteristics of some entity or phenomena.*
- *Quantitative data represents information about something through numerical values.*
- *Qualitative data is also known as categorical data and quantitative data is known as numerical data.*
- *Common data acquisition methods include surveys, questionnaire, interviews, observations and many others.*
- *High quality data is valid, reliable, and of ethical integrity.*
- *Data preprocessing refers to the process of making data appropriate for use by removing discrepancies in it.*
- *Some common data preprocessing techniques are : Data Cleaning, Data Reduction, Data Transformation, and Data Integration.*
- *Data cleaning is a process of identifying and correcting errors, inconsistencies, and anomalies in raw data.*
- *Data reduction is a process or a set of techniques used to reduce the size of a dataset while still preserving the most important information.*
- *Data transformation refers to the process of converting raw data into a suitable format or representation for analysis, visualization, or modeling.*
- *Data processing refers to manipulating, analyzing, and Interpreting data to extract meaningful information and derive insights.*

- *Data processing involves : Data Analysis and Data Interpretation.*
- *Data analysis is a process of apply many techniques on data to find/extract trends, correlations, outliers, and variations that convey a meaning or point to a specific result.*
- *Data analysis can be descriptive, diagnostic, predictive, and prescriptive.*
- *Descriptive analysis answers : What happened? ; Diagnostic analysis answers: Why did it happen?; Predictive analysis answers: What might happen in the future?; Prescriptive analysis answers: What should we do about it?*
- *Data interpretation involves making sense of the analyzed data, drawing conclusions, and deriving actionable insights to inform decision-making and problem-solving.*
- *Effective data interpretation is important for extracting actionable insights, informing decision-making, and driving organizational success.*

Solution Time

1. *Differentiate between qualitative and quantitative data. How are these alternatively known as?*

   Ans. Qualitative data deals with descriptions and qualities, while quantitative data deals with numeric values and quantities that can be measured. Qualitative data is also known as *Categorical data* and quantitative data is known as *Numerical Data.*

2. *Provide an example of qualitative data.*

   Ans. Customer reviews, interview transcripts, and case studies are examples of qualitative data.

3. *How is quantitative data collected ?*

   Ans. Quantitative data is collected through structured methods like surveys, sensors, and data collection instruments.

4. *Give an example of a predictive analysis question.*

   Ans. An example of a predictive analysis question is "What are the projected sales for the next quarter based on historical data trends ?"

5. *What is the purpose of data cleaning in data preprocessing ?*

   Ans. The purpose of data cleaning in data preprocessing is to remove errors, inconsistencies, and inaccuracies from the data.

6. *Give an example of a data processing operation.*

   Ans. An example of a data processing operation is data aggregation, which involves summarizing or rolling up data.

7. *Why is data interpretation important ?*

   Ans. Data interpretation is important for informed decision-making, risk management, cost control, performance assessment, and anticipating future trends based on data insights.

8. *Mention a method of data acquisition.*

   Ans. A method of data acquisition is using sensors to collect data, such as sensor readings or Internet of Things (IoT) data.

9. *Give some examples of various types of data analysis.*

   Ans. Some examples of various types of data analysis are :

   - **Descriptive.** The most popular ice cream flavors based on total sales.
   - **Diagnostic.** Customers buy more when there are promotions/discounts.
   - **Predictive.** Next year's mall traffic is expected to increase by 10% based on local population growth.
   - **Prescriptive.** Open a new store location in the fastest growing neighborhood.

10. *What is data normalization in data preprocessing ?*

Ans. Data normalization is a data preprocessing step that involves scaling or transforming data to a common range or format for consistent analysis.

Practical Session

Activity 1 : Trend Analysis : Analyzing Cafeteria Food Sales Trends

Ask students to figure out the trend of sales of food items in school canteen or cafeteria. Students can collect data on the daily or weekly sales of different food items in the school cafeteria over a period of time (e.g., one semester or one academic year). They should then analyze the data to identify trends, such as :

(*i*) *Which food items are most popular and least popular ?*

(*ii*) *How do sales vary across different days of the week or months ?*

(*iii*) *Is there an increasing or decreasing trend in the sales of certain items over time ?*

(*iv*) *How do sales trends correlate with factors like weather, school events, or holidays ?*

Solution. **Sample Data :**

The table below shows the daily sales (in units) of three food items (Pizza, Burgers, and Salads) in the school cafeteria for one month (20 school days).

Day	*Pizza*	*Burgers*	*Salads*
1	45	62	18
2	52	58	22
3	48	65	20
...	...	...	...
20	38	72	25

Interpretation and Results :

(*i*) The most popular item is **Burgers**, with the highest average daily sales of around 62 units.

(*ii*) Salads have the lowest average daily sales of around 20 units, indicating they are the least popular item.

(*iii*) There is a **slightly increasing trend** in Salad sales over the month, possibly due to health awareness campaigns or seasonal changes.

(*iv*) Burger sales peak on Fridays, while Pizza sales are relatively consistent throughout the week.

(*v*) Overall, there is a cyclical weekly pattern in sales, with higher sales on Fridays and lower sales on Mondays.

GLOSSARY

Data Cleaning A process of identifying and correcting errors, inconsistencies, and anomalies in raw data

Data Integration The process of combining data from multiple sources or formats into a unified dataset for analysis, reporting, or decision-making

Data Reduction A process or a set of techniques used to reduce the size of a dataset while still preserving the most important information

Data Transformation The process of converting raw data into a suitable format or representation for analysis, visualization, or modeling

Qualitative Data Data describing qualities or characteristics of some entity or phenomena

Quantitative Data Data representing information about something through numerical values

Assignment

1. Qualitative data helps answer questions like :
   (a) How many? (b) How much ? (c) Why? and How? (d) What percentage?
2. Quantitative data helps answer questions like:
   (a) What are the qualities ? (b) How many? and How much ?
   (c) What are the descriptions ? (d) What are the characteristics ?
3. Data types for qualitative data include:
   (a) Numbers and percentages (b) Text, audio, video, Images
   (c) Frequencies and ratings (d) Measurements and scores
4. Data types for quantitative data include :
   (a) Opinions and descriptions (b) Numbers, percentages, frequencies
   (c) Videos and images (d) Audio recordings
5. Which is an example of a data processing operation?
   (a) Data cleaning (b) Data aggregation (c) Data collection (d) Data visualization
6. Data interpretation is important for:
   (a) Informed decision-making (b) Compliance and reporting
   (c) Risk management (d) All of these
7. Descriptive analysis answers which question?
   (a) Why did it happen? (b) What might happen in the future ?
   (c) What happened ? (d) What should we do about it ?
8. Diagnostic analysis aims to answer which question?
   (a) What happened? (b) What might happen in the future?
   (c) What should we do about it? (d) Why did it happen?
9. Data acquired from sensors is an example of:
   (a) Qualitative data (b) Quantitative data (c) Descriptive data (d) Prescriptive data
10. Which of the following is a step in data preprocessing?
11. Qualitative data helps answer questions like _____ and _____ . ("why?" , "how?")
12. Quantitative data helps answer questions like _____ and _____ . ("how many?", "how much?")
13. The purpose of data cleaning in the preprocessing stage is to:
    (a) Analyze the data (b) Remove errors and inconsistencies
    (c) Interpret the data (d) Visualize the data
14. The process of combining data from multiple sources is called:
    (a) Data transformation (b) Data integration
    (b) Data cleaning (d) Data normalization
15. Which type of analysis is used to identify the root cause of a problem?
    (a) Descriptive analysis (b) Predictive analysis
    (c) Prescriptive analysis (d) Diagnostic analysis
16. Customer reviews are an example of ____ data. (qualitative)
17. Survey ratings are an example of _____ data. (quantitative)
18. Data acquired from sensors is an example of _____ data. (quantitative)

PART B

19. Data aggregation is an example of a _____ _____ operation. (data processing)
20. _____ analysis is commonly used in the finance industry. (Predictive)
21. _____ analysis provides recommendations for future actions. (Prescriptive)
22. Quantitative data deals with numeric values and quantities. (T)
23. Qualitative data deals with descriptions and qualities. (T)
24. Quantitative data helps answer questions like "why?" and "how?" (F)
25. Data types for quantitative data include text, audio, and videos. (F)
26. Data types for qualitative data include numbers, percentages, and frequencies. (F)
27. Data interpretation is important for informed decision-making. (T)
28. An example of a diagnostic analysis question is "Why did sales decrease?" (T)
29. What are the two categories of data ?
30. What is data acquisition ? What are common data acquisition methods ?
31. What is data preprocessing ?
32. Why is data cleaning important ?
33. What is data deduction ? What is its significance ?
34. Why is data transformed before processing ?
35. What is data processing ? What does it mainly do ?
36. What is data analysis and its purpose ?
37. What is data interpretation and its purpose ?
38. Write some data collection methods used for (*i*) qualitative data, (*ii*) quantitative data.
39. Write some data analysis techniques used for (*i*) qualitative data, (*ii*) quantitative data.
40. Define (*i*) predictive analysis (*ii*) diagnostic analysis. (*iii*) prescriptive analysis (*iv*) descriptive analysis.
41. Mention two methods used for qualitative data analysis.

PRACTICAL ASSIGNMENT

Activity Tracking Social Media Trends

Students can choose a particular topic, hashtag, or keyword related to their interests or current events, and collect data on the frequency of its usage or mentions on social media platforms like Twitter or Instagram over a specific period. Or they may also use the sample data given below that shows the daily frequency of tweets containing the hashtag ***#ClimateChange over two weeks.***

Day	Frequency	Day	Frequency
1	2,500	8	4,800
2	2,800	9	3,900
3	3,200	10	3,600
4	2,700	11	3,100
5	2,400	12	2,900
6	4,100	13	3,400
7	5,200	14	4,500

Using the collected data or the sample data, analyze the data to identify trends, such as :

(*i*) When did the topic/hashtag/keyword gain or lose popularity ?
(*ii*) Are there any periodic patterns or cycles in the usage trend ?
(*iii*) How do the trends correlate with real-world events or news ?
(*iv*) Can they predict future trends based on the historical data patterns ?

SESSION 3

Project Interactive Data Dashboard and Presentation

- Importance of Data Visualisation
- Introducing Tableau

3.1 INTRODUCTION

You have read about data visualisation and its importance in unit 2, session 7. Recall that Data visualisation refers to the process of representing data visually or graphically, by using visual elements like charts, graphs, diagrams and maps etc. This session will again talk about data visualisation, but in practical manner. In this session, you will learn to create different data visualisations using ***Tableau Public*** software.

3.2 IMPORTANCE OF DATA VISUALISATION

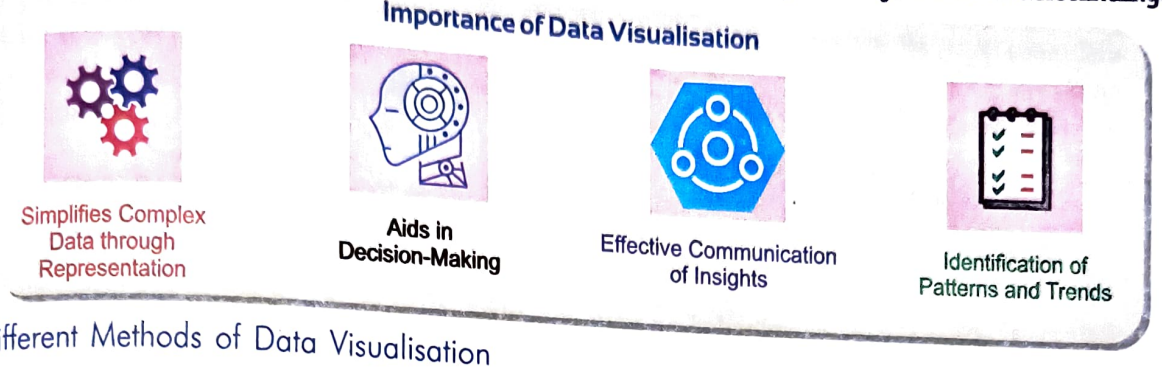
In today's world, data visualisation plays an important role in decision-making. It helps us understand information easily. Big sets of numbers and facts can be confusing. Visuals like ***charts*** and ***graphs*** make the information clear and simple. Through these visualisations, one can quickly see **patterns and trends** in the data. This helps us make good decisions based on the data.

Visualisations also help share complex information with others in a way that is easy to understand for everyone. In today's world, we have a lot of data. Data visualisation bridges the gaps between the plain raw data and the information to be driven from it.

We can summarise the significance of data visualisation as follows :

1. **Simplifies Complex Data through Representation.** Data visualisation helps in understanding complex data through visual depiction of data. It helps us find things in the information that are not obvious at first, such as *relationships among data and outliers*. Visuals like charts and graphs allow people to grasp complex data at a glance better than reading numbers and text alone.
2. **Aids in Decision-Making.** Data visualisations highlights the key points and important changes in the data, which aids in decision-making. Good visuals emphasize and **point out the most significant** *patterns, trends, and outliers* in the data through things like colours, sizes, and shapes.
3. **Effective Communication of Insights.** Visualization methods like charts, graphs, and maps can effectively summarize large datasets into a compact, high-density format. This condensed representation highlights the key insights while filtering out unnecessary details.

4. **Identification of Patterns and Trends.** Data visualisations help quickly see patterns and trends in the data, which communicate the understandings which otherwise would not have been that easy to detect. Patterns such as clusters, outliers, or trends can be easily spotted in visualizations, aiding in deeper analysis and understanding.



Different Methods of Data Visualisation

You have already about different data visualisation methods such as ***line chart, bar chart, bubble chart, XY chart, histogram*** and so on. Refer to section 7.3, session 7 of Unit 2 to recall these.

3.3 INTRODUCING TABLEAU

Tableau Public is a free software that helps you make stunning data visualizations easily. It's like magic for your data!

Imagine you have a bunch of numbers and facts. *Tableau Public* turns these into colourful pictures and charts that tell stories.

Note

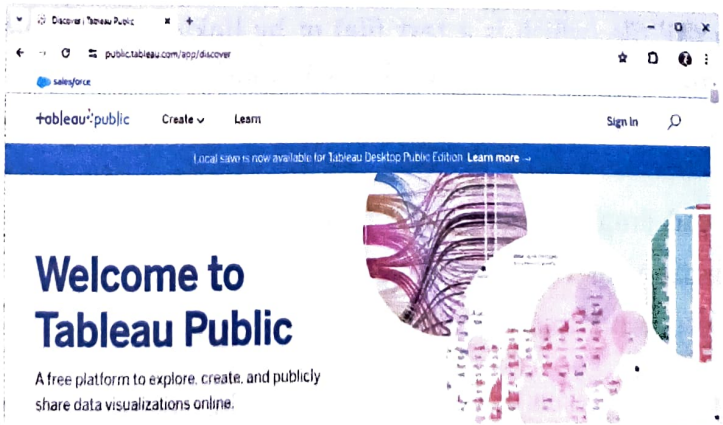
You should also known about **Data Stories** or **Data Story Telling** which are a way to convey data insights, using narratives and visualisations.

Using a narrative that has some context around data and recommended actions along with data visualisations, when you showcase the data insights, it is called a **data story** and the process of doing this is, **data story telling**.

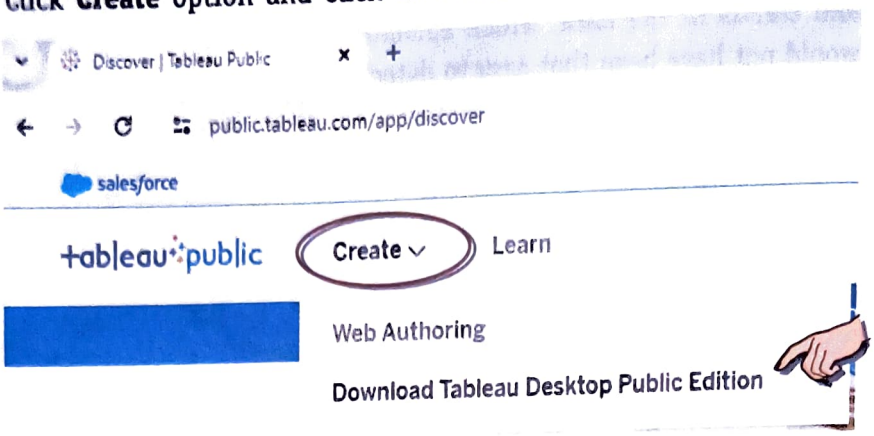
3.3.1 Downloading and Installing Tableau Public

Downloading Tableau Public is simple and free. For this, you need to do the following :

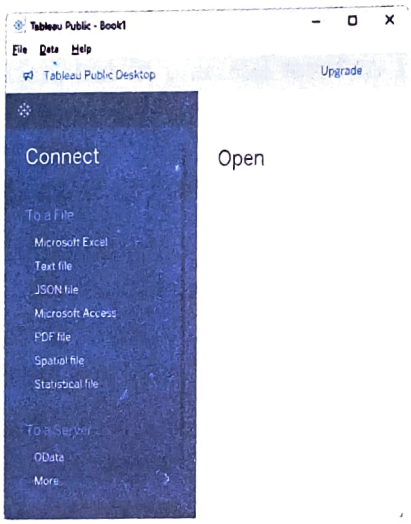
1. Go to the Tableau Public website : public.tableau.com



2. Click **Create** option and click on **Download Tableau Desktop Public Edition**



3. Once it gets downloaded on your computer, click on it and follow the instructions.
4. After the installation, open the Tableau desktop app.
5. It will show the following interface now.



PART B

3.3.2 Creating Data Visualizations in Tableau

Once you have Tableau Public installed, you can start creating visualizations right away. Let us see how. You need to follow the steps given below :

1. Open Tableau Public on your computer.
2. Now first thing you need to do is to connect your data by importing *a spreadsheet* or *a CSV file* (which is a text file) or by linking to *a database*. In the practical session, you will learn to do this practically.
3. Choose the type of visualization you want to create, like a bar chart, pie chart, or map.
4. Drag and drop your data fields onto the canvas to build your visualization.
5. Customize your visualization by adjusting colours, labels, and other settings.
6. Save your work and share it with others online.

The above given steps you will do practically in the practical session, that follows.

Practical Session

Data Visualisations using Tableau

Let us create Data visualisations from the data file vehicle sales counts that contains sales data for one year, of new and old/used vehicles. The data is stored in a text file with the name Vehicle_Sales_22-23.csv. (Please note CSV files are text files). This data file contains the following type of data :

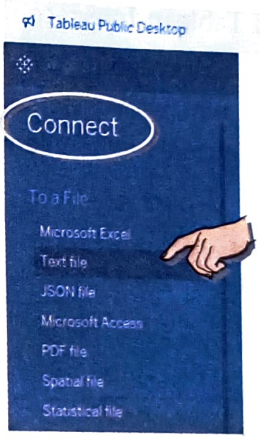
Year : The year of the sales data.
Month : The month of the sales data.
New : The number of new car sales for that month and year.
Used : The number of used car sales for that month and year.
Total Sales New : The total dollar value of new car sales for that month and year.
Total Sales Used : The total dollar value of used car sales for that month and year.

Step 1 : Understand the Data

Take a moment to familiarize yourself with the data. Go through the linked dataset. (refer to the details given above).

Step 2 : Connect to the Data Source

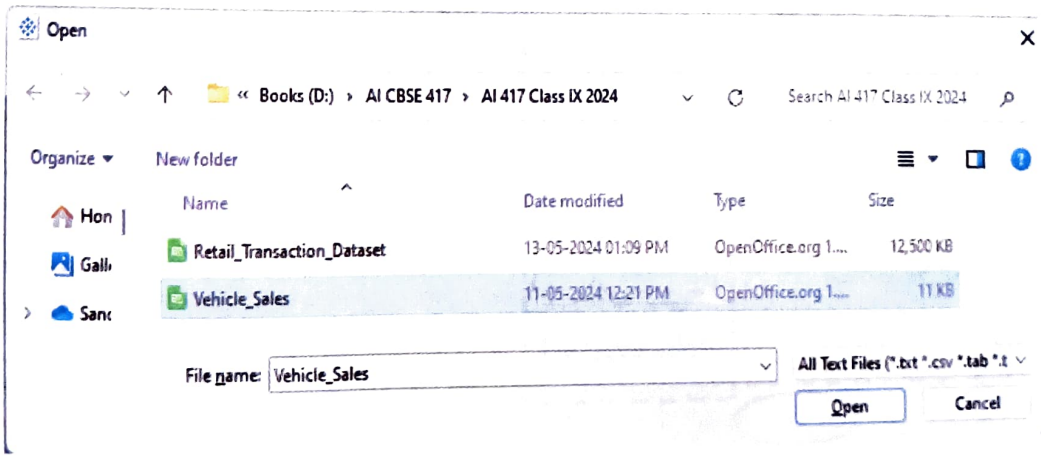
1. Open **Tableau Desktop.**



Open

2. Since our data is coming from a text file, Under **Connect,** click on **Text File.**

3. Navigate to the location where you saved the data file, select it and click **Open.**



4. Understand Tableau Terms First

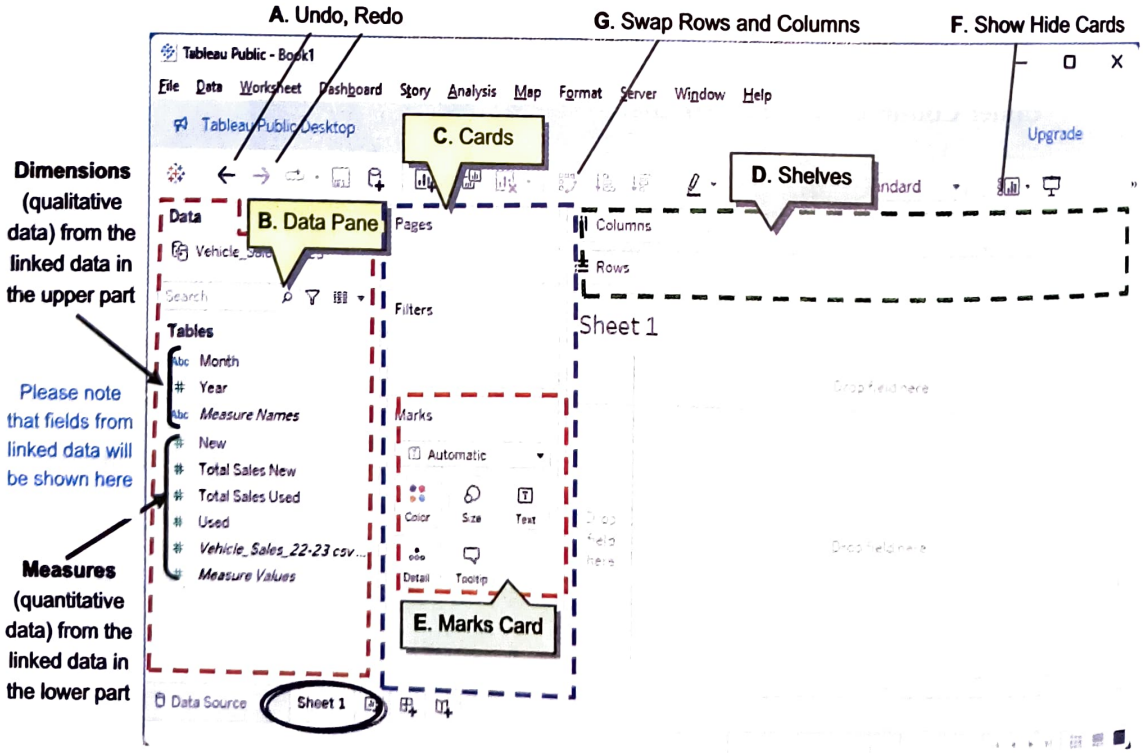
Before you start working on Tableau, you should understand its terms.

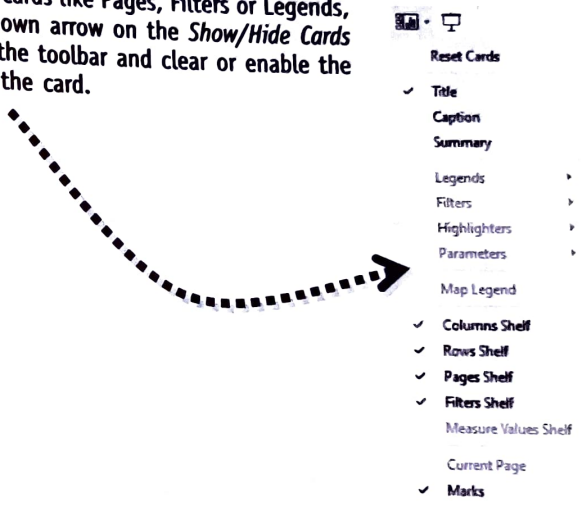
Term	Description
Aggregation	It refers to when row-level data are rolled up to a higher category, such as ***sum of sales*** or ***total profit*** or ***average sales*** etc. Tableau does aggregation automatically by offering many aggregation types/functions, such as ***Sum, Average, Maximum, Minimum*** and so on.
Dimension vs. Measure	**Dimensions** are qualitative data, such as a **name** or **date.** **Measures** are quantitative numerical data, **sales transactions** or **profit**. By default, Tableau **automatically classifies data as dimensions** or **measures**, based on their features and type. **Note** Data that is classified as a measure can be aggregated based on a given dimension, *e.g.*, **total sales (*Measure*) by region (*Dimension*).**
Continuous *vs.* Discrete	Continuous fields can contain an infinite number of values. This can be a range of values such as *sales within a specific date range* or *quantities*. Continuous fields are coloured green in Tableau. Discrete fields contain a finite number of values such as *country, state*, or *customer name*. Discrete fields are coloured blue in Tableau.
View or Viz	A view is a visualization or viz that you create in Tableau. A *viz.* could be a chart, a graph, a map, a plot, or even a text table.

PART B

5. Understand Tableau Interface

Before you start creating data visualisations, you should know about the tableau interface, its workspace etc. Following figure shows some of the commonly used workspace controls and elements of the **Sheet tab** (given at the bottom bar).

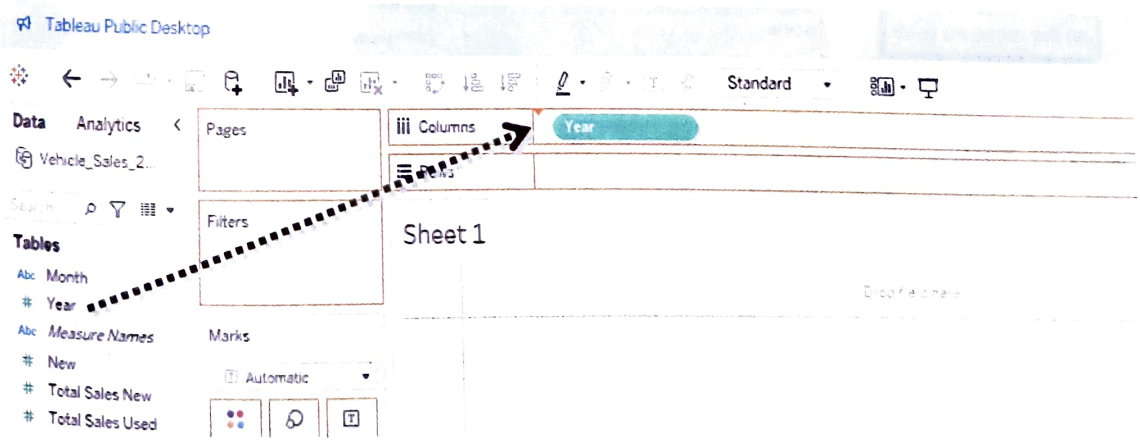


Element	Description
A – Undo and Redo	Use **Undo** to reverse the last action you performed in your workbook and **Redo** to repeat it. You can undo an unlimited number of times, back to the last time you opened the workbook, even after you saved your work.
B – Data pane	Displays the name of the data source that is included in the view, and the available fields as **dimensions** (qualitative data fields) and **measures** (quantitative data fields).
C– and D– Cards and Shelves	**Cards** are containers for the different controls that are available in Tableau. **Shelves** are a type of card control. When you drag fields from the Data pane to the **Columns** or **Rows** shelves, the data is added as a column or a row on the axes in your view.
E – Marks Card	The marks card located on the bottom shelf is one of the most commonly used tools in Tableau. It is used to give the visualization detail and context. When you drag fields from the ***Data pane*** to the ***Marks card***, you can control visual properties like type, colour, size, shape, and more. Only the marks in your view are impacted by this action; axes do not change.
F – Show/Hide Cards	To show or hide cards like Pages, Filters or Legends, click the drop-down arrow on the *Show/Hide Cards icon button* on the toolbar and clear or enable the check mark for the card. 
G – Swap Rows, Columns	Click this to swap Row with columns and vice versa, *i.e.*, it will change the row data with column data and vice versa.

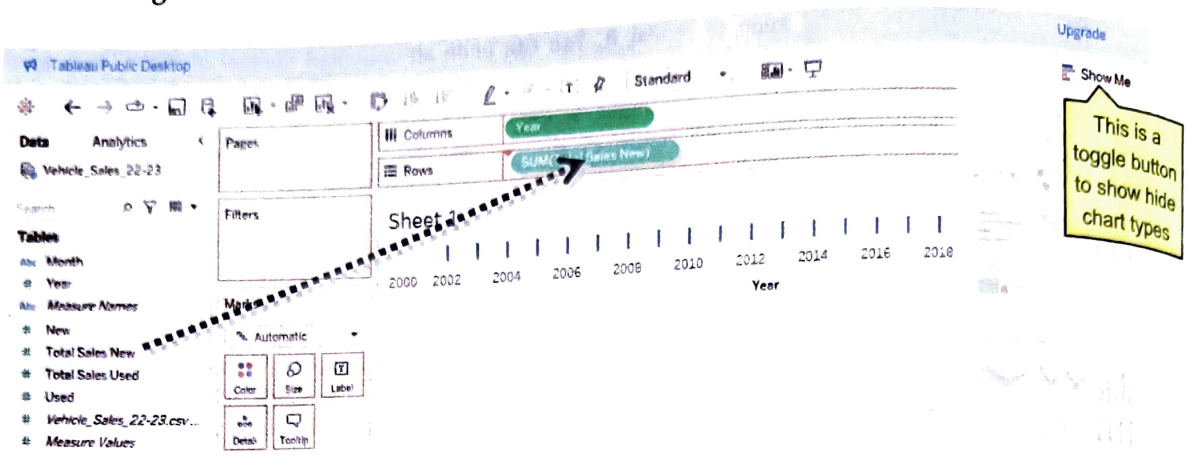
Step 3 : Creating Your First Chart

Let us now create you first chart. We shall do this by creating a proper sample chart.

1. Drag the "**Year**" *dimension* from the ***Data pane*** to the ***Columns shelf***.



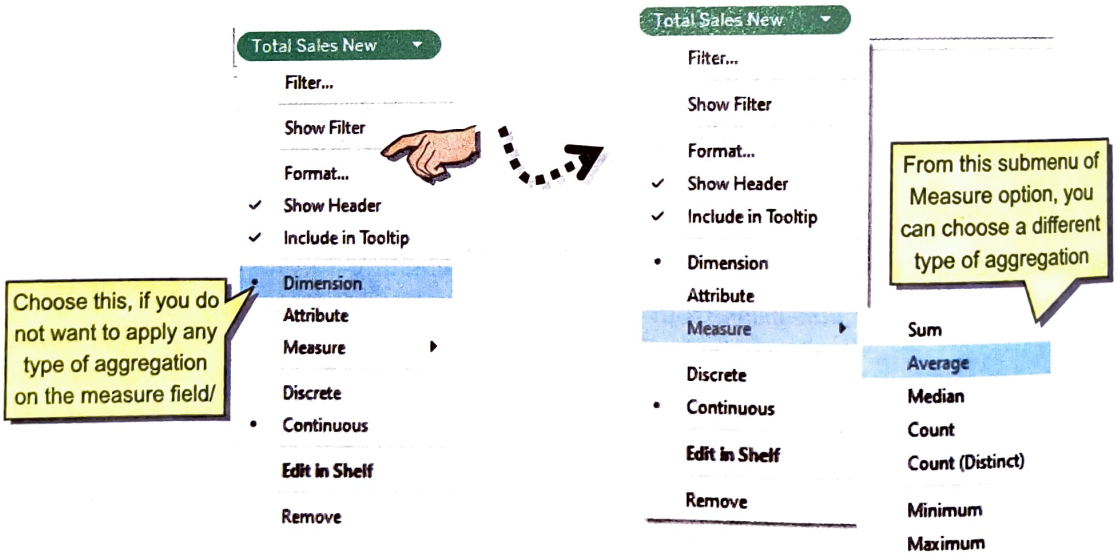
2. Drag the **"Total Sales New"** *measure* from the ***Data pane*** to the ***Rows*** *shelf.*



Notice now that Tableau automatically aggregated the dragged *measure* field as **Sum(Total_sales_new)** . If you do not want any type of aggregation function (such as *sum( ), average( )* etc.) and you want the field value as it is, or a different aggregation function/type, then do the following :

To keep the value as it is :

Click the tiny down arrow next in the aggregated field shown on the shelf (if you can't see this tiny down arrow, take your mouse cursor over it), and choose **Dimension** from the dropdown menu (see left figure below)



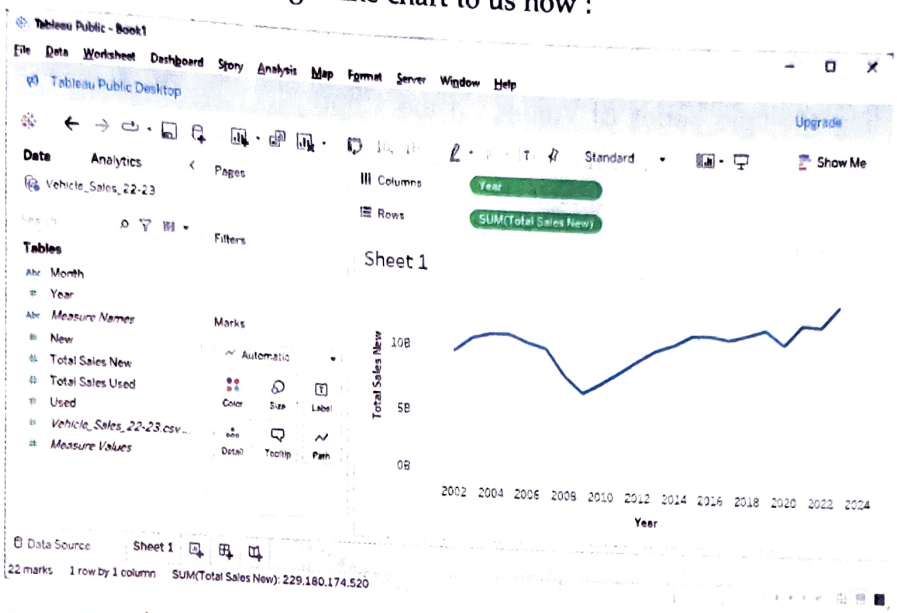
To change the aggregation function :

Click the tiny down arrow next in the aggregated field shown on the shelf, and click on **Measure** option to open its submenu. Choose the desired type of aggregation function from here. (see right figure above)

We have decided to go with the chosen aggregation, *i.e.*, **Sum(Total_Sales_new)**.

PART B

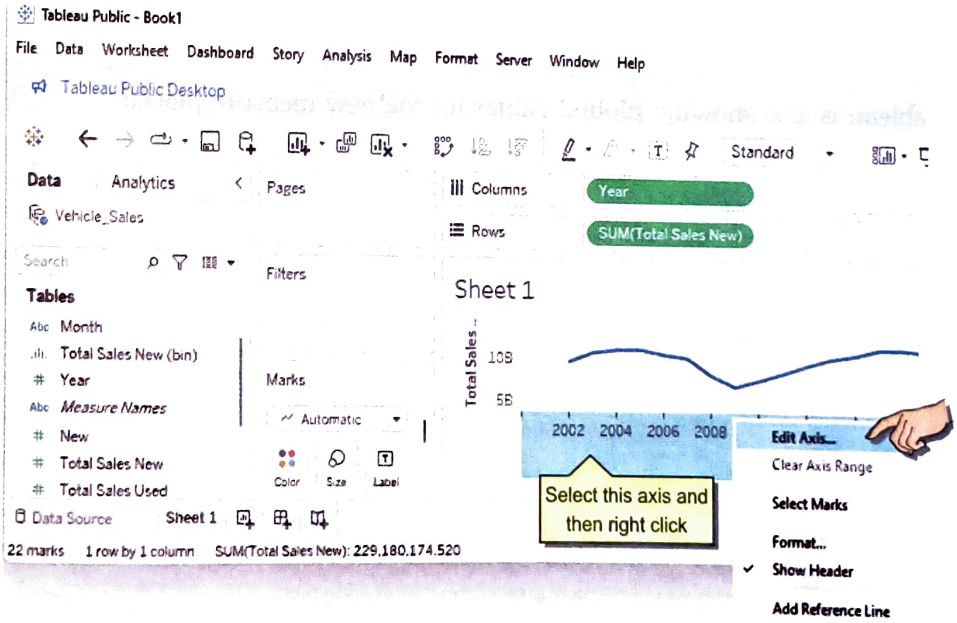
3. You should now see a chart chosen by Tableau displaying the total new car sales for each year. See it is showing a line chart to us now :



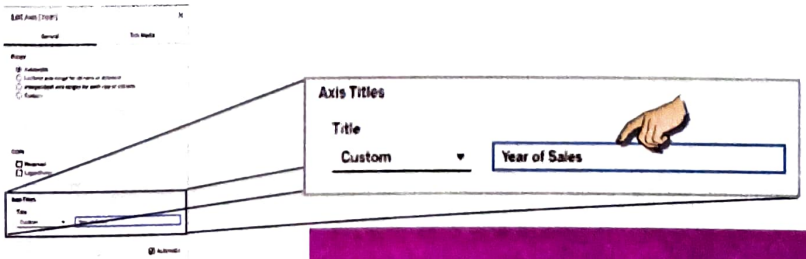
Step 4 : Customizing the Axes Title

You can now customise or format your chart. For this,

1. Click on the axis to be formatted or customised (*X*-axis or *Y*-axis) on the chart area. It will be selected and highlighted. Now right click on it and select **Edit Axis...** from the menu.



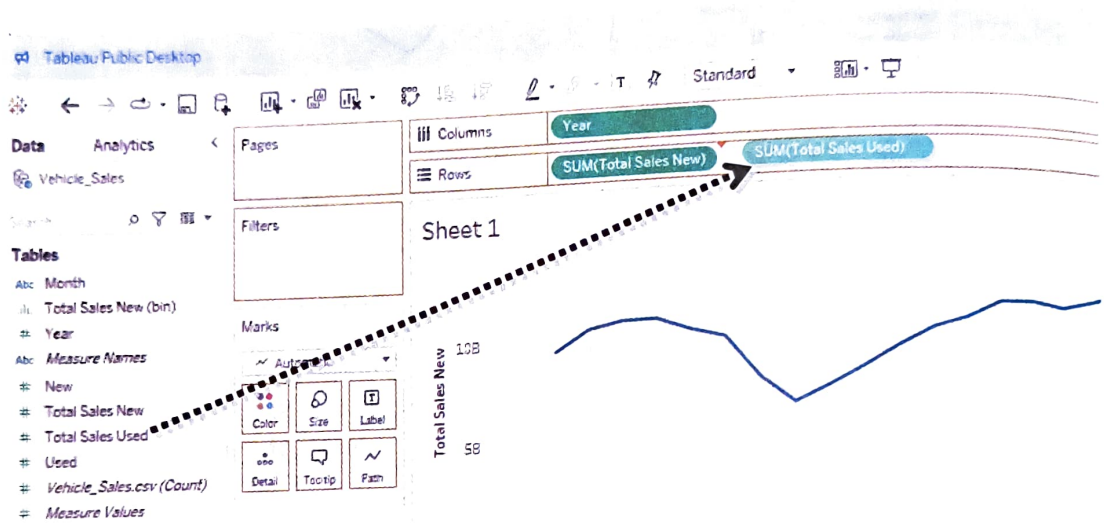
2. In the dialog that opens, we changed the title to ***Year of Sales***, by typing the custom title under the **Axis Titles** section.



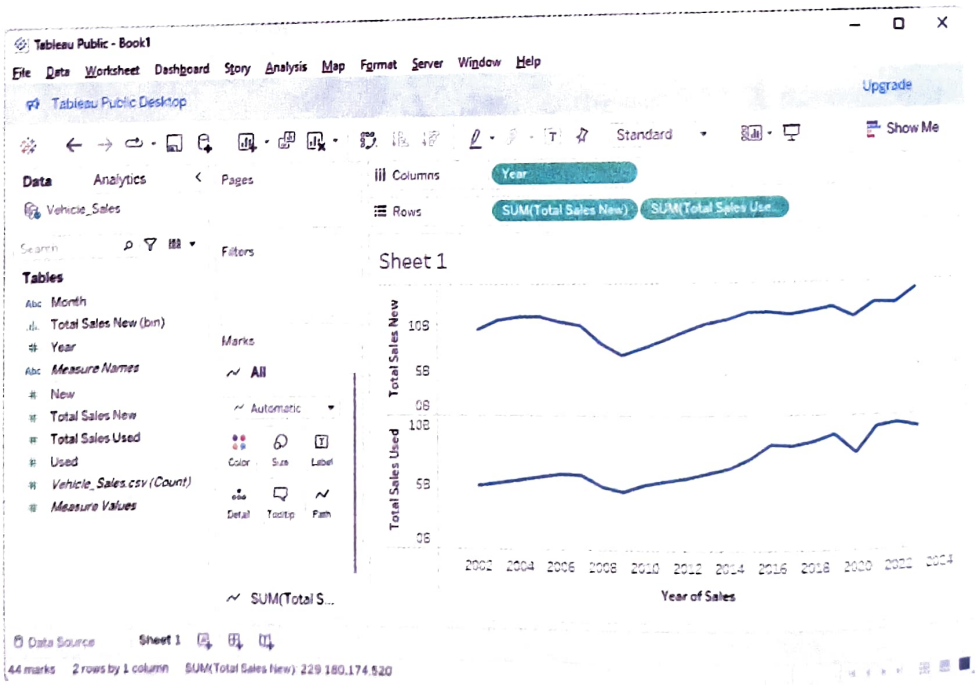
3. Similarly, you can select the vertical axis, right click on it and click Edit Axis... to customise it.
4. Close the dialog by clicking the close [X] button on top right.

Step 5 : Adding Multiple Series of Values on the Chart

Till now we have plotted only one series of value through the measure *Total_Sales_New's* sum( ). To add one more value, repeat the Step 2 (given above) with the new measure. We are adding Sum(Total_Sales_used), by dragging the measure Total_Sales_used to the Rows shelf and Tableau automatically used Sum( ) aggregation for it and places Sum(Total_Sales_used)



Now Tableau is also showing plotted values for the new measure placed



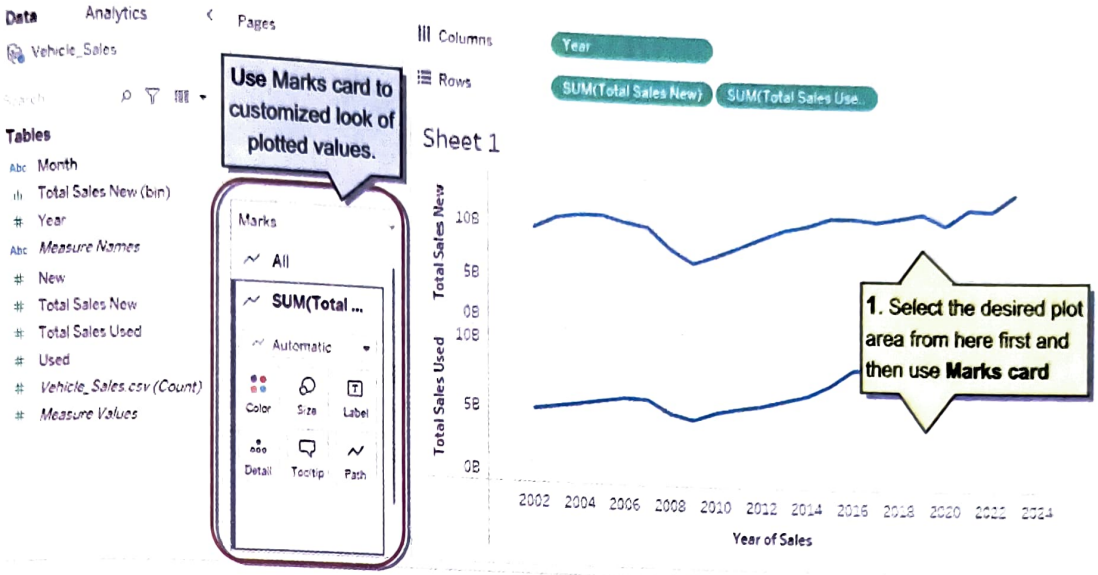
You can customise the titles for axes (vertical or horizontal) as we showed earlier.

PART B

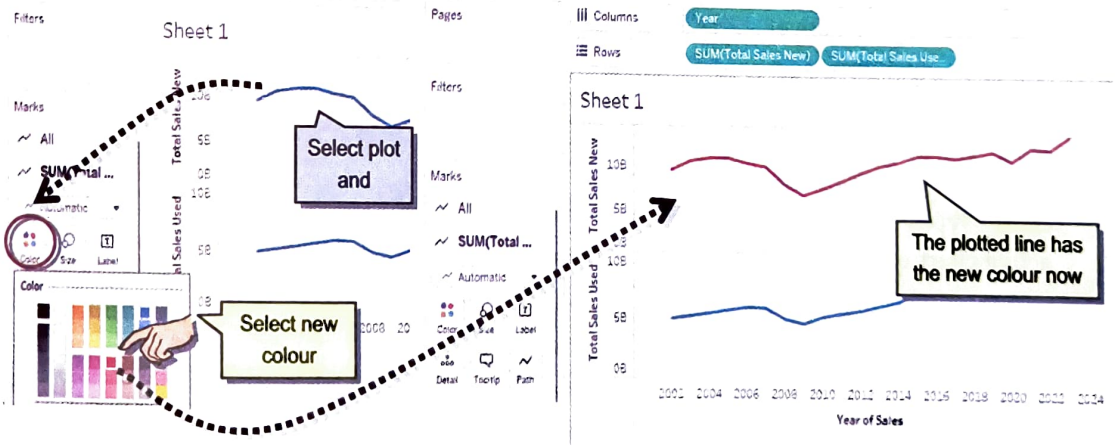
Step 6 : Customise the Plotted Values' Appearance

For this, select the plot area (of the desired ***measure***) on the chart.

You can customise the selected plot, **COLOR, Tooltip, Path** look (when the line should be dotted or dashed or full line) and more, by clicking on desired options from the **Marks Card.** (*see* below)



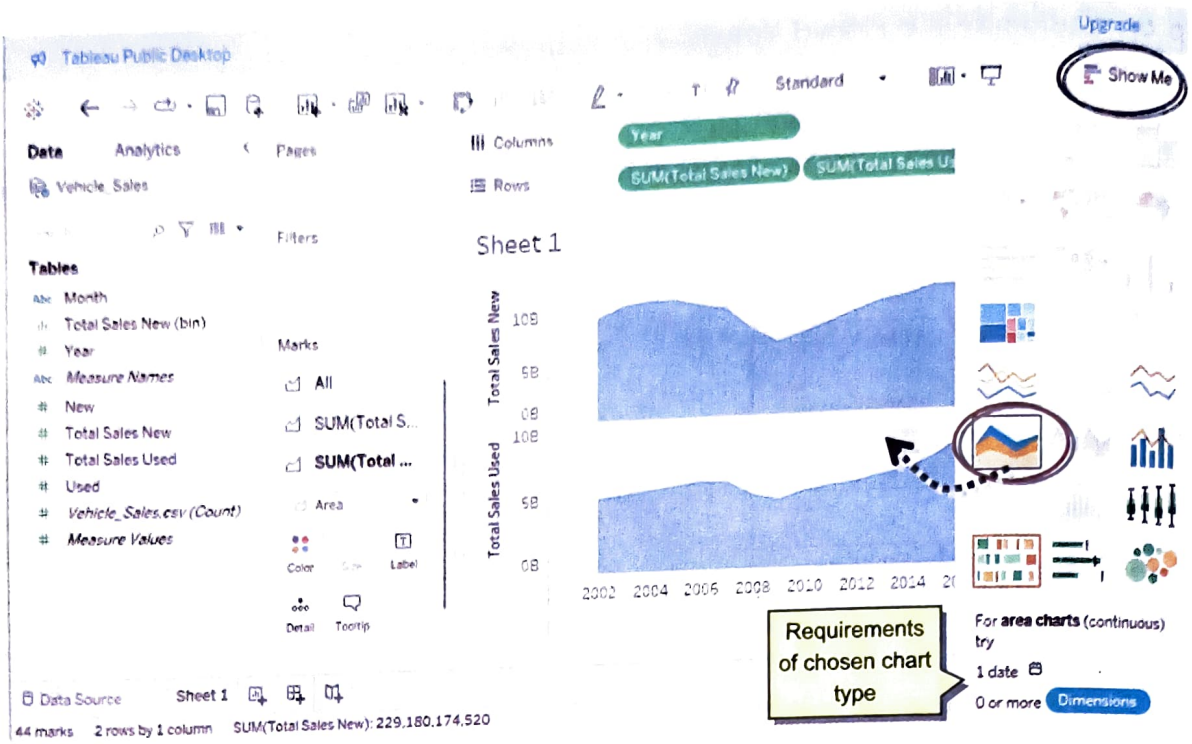
Let us change the colour for plotted line for ***Total_Sales_New*** to red. So, we first selected the plot area of ***Total_Sales_New*** and then clicked on **Color** on the **Marks card** and chose red colour. And it applied the new colour for the chosen plot.



Step 7 : Change Chart Type

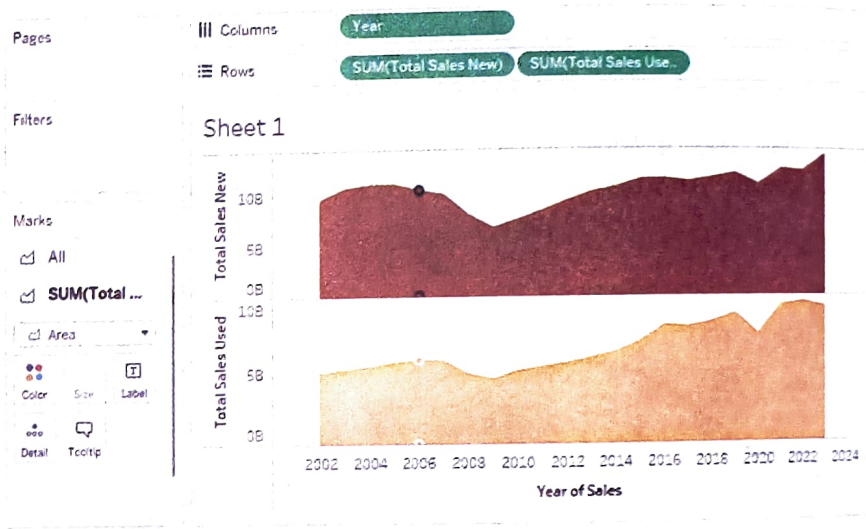
You can also change the chart type for your data visualisation. For this do the following :

1. Click on **Show Me** button (on top right) to list various chart types. Not all charts will be active, because different chart types require different combinations of *dimensions* and *measures*, (*e.g.*, Pie chart works with only one dimension and one measure, but horizontal bar chart requires multiple measures.)



PART B

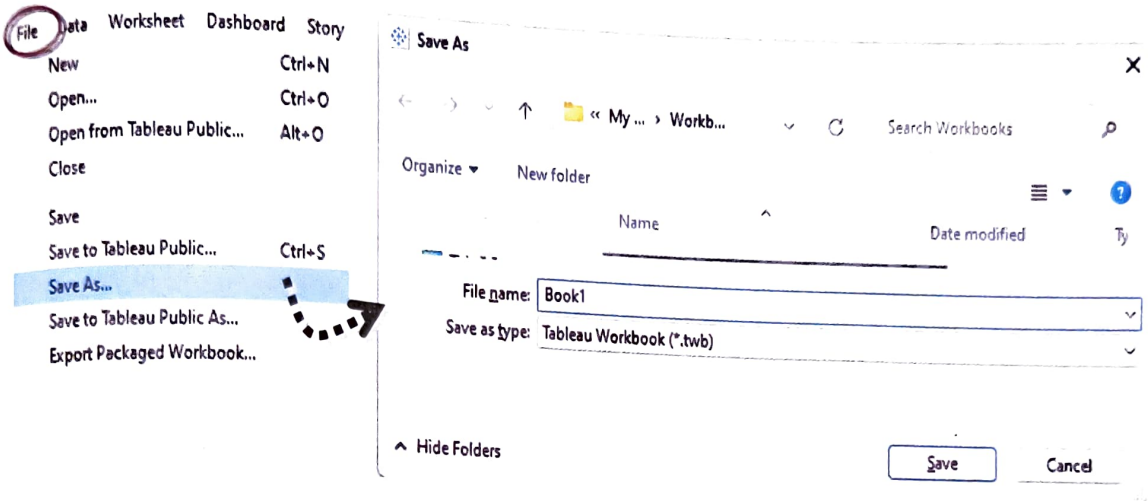
2. From the active charts, you can choose a different chart type and customise them as illustrated in step 6. For example, we chose ***Area Chart*** (see figure above) and then customised it. (see figure below)



3. You can choose different chart types from the active chart types.

Step 8 : Saving Your Work

1. Click on **File → Save As...** command.



2. Choose a location and filename for your Tableau workbook.
3. You can now share your workbook with others or continue exploring the data further.

Keep experimenting with the Tableau interface, now that you know the basics of *Data Visualisation using Tableau*.

Practical Assignment

For the following two assignments, it is advised to search for an appropriate dataset from Kaggle (www.kaggle.com). Kaggle is platform to share datasets publicly and privately. You can download a matching public dataset from here

1. **Visualizing Student Performance Data**

   Obtain data on the academic performance, such as test scores, grades, or attendance records. Then create various visualizations (*e.g.*, bar charts, line graphs, scatter plots) to represent and interpret the data, such as :

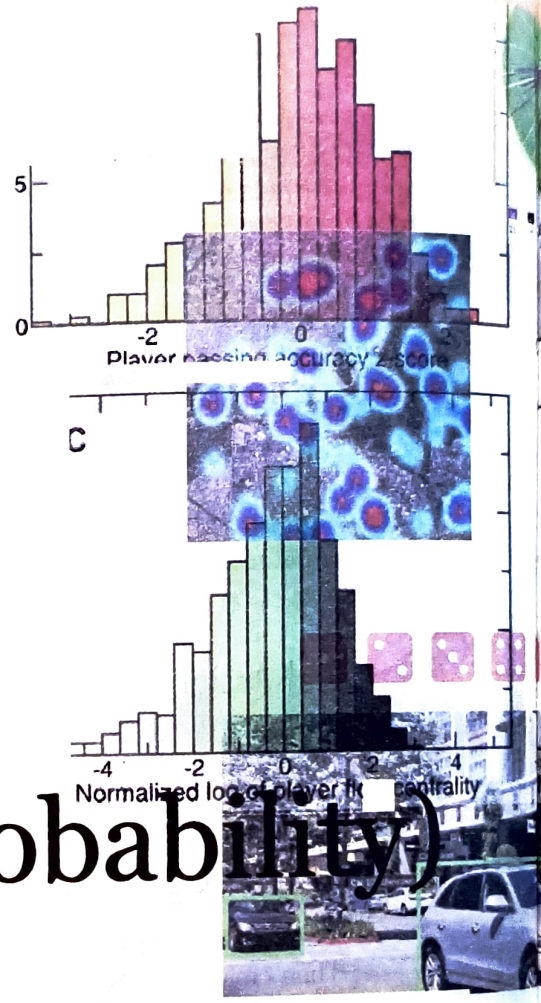
   - Comparing the performance of different subjects or topic areas
   - Identifying any correlations between factors like attendance and academic performance
   - Visualizing the distribution of scores or grades across the class or grade level
   - Tracking individual student progress over time using line graphs

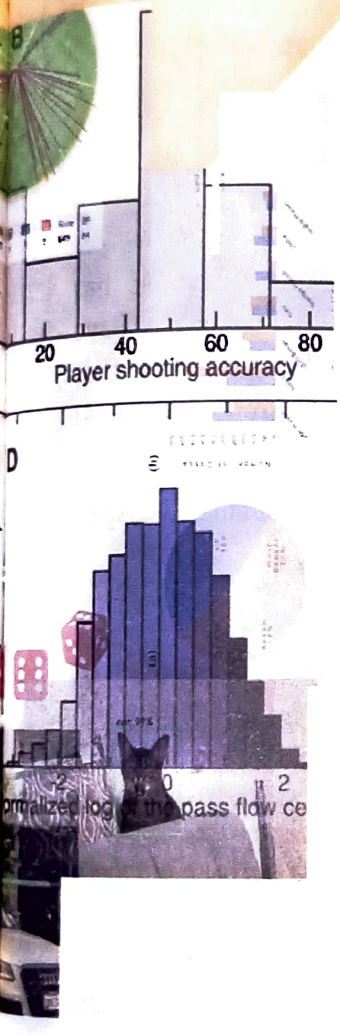
2. **Interpreting Environmental Data Visualizations**

   Explore online data sources or datasets related to environmental factors like air quality, water pollution, or carbon emissions. Then create visualizations (*e.g.*, heat maps, choropleth maps, or animated graphs) to represent and interpret the data, such as :

   - Identifying geographic hotspots or areas of concern
   - Visualizing temporal changes or trends over time
   - Comparing different regions or countries based on the environmental indicators
   - Interpreting the potential causes or impacts of the observed patterns

Math for AI (Statistics & Probability)





In This Unit

SESSION 1

Importance of Math for AI

- Finding Patterns in Numbers and Images
- Use of Math in AI
- Number Patterns, Picture Analogy, and AI

1.1 INTRODUCTION

The world you are in is undergoing industrial revolution 4.0. These days AI has transformed our lives to a reality we could not have imagined around 50 years ago. Have you ever wondered how is AI learning things, so easily ? What is the secret? Well, that secret is MATH. At the core of AI, lies a bedrock of mathematics, through which AI comprehends, learns, and predicts.

In this session, you will dive into the fascinating world of Math where numbers, images, and equations come together to make AI work its magic.

Note

At the core of AI, lies a bedrock of mathematics, through which AI comprehends, learns, and predicts.

1.2 FINDING PATTERNS IN NUMBERS AND IMAGES

Artificial Intelligence (AI) today can work with different types of data such as *numbers, images, audio, video, human interaction* and many more. From all these different types of data, AI learns and works by recognising or establishing some links in the data, which could be some common features or patterns. Whether it's analysing sales data to predict consumer trends or recognizing faces in photographs, the ability to identify patterns is fundamental to AI's functionality.

Examples of Patterns in Numbers

In numbers, patterns could be like repeating calculations, or some repeating difference of values or some mathematical function being applied and many more, *e.g.*,

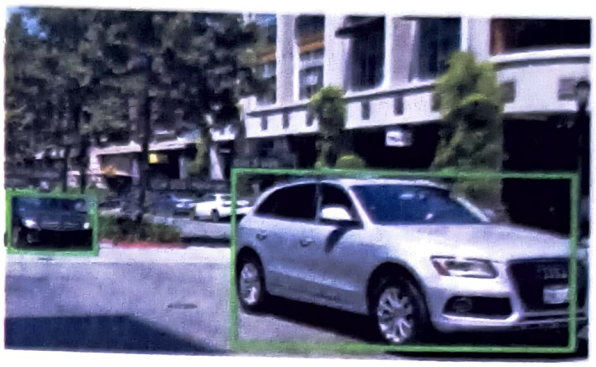
- *Consider the sequence of numbers* : 2, 4, 6, 8, ____ . The pattern is evident : each number increases by 2. Artificial Intelligence tries to find patterns in similar way.
- *Consider another sequence* : 0, 1, 2, 3, 5, 8, 13, ____ . The pattern is that 3rd number onwards, every number is the sum of previous two numbers.

Artificial Intelligence does the same thing but with tons of data, spotting patterns to make predictions or solve problems.

Examples of Patterns in Images

In images, visual patterns may be in the form of *recurring shapes, colours, or textures, e.g.,*

- AI can identify vehicles in an image by recognizing common features such as *wheels, bonnet, windshield or windscreen* etc. (see figure on the right side)



- AI can identify a cat in different images by recognizing common features such as *pointy ears* and *whiskers.* (see figure on the left side)

1.2.1 Importance of Pattern Recognition in AI

A key strength of AI models is their ability to identify patterns and relationships in data, both numerical and image-based.

- In **Numerical Data Analysis**, AI models can ***recognize trends, cycles, and correlations*** that may not be immediately apparent to humans.

  For instance, in sales forecasting, AI models can identify patterns in historical sales data, customer behaviour, and market conditions to predict future demand accurately.

- In **Image Data Analysis**, AI models can ***recognize visual patterns and relationships*** that humans may overlook or find difficult to discern.

  For example, in medical imaging, AI models can detect subtle patterns in X-ray or MRI images that indicate the presence of a particular condition or disease.

Note

A key strength of AI models is their ability to identify patterns and relationships in data, numerical, text- and image-based.

Given below are a few examples showcasing the importance of pattern recognition in AI :

- **Spam Detection.** By analysing the content and structure of emails, AI can identify patterns associated with spam messages and filter them out from your inbox.
- **Medical Diagnosis.** AI can analyse medical images such as X-rays or MRIs to detect patterns indicative of diseases like cancer or fractures.
- **Weather Forecasting.** By analysing historical weather data, AI can identify patterns to predict future weather conditions with a high degree of accuracy.
- **Financial Markets.** AI algorithms analyse stock market data to identify patterns that signal buying or selling opportunities for investors.

In each of these examples, math serves as the underlying framework that enables AI to identify and interpret patterns, leading to valuable insights and informed decisions.

1.3 USE OF MATH IN AI

Imagine AI as a super-smart robot that learns from data to make decisions or do tasks. But how does it learn? That's where math comes in handy! Math helps AI understand patterns in numbers and images, just like when you see a sequence of numbers or pictures and notice a rule or trend.

Some most-used math tools of AI are :

1. Statistics

This branch of math deals with *collecting, analysing, and interpreting* data. AI uses statistics *to summarize large datasets, identify trends, and make predictions*. For example, AI-powered recommendation systems use statistics to suggest movies or products based on your past preferences.

AI's commonly used statistical techniques include : *Classification, Clustering, Probability distributions, Hypothesis testing, Regression analysis*, and more.

Note

AI uses statistics to summarize large datasets, identify trends, and make predictions.

2. Linear Algebra

Linear algebra is the study of vectors, matrices, and linear transformations. It's like the language of AI, enabling machines to understand and manipulate data in multiple dimensions (multi-dimensional space). For instance, AI uses linear algebra *to process images, recognize patterns, and even generate realistic images from scratch.*

AI's commonly used linear algebra techniques include : *Matrix operations, Linear transformations, Regularisation*, and more.

Note

AI uses linear algebra in image processing, pattern recognition, and even in generating realistic images from scratch.

3. Probability

Probability is the study of uncertainty and randomness. AI uses probability theory to make decisions in uncertain situations. For example, self-driving cars use probabilistic models *to estimate the likelihood of different outcomes, such as whether a pedestrian will cross the road.*

AI's commonly used probability techniques include : *Bayes' theorem, Conditional probability random variables, Normalization,* and more.

Note

AI uses probability theory to analyse and eliminate uncertainty to make decisions in uncertain situations.

4. Calculus

Calculus deals with rates of change and accumulation. AI uses calculus to optimize processes and make incremental improvements over time. For instance, AI-powered search engines use calculus *to rank web pages based on their relevance to a user's query.*

AI's commonly used calculus techniques include : *Derivatives and gradients, Gradient descent, Limits and advanced logistic regressions*, and more.

Note

AI uses calculus to optimize processes and make incremental improvements over time, by adjusting parameters to minimize errors.

1.4 NUMBER PATTERNS, PICTURE ANALOGY AND AI

AI can work with numbers and images. AI looks for patterns in numbers, and analogies in pictures. Following table summarises how AI works with numbers and images, and makes use of them.

Table 1.1 *AI Working with Numbers and Images*

	AI Working with Numbers	***AI Working with Images***
What it looks for	AI studies lots of number data **to find relationships and rules.**	It tries **to see analogies, or similarities, between different pictures.**
How that helps	Patterns in numbers help AI make predictions from new number data.	Seeing analogies in pictures helps AI recognize objects, faces, or scenes.
Case Study	In a marketing research case study, a company is analysing the growth of their customer base over several months. By observing the pattern in the number of new customers each month, AI can predict future growth and plan their marketing strategies accordingly. Understanding number patterns helps make informed decisions about resource allocation and target setting.	AI using classification for animal pictures to analyse the animal behaviour. By identifying connections among images of different animals, AI can understand common traits and characteristics shared by certain groups of animals (*e.g.*, winged animals are birds). This knowledge helps them classify animals into categories and analyse their behaviours and habitats more effectively.

Let us have a look at some examples.

Example 1 *Look at the series of some numbers given below. Can you figure out the missing number? Also give the logic to support your answer.*

(*i*) 5, 10, 15, ____ , 25 (*ii*) 1, 4, 9, ____ , 25

(*iii*) 0, 1, 2, 3, 6, 11, 20, 37, ____ (*iv*) 1, 3,7, 15, 31, 63, ____

Solution.

(*i*) **It's 20.** Each number increases by 5.

(*ii*) **It's 16.** Each number is the square of consecutive integers (1^2, 2^2, 3^2, 4^2, 5^2).

(*iii*) **It's 68.** Each next number (barring first 3 numbers) is the sum of previous three numbers.

(*iv*) **It's 127.** Every next number is $2^n - 1$. First number is $2^1 - 1$, second is $2^2 - 1$, third is $2^3 - 1$... . So the seventh number would be $2^7 - 1$, which is 127.

Example 2 *Consider the pictures given below :*

Which of the following flower would fit in the last box ?

A **B** **C** **D** **E** **F**

Solution. Flower ***E*** or ***F*** would go. The pattern is that each flower is progressing in complexity.

By finding patterns in number data and analogies in picture data, AI can learn and get better at tasks like :

- Predicting things like sales or weather from number patterns
- Identifying objects, animals, or people from picture analogies
- Understanding human speech or text by recognizing language patterns (known as *natural language processing*)
- Playing games by strategizing from data patterns

and many more

Note

Looking for patterns in numbers and analogies in pictures is a core way AI learns and gets smarter over time. The better it recognizes these patterns, the more accurate its predictions and decisions become.

Under solved problems, you will find some practical activities dedicated to finding patterns in numbers, and analogies in pictures.

Check Point

Multiple Choice Questions (MCQs)

1. What role does math play in artificial intelligence (AI) ?
   (*a*) It is not important (*b*) It forms the foundation of AI
   (*c*) It is only used in advanced AI projects (*d*) It is used only for simple tasks in AI
2. What is the purpose of finding patterns in numbers and images for AI ?
   (*a*) To make AI more complicated (*b*) To slow down AI's learning process
   (*c*) To enable AI to make predictions and decisions
   (*d*) To confuse AI algorithms
3. Which branch of math deals with summarizing and analysing data ?
   (*a*) Algebra (*b*) Geometry (*c*) Statistics (*d*) Calculus

4. How does AI use linear algebra ?
   (*a*) To understand shapes and colours
   (*b*) To manipulate data in multiple dimensions
   (*c*) To calculate probabilities
   (*d*) To analyse historical weather data

5. What is the purpose of probability theory in AI ?
   (*a*) To eliminate uncertainty
   (*b*) To make decisions under uncertainty
   (*c*) To avoid making decisions
   (*d*) To make decisions based on facts only

6. Which of the following is an example of pattern recognition in images ?
   (*a*) Identifying numbers in a captcha
   (*b*) Solving algebraic equations
   (*c*) Calculating the area of a circle
   (*d*) Translating text from one language to another

7. Which of the following is an application of AI in healthcare ?
   (*a*) Predicting stock prices
   (*b*) Identifying diseases in medical images
   (*c*) Recommending movies
   (*d*) Analysing social media data

8. What is the primary purpose of probability theory in AI ?
   (*a*) To eliminate uncertainty
   (*b*) To predict future events
   (*c*) To analyse historical data
   (*d*) To manipulate data in multiple dimensions

9. How does AI use probability in self-driving cars?
   (*a*) To estimate the likelihood of accidents
   (*b*) To detect traffic signals
   (*c*) To measure distances between vehicles
   (*d*) All of the above

10. Which of the following is an example of pattern recognition in natural language processing ?
    (*a*) Identifying spam emails
    (*b*) Predicting stock prices
    (*c*) Translating text from one language to another
    (*d*) Analysing customer reviews

11. How does calculus contribute to the optimization of AI algorithms in robotics?
    (*a*) By calculating probabilities
    (*b*) By analysing spatial relationships
    (*c*) By adjusting parameters to minimize errors
    (*d*) By generating 3D models

12. How does AI use linear algebra in machine learning?
    (*a*) To predict future events
    (*b*) To analyse historical data
    (*c*) To represent data in multidimensional space
    (*d*) To perform statistical analysis

Fill in the Blanks

1. AI learns from _____ to make decisions.
2. Finding patterns in numbers and images helps AI to make _____.
3. _____ provides methods for analysing and interpreting data.
4. _____ helps AI to represent and manipulate data in multiple spaces.
5. _____ theory enables AI to reason under uncertainty.
6. Linear algebra enables AI to represent data in _____ spaces.
7. Calculus helps AI in optimizing algorithms by calculating _____.

8. AI algorithms use statistics to analyze data and identify _____.
9. Probability theory helps AI to make decisions based on probabilistic outcomes in _____ situations.
10. AI uses _____ to summarize large datasets.

True/False Questions

1. Math is not important for AI.
2. Linear algebra is not used in AI.
3. Probability theory is used to eliminate uncertainty in AI.
4. Calculus is only used for basic calculations in AI.
5. Finding patterns in numbers and images is not useful for AI.
6. AI algorithms uses statistics for finding trends in large datasets.
7. Calculus help AI in minimising errors by adjusting parameters.

Competency-Based Questions

Students in a school are working on various AI based projects, such as :

- Group A of students is developing an AI-based assistance called **Heya** to assist fellow classmates with math homework and tests.
- Group B of students is working on developing a model that can identify images based on certain features and distinguish it from surroundings.

PART B

1. What tool/competency will the **Heya** use for understanding the users' inputs and commands?
   (*a*) Multidimensional space (*b*) Probability
   (*c*) Finding linguistic patterns (*d*) Summarising data
2. **Heya** keeps track of earlier tests taken by students and gives them their summarised performance. Which of the following tools help it do this?
   (*a*) Linear Algebra (*b*) Statistics
   (*c*) Probability (*d*) Calculus
3. **Heya** identifies a pattern of performance from the students' past tests taken. What does this demonstrate?
   (*a*) Eliminating uncertainty (*b*) Finding trend
   (*c*) Recognising patterns (*d*) Process optimisation
4. Group B's tool can identify human figures in images. It can even work with somewhat blurred images. Which of the following tools help it identify the human figure image from the possible shapes it identified?
   (*a*) Linear Algebra (*b*) Statistics
   (*c*) Probability (*d*) Calculus
5. Group B's tool was used with a robot so that if sees a human, it raises hand to greet. Which of the following, in your opinion, is the most prominent feature(s) used in it?
   (*a*) Linear Algebra (*b*) Statistics
   (*c*) Probability (*d*) Calculus

Assertion & Reasoning Questions

Directions

In the following questions, a statement of assertion (A) is followed by a statement of reason (R). Mark the correct choice as :

(*a*) Both A and R are true and R is the correct explanation of A.
(*b*) Both A and R are true but R is not the correct explanation of A.
(*c*) A is true but R is false (or partly true).
(*d*) A is false (or partly true) but R is true.
(*e*) Both A and R are false or not fully true.

1. Assertion (A). Linear algebra is essential for AI.
   Reason (R). Linear algebra provides tools for representing and manipulating data in multiple dimensions, which is crucial for tasks like image processing and machine learning in AI.
2. Assertion (A). Probability theory is fundamental in AI-based recommendation systems.
   Reason (R). Probability theory enables AI to calculate the likelihood of different events and make informed decisions, which is essential for filtering and personalizing recommendations for users in uncertain situations.
3. Assertion (A). Calculus plays a significant role in training AI.
   Reason (R). Calculus is a useful tool to calculate gradients.
4. Assertion (A). Natural Language Processing through AI is all about identifying linguistic patterns in text data, facilitating tasks like sentiment analysis, language translation, and spam detection.
   Reason (R). Pattern recognition is vital for AI in natural language processing.
5. Assertion (A). AI works by identifying correlations, making predictions, or drawing conclusions about the underlying data distribution.
   Reason (R). Statistics is used in AI to minimize errors in correlations.

Let Us Revise

- *At the core of AI, lies a bedrock of mathematics, through which AI comprehends, learns, and predicts.*
- *A key strength of AI models is their ability to identify patterns and relationships in data, numerical, text- and image-based.*
- *AI uses statistics to summarize large datasets, identify trends, and make predictions.*
- *AI uses linear algebra to process images, recognize patterns, and even generate realistic images from scratch.*
- *AI uses probability theory to analyse and eliminate uncertainty to make decisions in uncertain situations.*
- *AI uses calculus to optimize processes and make incremental improvements over time, by adjusting parameters to minimize errors.*

Solution Time

1. *What are some math tools used by AI?*

   Ans. Some math tools used by AI are :

   Statistics. Helps AI make sense of data by summarizing and analysing it. For example, AI can predict weather patterns based on past data.

Linear Algebra. Deals with lines, planes, and shapes in multiple dimensions. AI uses this to understand relationships between different data points.

Probability. Helps AI make guesses based on likelihoods. For instance, AI in self-driving cars calculates the probability of a pedestrian crossing the road.

Calculus. Studies change and motion. AI uses calculus to optimize processes, like finding the shortest route on a map.

2. *How does AI use statistics to analyse data ?*

Ans. AI uses statistics to summarize and interpret data, identify trends, and make predictions based on patterns observed in the data.

3. *Give an example of a pattern in numbers and explain how AI can identify it.*

Ans. **Example :** A sequence of numbers increasing by 3 each time (*e.g.*, 2, 5, 8, 11, ...). AI can identify this pattern by observing the consistent increase of 3 between consecutive numbers.

4. *What is the purpose of linear algebra in AI ?*

Ans. Linear algebra helps AI represent and manipulate data in multiple dimensions, enabling tasks such as image processing, dimensionality reduction, and machine learning.

5. *Explain how probability theory helps AI make decisions.*

Ans. Probability theory enables AI to quantify uncertainty and make decisions based on probabilistic outcomes. By assigning probabilities to different events and updating them as new information becomes available, AI can make informed decisions in uncertain environments.

6. *How does calculus contribute to the optimization of AI algorithms ?*

Ans. Calculus helps AI optimize algorithms by calculating gradients and derivatives, allowing AI to adjust parameters to minimize errors and improve performance over time.

7. *Give an example how AI uses patterns in numbers and uses the learning to solve a problem.*

Ans. In a biology class, students are studying the growth patterns of bacterial colonies over time. By observing the number of bacteria present at regular intervals, they can identify trends and predict future growth rates. Understanding number patterns helps them make hypotheses about factors affecting bacterial growth and design experiments to test these hypotheses.

Practical Session

Practical Activity 1 : *Observing Number Patterns*

Objective : *To develop ability to recognize number patterns and find missing numbers within those patterns.*

Given some series of numbers below. Try to identify patterns and correlations of terms in them and give the missing number.

(*i*) Series 1: 1, 1, 2, 3, 5, 8, ___ , 21, 34

(*ii*) Series 2: 2, 3, 5, 7, ___ , 13, 17, 19, 23

(*iii*) Series 3: 1, 4, 9, ___ , 25, 36, 49, 64, 81

Solution.

(*i*) This sequence is generated by adding the two previous numbers to get the next number. So, 13 is the missing number.

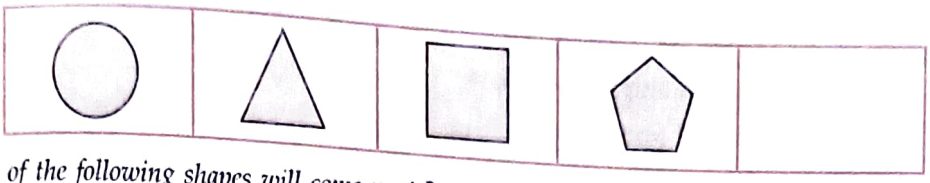
(*ii*) These are *Prime Numbers*. Prime numbers are numbers greater than 1 that have no positive divisors other than 1 and itself. So, 11 is the missing number.

(*iii*) These are *Square numbers*, which are the result of multiplying a number by itself. So, 16 is the missing number.

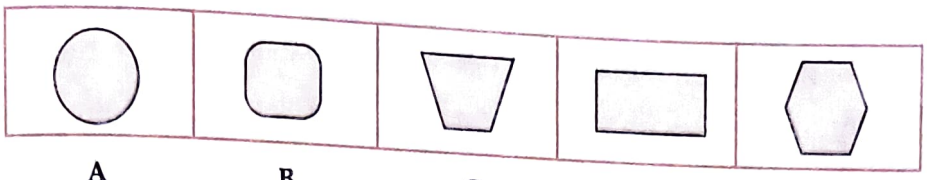
Practical Activity 2 : *Finding Connections Among Images*

Objective : *To develop ability to identify connections among images.*

Observe the set of shapes given below :



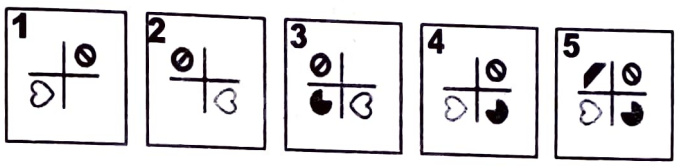
Which of the following shapes will come next ?



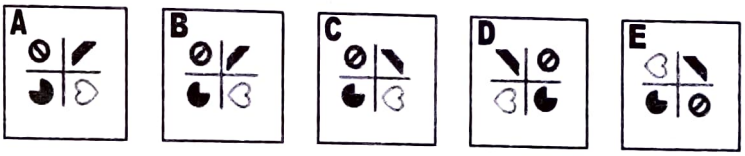
Solution. The shape E will come. The number of sides of the shapes is increasing by one each time. So, Image 5 should be a hexagon.

Practical Activity 3 : *Finding Connections Among Images*

Choose the image that completes the pattern :



Which of the shapes below continuous the sequence :



Solution.

The next frame after each step portrays a mirror image of the previous frame. In addition, every two steps a shape is added to the frame. Answer choice C is a mirrored version of frame 5 and is therefore the correct answer.

Assignment

1. Which branch of mathematics is essential for representing and manipulating data in multiple dimensions in AI ?

   (*a*) Algebra (*b*) Geometry (*c*) Calculus (*d*) Linear Algebra

2. What role does probability theory play in AI ?

   (*a*) Eliminating uncertainty (*b*) Finding patterns

   (*c*) Minimising errors (*d*) All of the above

3. Which mathematical concept helps AI algorithms optimize model parameters through gradient descent ?
   (*a*) Statistics (*b*) Calculus (*c*) Probability (*d*) Algebra
4. How does AI use statistics?
   (*a*) To represent data in multiple dimensions
   (*b*) To analyse shapes and properties
   (*c*) To summarize and interpret data
   (*d*) To reason under uncertainty
5. What mathematical concept enables AI to reason under uncertainty ?
   (*a*) Algebra (*b*) Geometry
   (*c*) Probability (*d*) Calculus
6. How does AI use linear algebra in image processing ?
   (*a*) To represent images as matrices (*b*) To analyse audio data
   (*c*) To generate 3D models (*d*) To perform text analysis
7. Which mathematical branch provides methods for collecting, organizing, and analysing data in AI ?
   (*a*) Algebra (*b*) Geometry
   (*c*) Statistics (*d*) Calculus
8. What is the primary purpose of calculus in AI?
   (*a*) To calculate probabilities
   (*b*) To find the shortest path between nodes
   (*c*) To optimize algorithms by adjusting parameters
   (*d*) To analyse spatial relationships
9. Linear algebra helps AI to represent and manipulate data in _____ spaces. (multiple)
10. Probability theory enables AI to reason under _____. (uncertainty)
11. Calculus helps AI optimize algorithms by calculating _____. (gradients)
12. AI algorithms use statistics to analyse data and identify _____. (trends)
13. Calculus contributes to the optimization of AI algorithms by adjusting _____. (parameters)
14. Statistics plays a crucial role in AI by providing methods for _____ large datasets. (analysing)
15. AI algorithms use statistics for analysing and summarising data. (T)
16. Probability theory is used to eliminate uncertainty in AI. (F)
17. Calculus is only used for basic calculations in AI. (F)
18. Finding patterns in numbers and images one prominent helpful feature for AI. (T)
19. Probability theory helps AI to make decisions based on facts only. (F)
20. Calculus is not used in self-driving cars. (F)
21. Linear algebra is only used for representing shapes in AI. (F)
22. How does linear algebra contribute to AI ?
23. Explain the role of probability theory in AI.
24. How does calculus help in optimizing AI algorithms ?
25. What is the significance of statistics in AI ?
26. Describe an example of finding patterns in numbers and images for AI.

PRACTICAL ASSIGNMENT

Activity 1 Exploring Number Patterns

Identify, analyse and state the number patterns from the multiple series of numbers given below and give the missing number

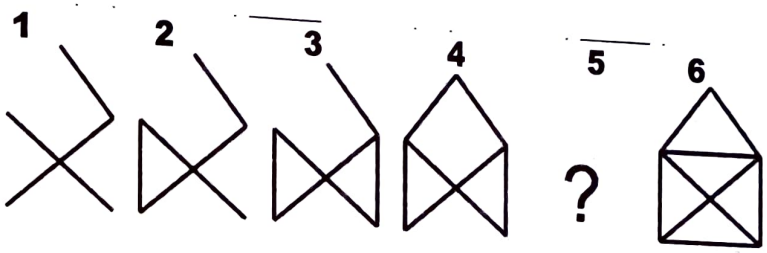
Series 1 : 3, 6, 9, ____ , 15, 18, 21, 24, 27 (12)

Series 2 : 2, 4, 8, 16, ____ , 64, 128, 256 (32)

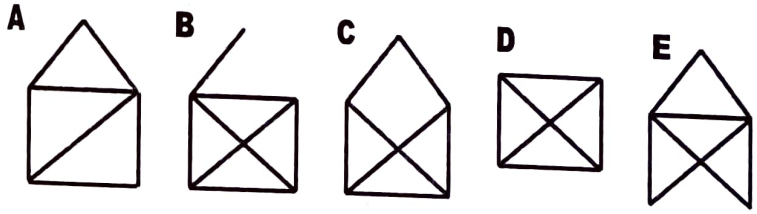
Series 3 : 1, 3, 6, 10, 15, ____ , 28, 36, 45 (21)

Activity 2 Exploring Image Patterns

Consider the images given below :



Which of the shapes below continuous the sequence :



Also give reason to support your answer.

SESSION 2

Statistics in Real Life

- What is Statistics ?
- Applications of Statistics

2.1 INTRODUCTION

You have learnt about some common math tools used in AI. Statistics is one of those important math tools, which helps AI accomplish a lot of difficult things. Statistics helps make sense of all the data and information. In this session, we'll explore the concept of statistics and its applications in real-life scenarios, along with some practical examples.

2.2 WHAT IS STATISTICS ?

Statistics is a branch of mathematics that deals with collecting, analyzing, interpreting, and presenting data. It involves ***gathering information, organizing it in a meaningful way, and drawing conclusions based on the patterns and trends observed in the data.*** Whether it's calculating the average score of a test, or determining the probability of rain tomorrow, statistics helps us make informed decisions by providing valuable insights into numerical information.

Statistics

1. Statistics is a field in which people work with real-world data.
2. The gathering, sorting, construction, and display of raw data are all covered by statistics.
3. Statistics is an applied science.

Note

Statistics is a branch of mathematics that deals with collecting, analyzing, interpreting, and presenting data, which helps us understand numerical information and make informed decisions based on it.

Why do we need Statistics ?

In our daily lives, we come across a lot of data or information. Statistics helps us :

- organize this data in a simple way
- understand patterns and trends in the data
- make decisions based on the data

Statistics helps us summarize and describe data in a meaningful way. Following table briefly summarises some basic statistical concepts and functions.

Table 2.1 *Basic Statistical Concepts*

Concept	Function	Purpose	Example
Frequency Distribution Table	Organizes data by showing how often each value or category occurs	Helps summarize and visualize the data	**Height (cm):** 140-144 (1), 145-149 (1), 150-154 (3), etc. In this group, most people have a height between 150-154 cm.
Mean	Calculates the *arithmetic average* of a data set	Measures the central tendency or typical value	Mean of 5, 7, 9, 11 = (5 + 7 + 9 + 11) / 4 = 8
Median	Finds *the middle value* in an ordered data set (takes average if there are more than one such values)	Measures the central tendency or middle value	Median of 5, 7, 9, 11 = 8 (average of 7 and 9, as there are 2 middle values : 7, 9)
Mode	Identifies *the most frequently occurring value*(s)	Measures most common value(s)	Mode of 5, 7, 9, 9, 11 = 9 (appears twice)
Range	Calculates the difference between the highest and lowest values	Measures the spread or dispersion of the data	Range of 5, 7, 9, 11 = 11 – 5 = 6

2.3 APPLICATIONS OF STATISTICS

Statistics has a wide range of applications in many real-life situations, *including disaster management, sports, disease prediction, weather forecast*, and many more.

2.3.1 Disaster Management

When a natural disaster strikes, such as an earthquake or hurricane or flood, we need data to assess things like :

- How many people are affected ?
- How many areas are affected ?
- How much damage has happened ?

For this, statistics are used ***to analyse data on population density, building structures, and infra-structure vulnerabilities, and assess the extent of damage*** (*e.g.*, see the heatmap created from disaster data in figure 2.1), estimate the number of affected people, and plan rescue operations.

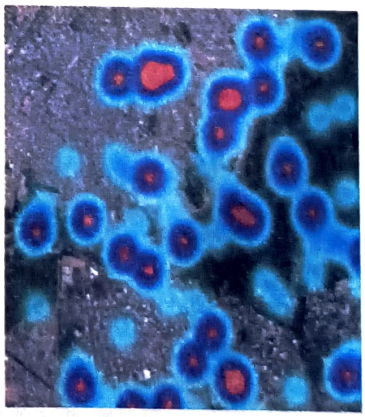


Figure 2.1 Heatmap generated from disaster data (a sample situation)

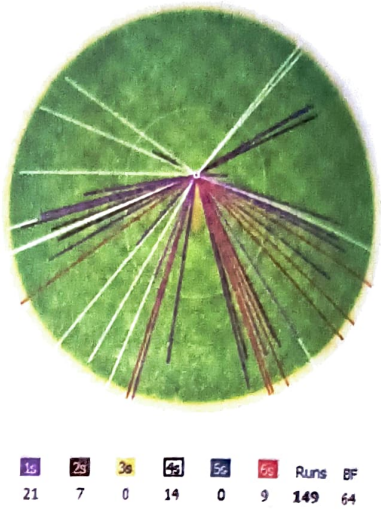
By analyzing historical data and current trends, emergency workers can

- understand the situation
- make informed decisions (such as, where to send help first ?, where to setup temporary shelters, how many rescue teams are needed?, and so on)

to minimize the impact of disasters

2.3.2 Sports

Statistics play a crucial role in sports. In sports, statistics tell us how well a player or team is performing. Scores, goals, runs - these are all statistics. Statistics helps *sports analysts, coaches and players* study these numbers for things like



- to evaluate players' strengths and weaknesses,
- identify key performance indicators,
- giving analytics about chosen parameters
- identify areas of improvement

and make strategic decisions during matches and for future games.

Whether it's calculating batting averages in cricket or shooting percentages in basketball, statistics provide valuable insights for coaches, players, and fans alike.

Note

Statistics helps in sports, from analyzing players' performance to predicting the outcome of games.

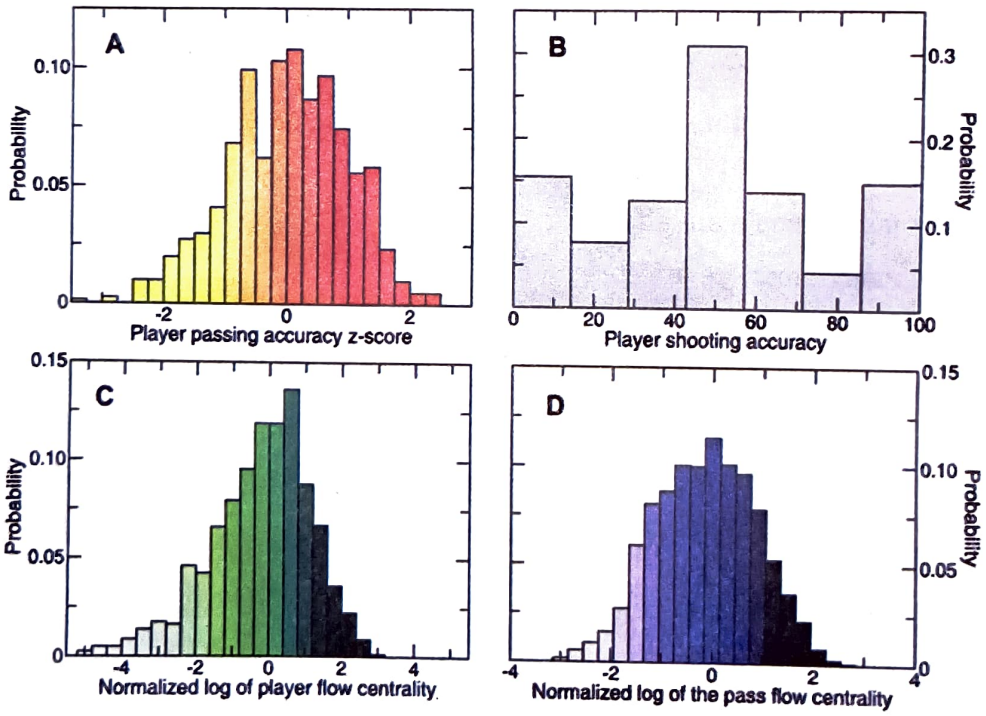


Figure 2.2 Statistics used in studying individual player's performance

Note

For example, in basketball, coaches use advanced analytics to evaluate players' shooting percentages, rebounding rates, and defensive efficiency, allowing them to identify areas for improvement and tailor their tactics accordingly.

PART B

2.3.3 Disease Prediction

In the field of healthcare, statistics are used to predict the spread of diseases, assess the effectiveness of treatments, and identify risk factors, and so on. Doctors use statistics to understand diseases and their outbreaks.

- How many people are sick ?
- What are their symptoms ?
- How frequent is the disease occurring ?
- Which areas are getting affected ?
- Which type of population (*e.g.*, young/old) is getting affected ?
- Which type of areas (*e.g.*, closed buildings, urban/rural etc.) are getting affected ?

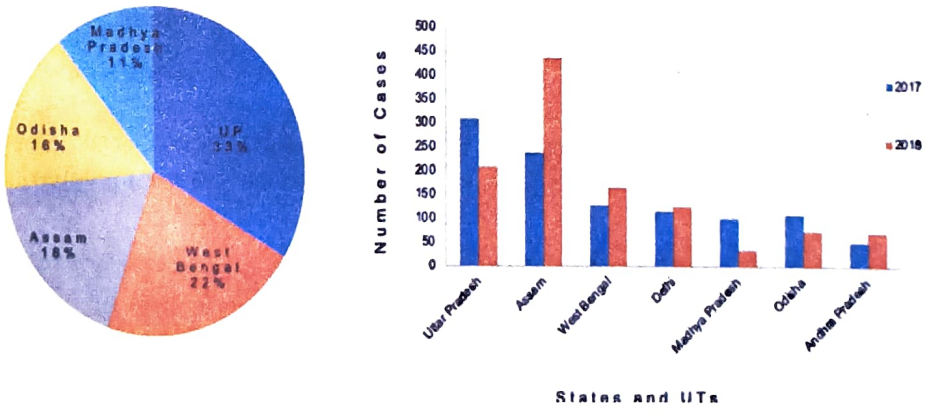
and so on.

Epidemiologists analyse data on *disease outbreaks, population demographics, and environmental factors* using statistics ***to find and understand the patterns and trends in disease data.***

This helps in

- developing models that can predict the likelihood of future outbreaks
- making and informing public health policies.
- taking proactive measures to prevent the spread of infectious diseases
- improving public health outcomes.

So, statistics can help predict disease outbreaks by identifying risk factors. It allows planning to control the spread of diseases.



(*a*) (*b*)

Figure 2.3 Sample statistical data related to some disease spread.

Note

During the COVID-19 pandemic, public health authorities used statistical models to track the spread of the virus, case counts, hospitalizations, and deaths. Using statistics, epidemiologists were able to identify hotspots of transmission, target interventions to high-risk populations, and guide policy decisions to mitigate the impact of the pandemic.

2.3.4 Weather Forecast

Statistics plays an important role in predicting about future weather. Meteorologists rely on statistical models to analyze weather data. Weather data is collected from weather stations, satellites, and other sources.

Using this data, meteorologists can :

- identify patterns and trends in temperature, precipitation, and atmospheric pressure
- develop mathematical models that simulate the behaviour of the atmosphere
- predict future weather conditions and patterns days or even weeks in advance
- issue weather forecasts to the public.

From planning outdoor activities to preparing for severe weather events, weather forecasts based on statistical analysis help individuals and communities make informed decisions to stay safe and protect their property.

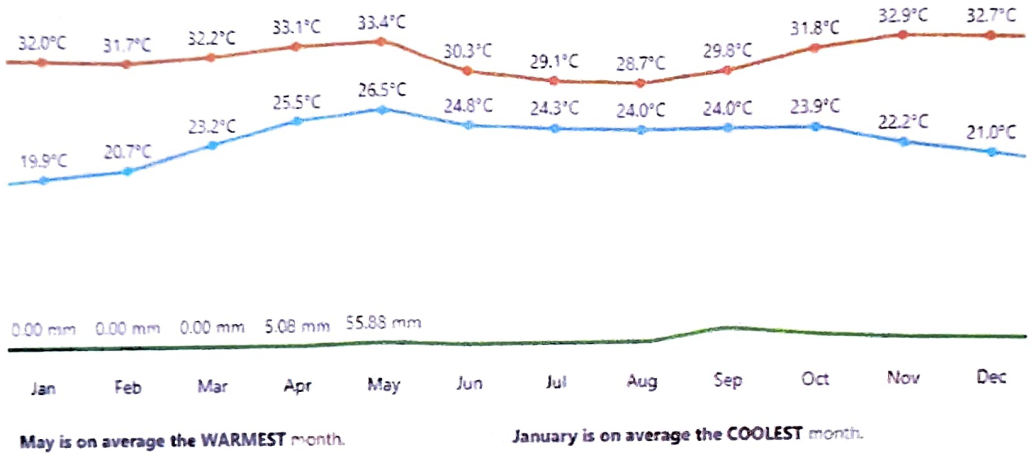


Figure 2.4 Sample statistical data about weather pattern in a city.

Note

Weather forecasters collect data on temperature, rainfall, wind speed and more. They use statistics to analyse patterns. This helps them predict the weather accurately.

Example ***Activity 1*** **:** Uses of Statistics

Let's look around and find examples where statistics is used.

Solution. Statistics can be used in many real-life situations.

Some of these could be :

- At a restaurant (number of customers, popular dishes)
- In a shopping mall (number of visitors, best-selling items)
- At school (number of students, test scores)
- On the roads (types of vehicles, which roads have heavy traffic)

and many more.

PART B

Check Point

Multiple Choice Questions

1. What is the primary purpose of statistics ?
   (*a*) To collect and analyze data
   (*b*) To solve mathematical equations
   (*c*) To study historical events
   (*d*) To predict future outcomes
2. In which field do sports analysts use statistics extensively ?
   (*a*) Analysing a player's performance
   (*b*) Identifying key factors
   (*c*) Highlighting facts
   (*d*) All of these
3. What role does statistics play in disease prediction ?
   (*a*) Identifying risk factors
   (*b*) Analyzing weather patterns
   (*c*) Designing new medications
   (*d*) Building bridges
4. How do meteorologists use statistics in weather forecasting ?
   (*a*) To analyze the density of forests
   (*b*) To test the quality of rain water
   (*c*) To analyze weather patterns
   (*d*) To design new buildings
5. In the disaster management field, statistics are used primarily for :
   (*a*) Planning rescue operations
   (*b*) Designing new buildings
   (*c*) Analyzing sports performance
   (*d*) Taking people to areas
6. What is the significance of data collection in statistics ?
   (*a*) It forms the basis for analysis
   (*b*) It is not important
   (*c*) It helps in designing new products
   (*d*) It is used for playing video games
7. How do epidemiologists use statistics in disease surveillance ?
   (*a*) To determine number of beds in hospitals
   (*b*) To check medical staff competence
   (*c*) To monitor disease trends
   (*d*) All of these

Fill in the Blanks

1. Statistics is the study of collecting, analyzing, interpreting, and presenting _____.
2. In sports, statistics are used to analyze players' performance and predict the outcome of _____.
3. Meteorologists use statistical models to analyze _____ data and issue weather forecasts.
4. Epidemiologists use statistics to monitor disease _____ and detect outbreaks.
5. Data collection is essential in statistics as it forms the basis for _____ and interpretation.
6. Statistics plays a crucial role in _____ management by assessing the extent of damage and planning rescue operations.
7. In disease prediction, statistics help in identifying _____ factors and assessing the likelihood of outbreaks.
8. Sports analysts use statistics to evaluate players' _____ and make strategic decisions during games.

True / False Questions

1. Statistics is primarily used for studying historical events.
2. Meteorologists use statistical models to analyze weather data and predict future weather conditions.
3. In sports, statistics cannot be used to analyze players' performance.
4. Disease prediction involves studying the risk factors for disease outbreak.
5. Data collection is not necessary for statistical analysis.
6. Sports analysts do not use statistics to predict the outcome of games.
7. Epidemiologists use statistics to monitor disease trends.

Competency-Based Questions

1. In the field of healthcare, Statistics plays role in assessing the extent of diseases and their outbreaks for _____ .
   (*a*) To predict the spread of diseases (*b*) Assess the effectiveness of treatments
   (*c*) To identify risk factors (*d*) All of these
2. Sports analysts use statistics to analyse players' performance. Which of the following is not a use of statistics?
   (*a*) Identify key performance indicators (*b*) Giving analytics about chosen parameters
   (*c*) Identify areas of improvement (*d*) Compressing files and archiving them
3. In disaster management, how do emergency responders use statistics to plan rescue operations ?
   (*a*) Analyse data on population density (*b*) Analyse infrastructure vulnerabilities
   (*c*) Assess the extent of damage (*d*) All of these
4. Meteorologists use statistics to analyze weather data, create statistical models to issue weather forecasts. Which of the following is not an application of that ?
   (*a*) Identify patterns and trends in temperature, precipitation, and atmospheric pressure
   (*b*) Print weather reports from saved copies
   (*c*) Develop mathematical models that simulate the behavior of the atmosphere
   (*d*) Predict future weather conditions and patterns days or even weeks in advance

Assertion & Reasoning Questions

Directions

In the following questions, a statement of assertion (A) is followed by a statement of reason (R). Mark the correct choice as :

(*a*) Both A and R are true and R is the correct explanation of A.
(*b*) Both A and R are true but R is not the correct explanation of A.
(*c*) A is true but R is false (or partly true).
(*d*) A is false (or partly true) but R is true. (*e*) Both A and R are false or not fully true.

1. Assertion (A). Statistics plays a crucial role in disaster management.
   Reason (R). By analyzing data on the extent of damage and estimating casualties, emergency responders can plan rescue operations effectively.

2. Assertion (A). Sports analysts use statistics to predict the outcome of games.
   Reason (R). By analyzing players' performance and historical data, sports analysts can forecast the results of upcoming matches.
3. Assertion (A). Meteorologists rely on statistical models to issue weather forecasts.
   Reason (R). Statistical analysis of weather data is useful for data visualization.
4. Assertion (A). Epidemiologists use statistical models to analyze disease data and predict the likelihood of outbreaks.
   Reason (R). Disease prediction uses medical science techniques and not statistics.
5. Assertion (A). Data collection is not necessary for statistical analysis and is not a part of statistics.
   Reason (R). Data collection forms the basis for statistical analysis and interpretation.

LET US REVISE

- *Statistics is a branch of mathematics that deals with collecting, analyzing, interpreting, and presenting data, which helps us understand numerical information and make informed decisions based on it.*
- *Statistics has a wide range of applications in many real-life situations, including disaster management, sports, disease prediction, weather forecast, and many more.*

Solution Time

1. *How is statistics applied in disaster management ?*

   Ans. Statistics is used to assess the extent of damage, estimate casualties, and plan rescue operations during natural disasters such as earthquakes and hurricanes. By analyzing data on population demographics, infrastructure vulnerabilities, and historical disaster events, emergency responders can make informed decisions to mitigate the impact of disasters and save lives.

2. *Give an example of how statistics is used in sports.*

   Ans. Sports analysts use statistics to analyze players' performance, predict the outcome of games, and make strategic decisions during matches. For example, in basketball, statistics such as shooting percentages, rebounds, and assists are used to evaluate players' contributions to their teams and identify areas for improvement.

3. *How do epidemiologists use statistics in disease prediction ?*

   Ans. Epidemiologists use statistical models to analyze disease data, identify patterns and trends, and predict the likelihood of future outbreaks. By collecting data on disease incidence, prevalence, and risk factors, epidemiologists can develop mathematical models that simulate the spread of infectious diseases and inform public health policies to prevent and control outbreaks.

4. *Explain the significance of data collection in statistics.*

   Ans. Data collection is essential in statistics as it forms the basis for analysis and interpretation. By collecting relevant data from various sources, statisticians can identify patterns, trends, and relationships in the data, leading to valuable insights and informed decision-making.

   Without accurate and reliable data, statistical analysis would not be possible, and the conclusions drawn from it would be unreliable.

5. *How do meteorologists use statistics in weather forecasting ?*

Ans. Meteorologists use statistical models to analyze weather data, identify patterns and trends, and predict future weather conditions.

By collecting data from weather stations, satellites, and other sources, meteorologists can develop mathematical models that simulate the behavior of the atmosphere and forecast weather patterns days or even weeks in advance. These forecasts help individuals and communities prepare for severe weather events and make informed decisions to protect lives and property.

6. *How does statistics contribute to disease surveillance ?*

Ans. Statistics plays a crucial role in disease surveillance by monitoring disease trends, detecting outbreaks, and assessing the effectiveness of control measures. Epidemiologists use statistical methods to analyze disease data collected from healthcare facilities, laboratories, and other sources, allowing them to identify patterns and trends in disease incidence, prevalence, and mortality.

By detecting unusual spikes or clusters of cases, epidemiologists can investigate potential outbreaks and implement targeted interventions to prevent the spread of infectious diseases and protect public health.

7. *Explain how statistics is used in predicting the outcome of sports events.*

Ans. Sports analysts use statistical models to analyze historical data on teams' performances, players' statistics, and other relevant factors to predict the outcome of sports events. By analyzing trends and patterns in the data, sports analysts can identify key factors that influence the outcome of games and develop predictive models that assign probabilities to different outcomes. These models help sports enthusiasts make informed decisions when betting on sports events and enable coaches and teams to strategize and prepare for upcoming matches.

PART B

Practical Session

Practical Activity : *Car Spotting and Tabulating*

In this activity, you will collect data and learn data organization.

(*i*) *Go outside and note down the number of cars of different colors that pass by in 10 minutes.*

Make a table like this and fill it up

Color	***Number of Cars***
Red	
Blue	
White	
:	

(*ii*) *Calculate the total number of cars you spotted.*

(*iii*) *For each color, calculate what percentage of the total it represents.*

(*iv*) *From the table, what do you think is the most popular colour for cars ?*

Solution.

(*i*) Collected sample data is :

Color	Number of Cars
Red	8
Blue	10
White	24
Black	15
Silver	8
Gray	5
Green	3
Yellow	2
Other	5

(*ii*) Total number of cars spotted = Sum of (number of Cars) = 80

(*iii*) The percentage of total cars for each color :

For **Red :** Percentage = (No. of Red Cars / Total No. of Cars) × 100
= (8 / 80) × 100 = 10%

For **Blue :** Percentage = (No. of Blue Cars / Total No. of Cars) × 100
= (10 / 80) × 100 = 12.5%

For **White :** Percentage = (No. of White Cars / Total No. of Cars) × 100
= (24 / 80) × 100 = 30%

For **Black :** Percentage = (No. of Black Cars / Total No. of Cars) × 100
= (15 / 80) × 100 = 18.75%

For **Silver :** Percentage = (No. of Silver Cars / Total No. of Cars) × 100
= (8 / 80) × 100 = 10%

For **Gray :** Percentage = (No. of Gray Cars / Total No. of Cars) × 100
= (5 / 80) × 100 = 6.25%

For **Green :** Percentage = (No. of Green Cars / Total No. of Cars) × 100
= (3 / 80) × 100 = 3.75%

For **Yellow :** Percentage = (No. of Yellow Cars / Total No. of Cars) × 100
= (2 / 80) × 100 = 2.5%

For **Other :** Percentage = (No. of Other Colored Cars / Total No. of Cars) * 100
= (5 / 80) × 100 = 6.25%

(*iv*) To find the most popular colour, we will find **mode,** *i.e.*, the most occurring value, which is **24,** for **white**. Thus, the most popular car colour is **white.**

Assignment

1. What of the following is not a use of statistics ?
   (*a*) To collect and analyze data
   (*b*) To organize data in a simple way
   (*c*) To study historical events
   (*d*) To understand trends in data
2. In which field do analysts use statistics extensively in real time (when it is happening) ?
   (*a*) Healthcare
   (*b*) Engineering
   (*c*) Sports
   (*d*) Agriculture
3. What role does statistics play in disaster management ?
   (*a*) Determining the no. of people affected
   (*b*) Determining the no. of areas affected
   (*c*) Identifying damage extent
   (*d*) All of these
4. Which of the following is not an application of statistics in sports ?
   (*a*) Evaluate players' strengths and weaknesses
   (*b*) Identify key performance indicators
   (*c*) Predicting future
   (*d*) Analytics
5. Which of the following situations can make use of statistics ?
   (*a*) Identifying patterns from historical data
   (*b*) Handling uncertain situations
   (*c*) Designing optimal solutions
   (*d*) Decision making based on data

PART B

6. _____ is the study of collecting, analyzing, interpreting, and presenting data. (Statistics)
7. In sports, statistics are used to analyze players' _____ and predict the outcome of games. (performance)
8. Meteorologists use _____ to analyze weather data and issue weather forecasts. (statistical models)
9. Statistics plays a crucial role in : disaster management by assessing the extent of _____ and planning rescue operations. (damage)
10. In disease prediction, statistics help in identifying risk factors and assessing the likelihood of _____. (outbreaks)
11. Sports analysts use statistics to evaluate players' performance and make _____ decisions during games. (strategic)
12. Weather forecasters rely on statistical models to analyze weather data and predict future _____ conditions. (weather)
13. Statistics is primarily used for storing data. (F)
14. Statistical models help in predicting future weather conditions, using past weather data. (T)
15. In sports, statistics can ruin a players' performance. (F)
16. Disease prediction does not involve the use of statistics. (F)
17. Statistics are just numbers. They cannot predict the outcome of games. (F)
18. Weather forecasters rely heavily on statistical models to issue weather forecasts. (T)
19. Statistics is useful for the field of medicine as it helps to monitor disease trends. (T)
20. What is the role of statistics in disaster management ?

21. Can statistics benefit individual players ?
22. How does math /statistics play a role in disease prediction ?
23. What role do statistical models play in weather forecasting ?

PRACTICAL ASSIGNMENT

Activity **Sports Performance Analysis**

In this activity, you will analyze the performance of athletes in a particular sport using statistical tools. By collecting data on various performance metrics, you will gain insights into the strengths and weaknesses of the athletes and identify areas for improvement.

Objective : To apply statistical analysis techniques to evaluate sports performance and make data-driven decisions.

Materials Needed : Stopwatch or timer, Pen and paper, Data collection sheet,

Instructions :

1. **Select a Sport:** Choose a sport that you are interested in analyzing, such as cricket, basketball, or Athletics.
2. **Define Performance Metrics:** Identify the key performance metrics that you will track during the activity. These may include: ***Total runs scored*** (for Cricket) or *Total points scored* (for basketball) or ***Race times*** (for Athletics) etc.
3. **Collect Data :** Choose a specific game, match, or event to observe.
   - Use a stopwatch or timer to record the duration of the game or event.
   - During the game, carefully observe the performance of the players and record relevant data on your data collection sheet. For example, record the number of runs scored by each player, or the time taken to complete a race etc.
   - Ensure that your data collection is accurate and consistent throughout the observation period.
4. **Interpret and analyse data:**
   (*a*) Once you have collected sufficient data, organize it into a structured format, such as a table or spreadsheet.
   (*b*) Calculate relevant statistics for each performance metric, such as :
   - Average runs/points scored per match
   - Average race time

   (*c*) Analyze the data to identify trends, patterns, and outliers (a distinct or abnormal data reading). Look for correlations between different performance metrics and draw insights from the data.

SESSION 3

Probability in Real Life

- Understanding Probability
- Calculating Probability of an Event
- Events in Probability
- Applications of Probability
- How AI Uses Probability

3.1 INTRODUCTION

What would you reply, if I ask you, "*What the chances are of it raining tomorrow?*" You might say, "***No chance***", or "***50:50***", or "***100% chance***". Do you what are we discussing? We are discussing the **probability** of raining tomorrow. **Probability** tells us how likely something is to happen. It helps us predict if an event will occur or not.

In this session we shall briefly talk about the concept of probability, some examples and how AI benefits from probability.

3.2 UNDERSTANDING PROBABILITY

Probability tells us how likely something is to happen. Probability helps us understand and predict the likelihood of events happening in real life. For example, probability of getting a six by rolling a dice once, is one out of six probable outcomes (see figure 3.1). By calculating probabilities, we can make better decisions and choices in various situations.

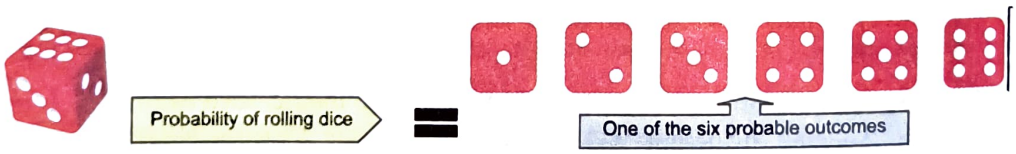


Figure 3.1 Concept of Probability

Probability is shown or represented as *a number between 0 and 1*.

- If the probability is **0**, it means the event cannot happen.
- If the probability is **1**, it means the **event will definitely happen**.
- If the probability is **between 0 and 1**, it means the **event might or might not happen**.

Probability

Probability is a way to measure the chance of something happening. It helps us predict the likelihood of different outcomes.

3.3 CALCULATING PROBABILITY OF AN EVENT

To calculate probability, we need to know the number of ways an event can happen and the total number of possible outcomes. And then we calculate ***probability*** as : *the number of ways an event can happen divided by the total number of possible outcomes.*

That is,

$$\text{Probability} = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}}$$

For example, if you have a bag with 3 red balls and 2 blue balls, the probability of picking a red ball is :

$$\text{Probability} = \frac{3}{3+2} = \frac{3}{5} = 0.6$$

Thus, the probability of picking a red ball (from the bag with 3 red plus 2 blue balls), is **0.6**

That is, the probability of picking a red ball is 3 out of 5, or 0.6, which can also be written as 60%.

Consider some more examples.

Example 1 *You have a box of chocolates with 10 chocolates. Out of these, 4 are milk chocolates, 3 are dark chocolates, and 3 are white chocolates. What will be the probability of picking a milk chocolate ?*

Solution. Probability of picking a milk chocolate

$$= \frac{\text{Number of milk chocolates}}{\text{Total number of chocolates}} = \frac{4}{(4+3+3)} = \frac{4}{10} = 0.4$$

The probability of picking a milk chocolate is 0.4, or 40%.

Example 2 *What is the probability of drawing a red diamond card from a deck of cards ?*

Solution. Probability of drawing a red diamond card

$$= \frac{\text{Number of red diamond cards}}{\text{Total number of cards}} = \frac{13}{52} = 0.25$$

The probability of drawing a red diamond card is 0.25, or 25%.

Note

Please note that **the sum of probabilities of all possible outcomes of an event would always be 1**, *e.g.*, after tossing a coin, probability of getting a head is 0.5 and probability of getting a tail is also 0.5. Sum of these two is 1 (0.5 + 0.5).

3.4 EVENTS IN PROBABILITY

In the context of probability, we refer to an **event** as a possible result or group of results that can occur from conducting an experiment or trial. An **experiment** is a process or activity that can be repeated multiple times, and each repetition leads to one of several specific outcomes (*i.e., events*). For example, we can roll a dice and it can generate an outcome, such as a one or a 3 or 6 etc. We can roll the dice repeatedly to get different outcomes.

So here,

- The **event** would be *getting a six or one* etc.
- The **experiment** would be *rolling the dice*.

The complete set of all potential outcomes that could happen in an experiment is called the *sample space*. So, an *event* can be considered a subset, or part, of the entire *sample space*. As per our earlier example of rolling a dice experiment,

- All possible outcomes would be : 1, 2, 3, 4, 5, 6
- Thus the **sample space** for rolling the dice experiment would be : 1, 2, 3, 4, 5, 6
- Any event (*e.g.*, getting 1, or 2 or a 6 etc.) will always be a subset of the sample space.

PART B

Let's talk about one more real-life example of use of probability in sports. In sports, probability can predict the outcome of a game. For example, if a basketball player makes 75 out of 100 free throws, the probability of making a free throw is :

$$\text{Probability} = \frac{75}{100} = 0.75 \text{ or } 75\%$$

Probability

In probability, an **event** represents one or more possible results from an experiment. An **experiment** is something that can be repeated, with defined potential outcomes each time. The **sample space** includes all possible outcomes that could occur in that experiment.

3.4.1 Types of Events

There are different types of events in probability. The most common events are :

1. **Certain Events**: These are events that will definitely happen. For example, *the sun rising in the east* is a **certain event**. Certain events have the probability of 1.
2. **Impossible Events**: These are events that will never happen. For example, rolling a 7 on a standard six-sided dice is impossible. Impossible events have the probability of 0.
3. **Likely Events**: These are events that have a high chance of happening. For example, if you have 90 red marbles and 10 blue marbles in a bag, picking a red marble is a likely event. Likely events are high probability events.
4. **Unlikely Events**: These are events that have a low chance of happening. For example, picking a blue marble from the same bag with 90 red marbles and 10 blue marbles is an unlikely event. Unlikely events are low probability events.

5. **Equally Likely Events:** These are events that have the same chance of happening. For example, when you flip a fair coin, getting heads or tails are equally likely events.

> **Note**
>
> An event is a subset or a grouping of some possible outcomes from the overall sample space.

Example 3 *Identify the type of event. Let's do a small exercise to see if you can identify the type of event. We are list some events below. Answer if the given event is this* ***certain, impossible, likely, unlikely****, or* ***equally likely*** *?*

(i) *Flipping a coin and getting heads.*

(ii) *Rolling a dice and getting a number greater than 6.*

(iii) *Picking a red ball from a bag with 1 red ball and 9 blue balls.*

(iv) *Choosing a day of the week and it being Monday.*

(v) *The sun rising tomorrow.*

Solution.

(i) Flipping a coin and getting heads. **Equally likely** event (*because getting heads or tails are equally likely events when flipping a fair coin*).

(ii) Rolling a dice and getting a number greater than 6. **Impossible** event (*because a standard dice only has numbers 1 to 6*).

(iii) Picking a red ball from a bag with 1 red ball and 9 blue balls. **Unlikely** event (*because there is only 1 red ball and 9 blue balls*).

(iv) Choosing a day of the week and it being Monday. **Equally likely** event (*because any day of the week has the same chance of being chosen*).

(v) The sun rising tomorrow. **Certain** event (*because the sun rises every day*).

3.5 APPLICATIONS OF PROBABILITY

Probability helps us predict how likely an event will occur. We calculate probabilities and use them in many everyday situations. Let us have a look at various applications of probability in real life:

3.5.1 Probability in Sports

Sports are full of uncertain events, and probability helps teams and players make better decisions. Probability is applied in Sports for various things. Some of these are :

1. Predicting Game Outcomes

Sports analysts use probability to predict the outcome of games. They look at past performances, player statistics, and other factors to calculate the probability of a team winning or losing. For example, if a basketball team has won 80 out of their last 100 games, the probability of them winning the next game is :

$$\text{Probability} = \frac{80}{100} = 0.8 \text{ or } 80\%$$

2. Player Performance

Coaches use probability to decide which players to put in the game. They look at the players' past performances to see how likely they are to score points, make passes, or block shots. For example, if a soccer player has scored in 10 out of 20 games, the probability of them scoring in the next game is :

$$\text{Probability} = \frac{10}{20} = 0.5 \text{ or } 50\%$$

3. Game Strategies

Probability helps teams choose the best strategies. For example, in the game of Cricket, a team might use probability to decide if they should put fielders at slips. If the probability of getting a catch at a slip point is high, they might go for it. If it is low, they might not put fielders there.

3.5.2 Probability in Weather Forecasts

Weather forecasts use probability to predict the weather. Meteorologists, the scientists who study the weather, use data from weather stations, satellites, and radars to make their predictions. Probability is applied in weather forecast for various things.

Some of these are :

1. Chance of Rain

When you hear that there is a 70% chance of rain, it means that in 70 out of 100 similar situations, it rained. This helps people decide if they should carry an umbrella or plan indoor activities.

2. Temperature Predictions

Meteorologists use probability to predict temperatures. They look at past data and current weather patterns to estimate the temperature. For example, *if the average temperature on a specific day in the past years was around 25°C, they might predict a similar temperature for this year.*

3. Severe Weather Alerts

Probability helps in predicting severe weather like *storms, tornadoes, and hurricanes.* By analysing weather patterns, meteorologists can estimate the probability of these events and issue warnings to keep people safe. For example, *if there is a high probability of a hurricane hitting a coastal area, people can evacuate in time.*

3.5.3 Probability in Traffic Estimation

Traffic estimation uses probability to help manage traffic flow and reduce congestion. Traffic engineers and planners use data from road sensors, cameras, and historical traffic data to estimate the traffic strength, peak times, and so forth. Using this data, the probability techniques are applied to handle uncertainty, and find patterns to make predictions for future traffic estimates.

1. Predicting Traffic Jams

By analysing traffic data, engineers can predict when and where traffic jams are likely to happen. For example, *if a road is usually congested during rush hour, they can estimate the probability of a traffic jam at that time.* This helps drivers plan their trips better.

2. Travel Time Estimation

GPS systems use probability to estimate travel times. They look at current traffic conditions and historical data to predict how long it will take to get from one place to another. For example, *if it usually takes 30 minutes to travel a certain route but there is a 70% chance of heavy traffic, the GPS might estimate a longer travel time.*

3. Accident Predictions

Traffic planners use probability to predict where accidents are likely to happen. They look at data on road conditions, weather, and driver behaviour. If they find a high probability of accidents in a certain area, they can make changes to improve safety, like adding traffic lights or speed bumps.

INFOBOX 3.1

Probability Applications

Probability is vital in many other fields like :

- Business. Estimating sales, product demand, investment risks etc.
- Manufacturing. Probability of defective products, quality control.
- Healthcare. Likelihood of disease, treatment effectiveness etc.
- Insurance. Probability of events like accidents, natural disasters etc.

3.6 HOW AI USES PROBABILITY ?

Artificial Intelligence (AI) relies heavily on probability to make decisions, learn from data, and predict outcomes, especially in uncertain situations.

Let's explore how probability helps AI.

1. Making Decisions

Artificial Intelligence systems often have to choose the best action from many possible options. To make these decisions, AI uses probability to evaluate the likelihood of different outcomes. For instance, a self-driving car uses probability to decide whether to stop, slow down, or speed up, based on the probability of obstacles on the road.

2. Learning from Data

Artificial Intelligence learns from data using techniques called *machine learning*. Probability helps AI understand patterns in the data and make predictions. For example, if an AI system is trained to recognize images of cats and dogs, it calculates the probability that a new image is a cat or a dog based on the patterns it learned.

3. *Predicting Outcomes*

Artificial Intelligence uses probability to predict future events. For example, in weather forecasting, AI systems analyse historical weather data to predict the probability of rain, snow, or sunshine on a given day. Similarly, in finance, AI predicts stock market trends based on the probability of different economic factors.

4. *Handling Uncertainty*

Probability helps AI handle uncertainty in the real world. Real-life data is often messy and incomplete, and AI must make decisions based on this imperfect information. Probability allows AI to make the best possible decisions even when there is uncertainty.

5. *Improving Accuracy*

By using probability, AI systems can improve their accuracy. For example, in medical diagnosis, AI analyses patient data to calculate the probability of different diseases. This helps doctors make more accurate diagnoses and provide better treatment.

6. *Learning and Adapting*

Probability helps AI systems learn from new data and adapt to changing environments. For instance, a recommendation system continuously updates its probabilities based on new user interactions, improving its recommendations over time.

PART B

Check Point

Multiple Choice Questions

1. In probability, an _____ represents one or more possible results of some action taking place.
   (*a*) Experiment (*b*) Event
   (*c*) Sample space (*d*) None of these
2. In probability, an _____ is something that can be repeated, with defined potential outcomes each time.
   (*a*) Experiment (*b*) Event
   (*c*) Sample space (*d*) None of these
3. In probability, the _____ includes all possible outcomes that could occur in that experiment.
   (*a*) Experiment (*b*) Event (*c*) Sample space (*d*) None of these
4. What is the probability of getting a head when tossing a fair coin?
   (*a*) 0.25 (*b*) 0.5 (*c*) 0.75 (*d*) 1
5. If the probability of an event is 0, it means the event is:
   (*a*) Impossible (*b*) Certain (*c*) Likely (*d*) None of these
6. If the probability of an event is 1, it means the event is:
   (*a*) Impossible (*b*) Certain
   (*c*) Likely (*d*) None of these

7. Which of the following is an impossible event ?
   (*a*) Rolling a 7 on a standard die
   (*b*) Picking a red ball from a bag with only red balls
   (*c*) Drawing an Ace from a deck of cards
   (*d*) Flipping a coin and getting heads

8. What type of event has a 100% probability of occurring ?
   (*a*) Impossible event
   (*b*) Certain event
   (*c*) Unlikely event
   (*d*) Equally likely event

9. Which of the following is an example of an equally likely event ?
   (*a*) It raining tomorrow
   (*b*) Flipping a coin and getting heads or tails
   (*c*) Winning the lottery
   (*d*) Scoring a goal in soccer

10. What is the sum of probabilities of all possible outcomes of an event ?
    (*a*) 0 (*b*) 0.5 (*c*) 1 (*d*) 2

11. Which of the following best describes a likely event?
    (*a*) It has a high chance of happening.
    (*b*) It has a low chance of happening.
    (*c*) It is certain to happen.
    (*d*) It is impossible to happen.

12. In weather forecasting, a 60% chance of rain means:
    (*a*) It will definitely rain.
    (*b*) It is equally likely to rain or not rain.
    (*c*) It is more likely to rain than not to rain.
    (*d*) There is no chance of rain.

13. In AI, probability is used to:
    (*a*) Design circuits
    (*b*) Make predictions and decisions
    (*c*) Write code
    (*d*) Perform physical tasks

14. Which of the following uses probability to improve safety on roads?
    (*a*) Predicting accidents
    (*b*) Designing car seats
    (*c*) Painting road signs
    (*d*) Building bridges

15. Probability is used in sports to:
    (*a*) Calculate scores
    (*b*) Predict game outcomes
    (*c*) Design uniforms
    (*d*) Set ticket prices

16. What type of event has an equal chance of happening or not happening?
    (*a*) Certain event
    (*b*) Impossible event
    (*c*) Likely event
    (*d*) Equally likely event

Fill in the Blanks

1. _____ is the study of how likely an event is to happen.
2. The probability of flipping a fair coin and getting tails is _____.
3. An event that will never happen is called an _____ event.
4. If a die is rolled, the probability of getting a number greater than 4 is _____.
5. A certain event has a probability of _____.
6. If there are 10 red balls and 5 blue balls in a bag, the probability of picking a blue ball is _____.

7. A weather forecast predicting a 70% chance of rain means there is a _____ probability of rain.
8. Coaches use probability to decide which _____ to put in the game.
9. Traffic estimation helps in predicting _____ times.
10. In AI, probability is used to make _____.
11. If a basketball player has a 70% success rate, the probability of making the next shot is _____.

True/False Questions

1. The probability of an impossible event is 1.
2. If a dice is rolled, the probability of getting a 3 is 1/6.
3. In weather forecasts, a 100% chance of rain means it will definitely rain.
4. AI uses probability to predict outcomes based on data.
5. An event that is certain to happen has a probability of 0.
6. Probability is used in sports to calculate game scores
7. A traffic estimation system can predict travel times using probability.
8. An equally likely event has a probability of 1.
9. The sum of probabilities of all possible outcomes is always 1.
10. The probability of rolling an even number on a die is 1/3.
11. AI does not use probability to improve accuracy.
12. In traffic estimation, probability helps predict accident locations.
13. Probability helps in making better decisions in everyday life.
14. The probability of a certain event is 1.
15. Probability helps meteorologists predict rainfall.
16. An event that cannot happen is called a certain event.
17. Traffic engineers use probability to predict traffic jams.
18. If the probability of an event is 0, it is an impossible event.
19. In sports, probability can predict player performance.
20. The probability of a likely event is very low.

Competency Based Questions

1. A basketball player has been successful in scoring free throws in 80 out of 100 attempts. Based on this performance, a coach is predicting the player's performance in the next game.

   What is the probability of the player making the next free throw?

   (*a*) 0.8 (*b*) 0.5

   (*c*) 0.2 (*d*) 1.0

2. A weather forecasting model predicts a 60% chance of rain tomorrow based on current weather patterns and historical data.

   What does a 60% chance of rain mean?

   (*a*) It will definitely rain tomorrow.

   (*b*) There is no chance of rain tomorrow.

   (*c*) It is more likely to rain than not to rain tomorrow.

   (*d*) It is equally likely to rain or not rain tomorrow.

3. A traffic estimation system uses historical traffic data to predict that there is a 70% chance of a traffic jam on a specific road during rush hour.
What action should drivers take based on this prediction?
(*a*) Ignore the prediction and drive as usual.
(*b*) Plan an alternative route to avoid the road.
(*c*) Drive faster to avoid the traffic jam.
(*d*) Wait until the traffic jam clears.

4. An AI system predicts that there is an 85% probability that a patient will develop a certain condition based on their medical history and current health data.
What should doctors do with this information?
(*a*) Ignore the prediction.
(*b*) Monitor the patient closely and take preventive measures.
(*c*) Discharge the patient immediately.
(*d*) Assume the patient is already healthy.

5. A financial analyst uses probability to predict that there is a 75% chance that a stock's price will increase over the next month.
Based on this prediction, what should an investor consider doing?
(*a*) Selling all their stocks immediately.
(*b*) Ignoring the prediction and not making any changes.
(*c*) Considering buying more of the stock.
(*d*) Assuming the stock will decrease in value.

6. An outdoor event is planned, and the weather forecast shows a 30% probability of rain.
How should the event organizers use this information?
(*a*) Cancel the event.
(*b*) Proceed with the event without any preparations.
(*c*) Have contingency plans in case it rains.
(*d*) Assume it will definitely rain.

Assertion & Reasoning Questions

Directions

In the following questions, a statement of assertion (A) is followed by a statement of reason (R).
Mark the correct choice as :

(*a*) Both A and R are true and R is the correct explanation of A.
(*b*) Both A and R are true but R is not the correct explanation of A.
(*c*) A is true but R is false (or partly true).
(*d*) A is false (or partly true) but R is true.
(*e*) Both A and R are false or not fully true.

1. Assertion (A). The probability of getting a head when flipping a fair coin is 0.5.
Reason (R). A fair coin has two equally likely outcomes.
2. Assertion (A). An event with a probability of 1 is considered impossible.
Reason (R). An event with a probability of 1 is certain to occur.

3. Assertion (A). The sum of probabilities of all possible outcomes of an event is always 1.
   Reason (R). Probability measures the likelihood of all possible outcomes together.
4. Assertion (A). The probability of drawing a red card from a standard deck of 52 cards is 1/2.
   Reason (R). A standard deck of cards has an equal number of red and black cards.
5. Assertion (A). Weather forecasts use probability to predict the chance of rain.
   Reason (R). Meteorologists use historical weather data and current conditions to calculate probabilities.
6. Assertion (A). The probability of getting an odd number when rolling a standard dice is 1/2.
   Reason (R). A standard dice has three odd numbers (1, 3, 5).

LET US REVISE

- *Probability is a way to measure the chance of something happening. It helps us predict the likelihood of different outcomes.*
- *In probability, an event represents one or more possible results from an experiment.*
- *In probability, an experiment is something that can be repeated, with defined potential outcomes each time.*
- *In probability, the sample space includes all possible outcomes that could occur in that experiment.*
- *An event is a subset or a grouping of some possible outcomes from the overall sample space.*

PART B

Solution Time

1. *What is probability ?*

   Ans. Probability is the measure of how likely an event is to happen.

2. *What is an impossible event ?*

   Ans. An impossible event is an event that cannot happen and has a probability of 0.

3. *What is a certain event ?*

   Ans. A certain event is an event that will definitely happen and has a probability of 1.

4. *How is an event related to sample space ?*

   Ans. In simple terms, an event covers one or more of the potential ways an experiment could turn out, taken from all the possibilities in the sample space for that experiment or trial. Defining the events and sample space is crucial for calculating probabilities.

5. *If you roll a dice, what is the probability of getting a 3 ?*

   Ans. 3 is one probable outcome out of six possible outcomes. So,

   $$\text{Probability of getting a 3} = \frac{1}{6} \approx 0.167 \text{ or } 16.7\%$$

   This is because a standard dice has 6 sides, each with an equal chance of landing face up.

6. *If you draw a card from a standard deck of 52 cards, what is the probability of drawing an Ace?*

   Ans. There are 4 Aces in a deck of 52 cards. So,

   $$\text{Probability of drawing an Ace} = \frac{4}{52} = \frac{1}{13} \approx 0.077 \text{ or } 7.7\%$$

7. *How is probability used in sports ?*

Ans. Probability is used in sports to predict game outcomes, player performance, and decide game strategies.

8. *How do meteorologists use probability ?*

Ans. Meteorologists use probability to predict weather conditions like rain, temperature, and severe weather events.

9. *How is probability used in traffic estimation ?*

Ans. Probability is used in traffic estimation to predict traffic jams, travel times, and accident locations.

10. *What does a 70% chance of rain mean ?*

Ans. A 70% chance of rain means that there is a high probability that it will rain.

11. *What is an equally likely event ?*

Ans. An equally likely event is one where all possible outcomes have the same chance of occurring.

12. *How does probability help in making decisions?*

Ans. Probability helps in making decisions by providing a measure of the likelihood of different outcomes, allowing for more informed choices.

Practical Session

1. *A jar contains 5 red marbles, 3 blue marbles, and 2 green marbles. What is the probability of picking a blue marble?*

Solution. First, count the total number of marbles : $5 + 3 + 2 = 10$

The number of blue marbles is 3.

So, the probability is :

$$\text{Probability} = \frac{3}{10} = 0.3 \text{ or } 30\%$$

2. *There are 20 students in a class, and 12 of them are girls. What is the probability of picking a girl's name out of a hat?*

Solution. The total number of students is 20, and the number of girls is 12.

So, the probability is :

$$\text{Probability} = \frac{12}{20} = 0.6 \text{ or } 60\%$$

3. *In a school of 500 students, 150 students play football. What is the probability that a randomly selected student plays soccer ?*

Solution. The total number of students is 500, and the number of students who play football is 150.

So, the probability is :

$$\text{Probability} = \frac{150}{500} = 0.3 \text{ or } 30\%$$

4. *A spinner has 8 equal sections, numbered 1 to 8. What is the probability of the spinner landing on an even number ?*

Solution. There are 4 even numbers (2, 4, 6, 8) out of 8 sections. So, the probability is :

$$\text{Probability} = \frac{4}{8} = 0.5 \text{ or } 50\%$$

GLOSSARY

Probability A measure of the likelihood that an event will occur.

Event Something that happens or could happen.

Sample Space Total Possible Outcomes; All the outcomes that could possibly happen.

Certain Event An event that is sure to happen.

Impossible Event An event that cannot happen.

Likely Event An event that has a good chance of happening.

Unlikely Event An event that has a small chance of happening.

Equally Likely Events Events that have the same chance of happening

Assignment

PART B

1. If a dice is rolled, what is the probability of getting an even number?
   (*a*) 1/2 (*b*) 1/3 (*c*) 1/4 (*d*) 1/6
2. In a traffic estimation system, probability helps in:
   (*a*) Designing new cars (*b*) Predicting travel times
   (*c*) Playing music (*d*) Cooking food
3. A certain event has a probability of:
   (*a*) 0 (*b*) 0.5 (*c*) 1 (*d*) 0.75
4. What does a probability of 0 mean?
   (*a*) The event is certain to happen. (*b*) The event is impossible.
   (*c*) The event is likely. (*d*) The event is equally likely.
5. The probability of drawing a King from a deck of cards is:
   (*a*) 1/13 (*b*) 1/4 (*c*) 1/52 (*d*) 1/26
6. Probability helps meteorologists predict:
   (*a*) Earthquakes (*b*) Rainfall (*c*) Volcanoes (*d*) Tides
7. Which event has a 25% chance of occurring?
   (*a*) Rolling a 2 on a standard die
   (*b*) Drawing a heart from a deck of cards
   (*c*) Flipping a coin and getting heads
   (*d*) Picking a red marble from a bag with 4 marbles (1 red, 3 blue)
8. If the probability of it raining tomorrow is 0.2, it means:
   (*a*) It will definitely rain. (*b*) It is unlikely to rain.
   (*c*) It is certain not to rain. (*d*) It is equally likely to rain or not rain.
9. AI uses probability to:
   (*a*) Draw pictures (*b*) Make weather forecasts
   (*c*) Write music (*d*) Design buildings

10. An event with an equal chance of happening or not happening is called an _____ event. (equally likely)
11. The sum of probabilities of all possible outcomes is always _____. (1)
12. The probability of rolling a 5 on a standard die is _____. (1/6)
13. Meteorologists use probability to predict _____. (weather)
14. Traffic engineers use probability to predict traffic _____. (jams)
15. A probability of 0 means an event is _____. (impossible)
16. Probability helps in making better _____. (decisions)
17. An event that is certain to happen is called a _____ event. (certain)
18. The probability of drawing a King from a deck of cards is _____. (1/13)
19. The probability of a fair coin landing on heads is _____. (0.5)
20. The probability of flipping a coin and getting heads or tails is 0.5. (F)
21. The probability of drawing an Ace from a deck of cards is 1/4. (F)
22. A likely event has a low chance of happening. (F)
23. Probability is not used in weather forecasting. (F)
24. AI uses probability to make decisions based on data. (T)
25. A 50% chance of rain means it is equally likely to rain or not rain. (T)
26. What is probability ? How is it useful ?
27. Define these terms : (*i*) event), (*ii*) experiment, (*iii*) sample space
28. What are different types of events? What is their probability ?
29. What is the probability of drawing an Ace from a standard deck of cards ?
30. What is the sum of probabilities of all possible outcomes of an event ?
31. What is the probability of flipping a coin and getting heads ?
32. What is the probability of rolling a 4 on a standard die ?

PRACTICAL ASSIGNMENT

1. **Identify the Probabilities**

   A weather report states :

   "There is a 40% chance of rain tomorrow and a 60% chance of sun."

   (*i*) What is the probability of rain ?

   (*ii*) What is the probability of no rain ?

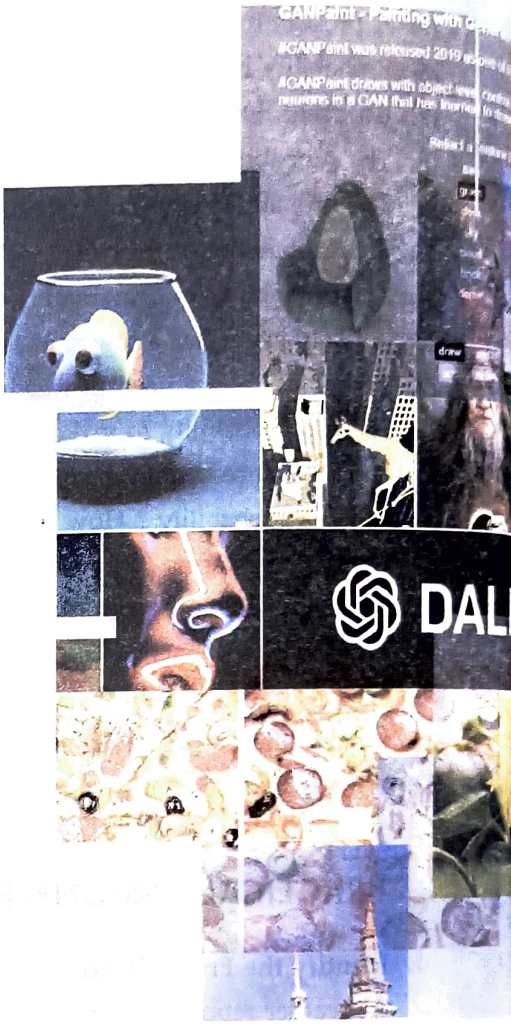
   (*iii*) Is rain impossible or certain to occur ?

2. Think of a real-life situation where probability is used, for example :
   - Probability of getting a parking spot
   - Probability of winning a game or contest
   - Probability of being late for school

   Discuss with your friends and identify the possible events and their probabilities.

Unit IV

Introduction to Generative AI



In This Unit

SESSION 1

Introducing Generative AI World

Do you know what is *Generative AI* ? To introduce it, just know that it is the creative form of AI. Let us now notice how creative can Generative AI be, through the following activity.

Activity 1 Can you spot the real *vs* AI-generated image ?

Carefully look at the following two images.

A  B

Identify which of these images is real and which one is created through generative AI ?

Solution. *Real Image is :* A *Generative AI image is :* B

Activity 2 Can you spot the real *vs* AI-generated image ?

Carefully look at the following two images.

C  D

Identify which of these images is real and which one is created through generative AI ?

Solution. *Real Image is :* C *Generative AI image is :* D

Activity 3 Can you spot the real *vs* AI-generated image ?

Carefully look at the following two images.

E

F

Identify which of these images is real and which one is created through generative AI ?

Solution. *Real Image is* : **E** *Generative AI image is* : **F**

Isn't it surprising how generative AI can create entirely new images that look real ?

After looking at these images, you must have figured out that *Generative AI* is a form of Artificial Intelligence that can create or generate some real looking images. Yes, you are right, but generative AI is way more than that. It can just not only create images, it can also create articles and text just like a human can, or it can even generate music or some real looking videos.

The following session, *Introduction to Generative AI* would talk about what it is, how it works, its advantages, challenges, and so forth.

SESSION 2

Introduction to Generative AI

- What is Generative AI ?
- Generative AI vs Conventional AI
- How Generative AI Works ?
- Generative AI Tools
- Ethical Considerations of Using Generative AI

2.1 INTRODUCTION

You all are aware that AI, *Artificial Intelligence*, has taken the world by storm. Everyone is aware of it and everyone is using and benefitting from it directly or indirectly. There are different types of AI that work in different ways for different purposes.

AI belongs to one of these *three* main types :

(*i*) **Narrow AI.** This AI can only perform specific, narrow tasks *like image recognition/ creation or language/text processing*. Virtual assistants like *Siri* or *Alexa* are forms of narrow AI focused on understanding voice commands.

(*ii*) **Reactive Machines.** These AI systems can only react to current situations and stimuli, without learning or building memories. Examples include *spam filters* and *recommendation systems* that suggest content based on your current behavior.

(*iii*) **Limited Memory AI.** This AI can store knowledge, learn from it, and then use that learning to perform new tasks and make predictions. It's used in applications like *chatbots*, *self-driving cars* etc. This is one of the most common types of AI today.

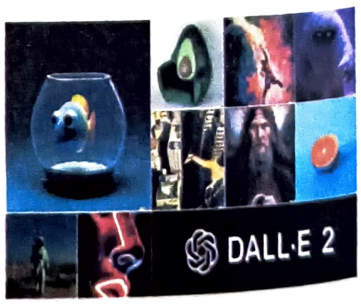
Generative AI is a type of narrow AI. In this session, we shall discuss what Generative AI is, its types, tools, benefits and limitations of generative AI, and ethical considerations of using generative AI.

Note

Generative AI is a type of narrow AI, that can only perform specific, narrow tasks.

2.2 WHAT IS GENERATIVE AI ?

Generative AI is a type of artificial intelligence that can create new content of different types, such as ***images***, ***music***, ***text***, or even ***videos***. Imagine you have a robot that can paint pictures, write stories, or compose songs. That's what Generative AI can do! It does this by learning patterns from a large amount of existing data using special algorithms and then create something new that is similar.



Another example use of generative AI could be this. For example, if you train a generative AI on millions of pictures of cats, it can then generate brand new images of cats that look real but have never existed before. It essentially learns what a cat looks like by studying the training data.

Note

Popular tools like ***ChatGPT*** and ***DALL-E*** are based on generative AI.

Generative AI

Generative AI is a type of artificial intelligence that can create new content of different types, such as ***images***, ***music***, ***text***, or even ***videos***.

2.3 GENERATIVE AI vs CONVENTIONAL AI

Regular AI or *conventional AI* is good at classifying and predicting based on existing data. But it can't create brand new data from scratch. However, generative AI is creative and can create new things. It can take what it has learned and use that to make something totally new and original, much like an artist paints a new picture.

Conventional AI follows rules to perform tasks. For example, a calculator uses math rules to solve problems. In contrast, Generative AI creates new things, *e.g.*, DALL-E creating an image. Generative AI doesn't just follow rules; it learns patterns from data and uses these patterns to generate new content.

Following table lists some differences in Generative AI and Conventional AI.

Table 2.1 *Generative AI vs. Conventional AI*

S.No.	***Conventional AI***	***Generative AI***
1.	It is good at classifying and predicting based on existing data.	It is good at creating new content of different types.
2.	It cannot create brand new data from scratch.	It can create brand new data from scratch.
3.	It follows rules to perform tasks.	It doesn't just follow rules; it learns patterns from data and uses these patterns to generate new content.
4.	It is limited to doing already planned jobs and tasks.	It is more creative and flexible and can create new things.
5.	It is suitable task-specific responses, *e.g.*, real-time response in automation, or vehicles etc.	It can create innovative and creative content, and it can create a variety of things.
6.	It requires clear rules and standards.	It can create content without preset clear rules.

Note

Regular AI is good at classifying and predicting based on existing data, but can't create new data or things. However, generative AI is creative and can create new things.

2.3.1 How Generative AI Works ?

Generative AI works by learning from data. Generative models work by first being trained on a large dataset, like millions of images or billions of words of text data. During training, the model learns the patterns and characteristics of the data. For images, it might learn features like shapes, textures, colours etc.

Then when you provide the trained model with a text prompt or seed data, it uses what it learned to generate brand new content that matches the patterns in a realistic way. It's like the model has absorbed and understood the ***rules*** *or* ***statistics*** behind what a real image or text should look like. So its generated output matches those learned patterns.

Thus, we can summarise that generative AI works by following these steps :

(*i*) **Training with Data.** The AI looks at lots of examples (*e.g., pictures of cats*).

(*ii*) **Learning Patterns.** It learns the patterns in these examples (*e.g., fur, eyes, and whiskers*).

(*iii*) **Generating.** It uses these patterns to create new images (*e.g,,* of cats that look realistic).

2.4 TYPES OF GENERATIVE AI

Generative AI comes in different types. Few main types of generative AI models are being listed below :

1. Generative Adversarial Networks (GANs)

GANs are like a game between two AI systems (two competing AI-based *neural networks*) : *the Generator* and *the Discriminator*. The *Generator* creates new content, and the *Discriminator* tries to guess if the content is real or fake. They learn from each other until the Generator becomes very good at creating realistic content.

2. Variational Autoencoders (VAEs)

VAEs learn to compress data into smaller bits and then reconstruct it. This helps them understand the data's underlying patterns and generate new, similar content.

3. Transformers

Transformers are used in language models like ChatGPT. They learn from lots of text data and can generate new text that looks like a human's work.

4. Diffusion Models

Diffusion models learn data distributions to generate things like photorealistic images from text descriptions.

2.5 EXAMPLES OF GENERATIVE AI

Generative AI is used these days in many fields. Some real-life examples of how Generative AI is used are :

(*i*) **Image Generation.** AI can create realistic images of people who don't exist. This is useful in movies and video games.

(*ii*) **Music Composition.** AI can compose new music pieces by learning from existing songs.

(*iii*) **Text Generation.** AI can write stories, articles, and even chat with you like a human.

(*iv*) **Video Creation.** AI can generate new videos or enhance existing ones by adding details.

Following figure 2.1 shows some common use cases of generative AI.

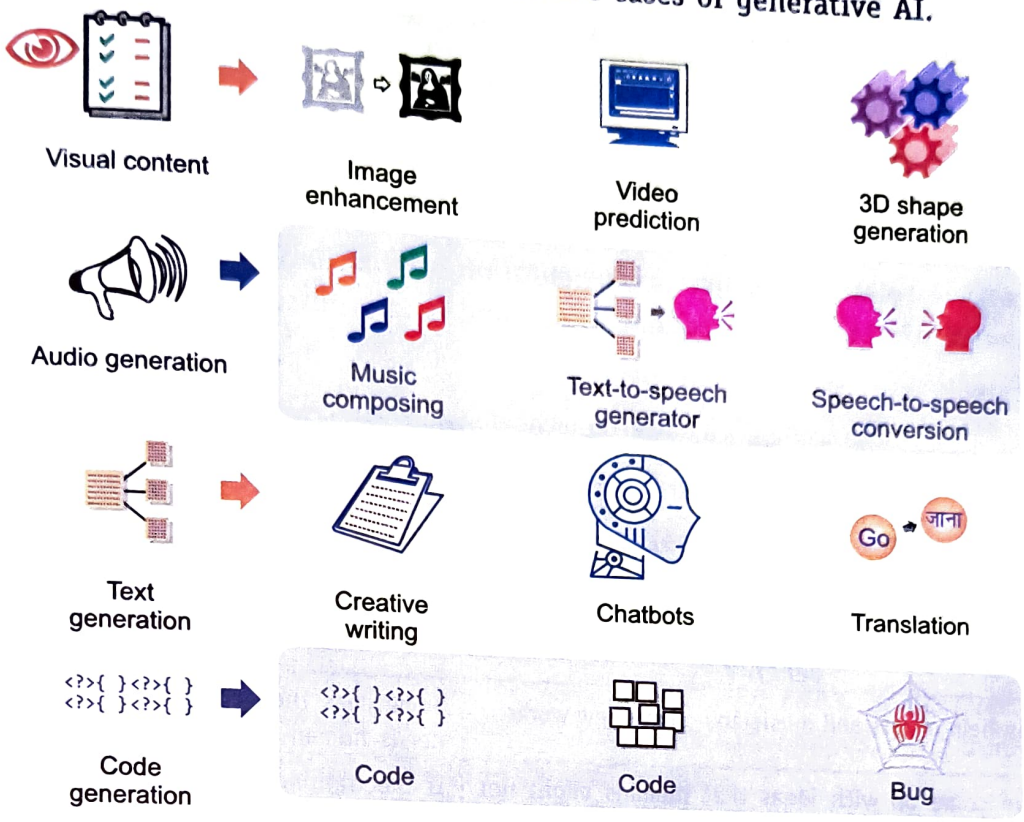


Figure 2.1 Generative AI use cases

2.6 BENEFITS AND LIMITATIONS OF USING GENERATIVE AI

Generative AI has made fast inroads in our daily work and activities. There are many benefits of using generative AI models, but there are certain limitations too.

For example, generative AI offers these **benefits** :

- **Creativity.** It can aid creativity and automate content creation for things like marketing materials, artwork, code generation and more.
- **Efficiency.** It can save time by automating routine tasks like generating product descriptions or design mockups.
- **Innovation.** Generative AI can explore new ideas and concepts that humans may not think of.
- **Data Generation.** It allows companies to quickly create synthetic data like images to train other AI systems.

Note

In science and research, generative models can boost discovery by generating new molecular structures or nano-patterns.

However, Generative AI also has some limitations :

- **Quality.** The output can be nonsensical, biased, or contain mistakes since the model just learns patterns.

- **Bias.** Models can unintentionally learn societal biases present in their training data, which will impact their outcomes (which will be then biased outcomes).
- **Ethics.** There are potential ethical issues around things like deepfakes or impersonating real people/data.
- **Quality.** AI-generated content may not always be high quality or factual. So users must check facts and ensure the correctness of data before using it, *e.g.*, if you got an article related to *policy implementation related to SDG in your area*, then you must verify the produced content's facts and correctness.

> Note
>
> So while it's a powerful technology, generative AI needs careful monitoring to ensure responsible use.

Following table lists the benefits and limitations of Generative AI.

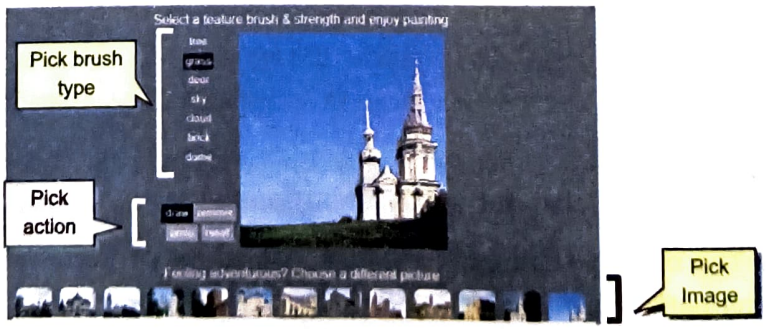
Table 2.2 *The Benefits and Limitations of Generative AI*

S.No.	Benefits	Limitations
1.	It can help artists and musicians create new works.	Sometimes the generated content isn't perfect and needs human improvement.
2.	It can come up with ideas that humans might not think of.	If the training data has biases, the AI will also be biased.
3.	It can generate content quickly, saving time and effort.	There are ethical concerns about using AI to create fake content.
4.	Generative AI can create synthetic data in bulk, which proves useful in situations where bulk data is required, such as training AI systems.	AI-generated content may not always be high quality or factual.

Activity 1 Use AI to Generate Real Looking Image using GAN Paint

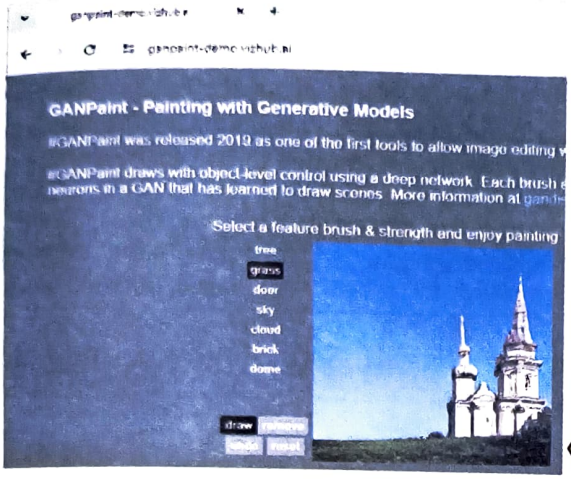
Try using *GAN Paint*, a tool that lets you modify images with AI. You can add *trees, buildings*, and more to a picture just by clicking! Use URL : ganpaint.io

1. Go to **ganpaint.io** in a web browser.
2. It will show a list of images in the lower side and some predefined paint tools/brushes on the left of image such as to paint *grass, trees, clouds* etc. (*see* below)



3. Explore with various tools, actions and images.

Following images show the *before* and *after* using cloud brush.



Before

After

2.7 GENERATIVE AI TOOLS

Generative AI tools leverage machine learning models like *GANs, VAEs, transformers,* and *diffusion models* to create new synthetic data outputs like *images*, *text*, *audio*, and more. These tools have a wide range of potential applications.

Various Generative AI tools are given below :

1. Image Generation Tools

These tools provide features (realistic digital brushes) or use prompts to generate photorealistic images and artworks. Some examples of Generative AI image generation tools are being listed in the table below.

Table 2.3 *Generative AI Image Generation Tools*

	Tool	Description	Application	Example
1.	GAN Paint	*GAN Paint* is a tool that allows users to modify images by adding or removing objects like trees, buildings, or clouds.	It is useful for artists/ designers to quickly prototype changes in images without manually editing.	An architect can use GAN Paint to visualize how adding trees might change the look of a building's surroundings.
2.	DeepArt	*DeepArt* uses AI to turn photos into artworks by applying styles of famous artists.	It helps artists / hobbyists create unique art pieces from ordinary photographs.	A user can transform a vacation photo into an artwork that looks like it was painted by a greater artist.

Some other Generative AI image generation tools are : *DALL-E , MidJourney, Stable Diffusion, Jasper Art, Prompt Hunt, NightCafe, AutoDraw, Designs.ai*, and many more.

2. Text Generation Tools

Generative text generation tools can write human-like text on almost any topic. Examples of Generative AI text generation tools are being listed in the table below.

Table 2.4 *Generative AI Text Generation Tools*

	Tool	Description	Application	Example
1.	ChatGPT, Claude.ai, GEMINI, CopySmith	These are examples of conversational AI that generates human-like text based on the input (prompt) it receives.	These are used for creating content, customer service chatbots, and even for educational purposes.	Businesses use them to provide 24/7 customer support through chatbots, answering common queries from customers.
2.	AI Dungeon	AI Dungeon is a text-based game that uses AI to generate interactive stories based on the player's choices.	It provides an interactive and engaging storytelling experience for entertainment and education.	Players can create and explore their own unique adventures, with the AI generating new scenarios based on their decisions.

Some other Generative AI text generation tools are : *DuetAI, Bard, AnyWord,* and many more.

3. Music Generation Tools

Generative music generation tools can create real music of different genres. Examples of Generative AI music generation tools are being listed in the table below.

Table 2.5 *Generative AI Music Generation Tools*

	Tool	Description	Application	Example
1.	Amper Music	Amper Music is a platform that allows users to create custom music tracks using AI.	It is ideal for video creators, marketers, and musicians to produce royalty-free music quickly.	A YouTuber can generate background music for their videos without worrying about copyright issues.
2.	AIVA (Artificial Intelligence Virtual Artist)	AIVA composes original music by learning from existing compositions.	It is used in creating soundtracks for films, video games, & other multimedia projects.	A filmmaker can use AIVA to generate an original score for a movie scene.

Some other Generative AI music generation tools are : *Boomy, SOUNDRAW, Soundful, Loudly, Musicfy, FM,* and many more.

4. Video Generation Tools

Generative video generation tools can generate entirely new video clips from text. Examples of Generative AI video generation tools are being listed in the table below.

Table 2.6 *Generative AI Video Generation Tools*

	Tool	Description	Application	Example
1.	Deepfake Technology	Deepfake technology uses AI to create realistic videos where people appear to do or say things they never did.	It can be used for entertainment, educational content, and even in political satire (though it has significant ethical considerations).	In movies, deepfake technology can be used to bring historical figures to life or to de-age actors.
2.	Synthesia	Synthesia allows users to create AI-generated videos with customizable avatars.	It is useful for creating training videos, marketing content, and personalized messages.	A company can use Synthesia to create a series of training videos featuring a consistent virtual presenter.

Some other Generative AI video generation tools are : *Colossyan, Hour One, D-ID, Elai, HeyGen, Runway, Pictory, Deepbrain AI, InVideo,* and many more.

Benefits of Generative AI Tools

Generative AI tools have become very popular as they offer the following benefits :

(*i*) **Creativity Enhancement.** They help artists, writers, and musicians by providing new ways to create content.

(*ii*) **Efficiency.** They can produce high-quality content quickly, saving time and effort.

(*iii*) **Accessibility.** They make complex content creation accessible to people without specialized skills.

(*iv*) **Personalization.** They allow for the creation of personalized content tailored to individual preferences and needs.

2.8 ETHICAL CONSIDERATIONS OF USING GENERATIVE AI

As generative AI becomes more advanced and widespread, we need to think about the ethical implications. Although Generative AI tools offer numerous benefits, they also raise important ethical issues, such as :

- **Misinformation.** AI-generated content can be used to create fake news or misleading information. For example, during elections, many fake news or misleading content, created through Generative AI, were circulated which some people could not suspect and believed in.
- **Bias.** If the AI is trained on biased data, it can produce biased content. For instance, a generative AI ad for hiring people displayed high-paying positions to males more often than to women, because of the data it was trained with, which had biases against women.
- **Privacy.** Using personal data to generate content can raise privacy concerns. For example, you must have read about or seen some fake videos circulating of some famous people. People's personal data such as information, pictures etc. can be misused by generative AI to create fake content, which is a big threat to privacy.
- **Copyright Violations.** AI-generated content might unintentionally infringe on existing copyrights. For example, a generative AI tool created a script for a movie inspired by a novel, which is copyrighted. This is copyright infringement.

Thus, before using the generative AI tools, one must follow and ensure the following ethical considerations :

(*i*) The users must ensure that the AI training data is fair and unbiased.

(*ii*) The users must protect individual privacy and prevent misuse like spreading misinformation.

(*iii*) Users must consider challenges like automated generation of malicious code or explicit content and must have ways to handle it.

(*iv*) Users must honour intellectual property and licensing. They must ensure that the AI-generated content is not violating this.

To help guide ethical AI development, groups like the OECD (*Organisation for Economic Co-operation and Development*) have proposed principles around transparency, robustness, and accountability for these systems.

Technology companies deploying generative AI will need clear policies and practices to identify and mitigate ethical risks. Humans should always have oversight over these powerful AI systems.

Check Point

Multiple Choice Questions (MCQs)

1. What is Generative AI ?
   (*a*) A type of AI that follows rules to perform tasks
   (*b*) An AI used for data analysis
   (*c*) A type of AI that creates new content (*d*) A robot that can paint pictures
2. Generative AI can create:
   (*a*) Only text (*b*) Only images
   (*c*) Only music (*d*) Images, text, music, and videos
3. Which of these is an example of narrow AI ?
   (*a*) A chess engine (*b*) Siri or Alexa
   (*c*) A spam filter (*d*) Both (*a*) and (*b*)
4. Which of the following is a tool that turns photos into artworks ?
   (*a*) GAN Paint (*b*) DeepArt (*c*) ChatGPT (*d*) AI Dungeon
5. What are Generative Adversarial Networks (GANs) composed of ?
   (*a*) Generator and Discriminator (*b*) Encoder and Decoder
   (*c*) Painter and Viewer (*d*) Player and Opponent
6. Which Generative AI tool is used for generating human-like text ?
   (*a*) DeepArt (*b*) Amper Music (*c*) ChatGPT (*d*) Synthesia
7. What does VAE stand for in Generative AI ?
   (*a*) Variational Adversarial Encoder (*b*) Virtual Artificial Environment
   (*c*) Variational Autoencoder (*d*) Visual AI Encoder
8. Which tool allows users to create custom music tracks using AI ?
   (*a*) AIVA (*b*) DeepArt (*c*) Amper Music (*d*) GAN Paint
9. Deepfake technology is mainly used for :
   (*a*) Creating realistic videos (*b*) Composing music
   (*c*) Writing stories (*d*) Analyzing data
10. Which of the following is a key benefit of Generative AI ?
    (*a*) Reduces creativity (*b*) Produces low-quality content
    (*c*) Enhances efficiency (*d*) Decreases accessibility
11. A tool that creates AI-generated videos with customizable avatars is called :
    (*a*) ChatGPT (*b*) Synthesia (*c*) AI Dungeon (*d*) GAN Paint

PART B

12. Ethical concerns of Generative AI include :
(a) Increasing creativity (b) Speeding up processes
(c) Misinformation (d) Enhancing accessibility

13. Which type of AI learns from text data to generate new text ?
(a) GAN (b) VAE (c) Transformer (d) Deepfake

14. A common use of AIVA is :
(a) Writing articles (b) Generating realistic images
(c) Composing music (d) Creating video games

15. Which AI tool is used to modify images by adding or removing objects ?
(a) GAN Paint (b) DeepArt (c) ChatGPT (d) Amper Music

16. Generative AI tools can help artists by :
(a) Making content creation slower (b) Increasing the complexity of artwork
(c) Providing new ways to create content (d) Reducing the need for creativity

17. Which of the following is NOT a limitation of Generative AI ?
(a) Quality issues (b) Bias in data
(c) Creativity enhancement (d) Ethical concerns

18. An AI tool that allows users to play text-based interactive stories is :
(a) Synthesia (b) GAN Paint (c) AI Dungeon (d) DeepArt

19. A common ethical issue with deepfake technology is :
(a) Improving video quality (b) Creating personalized content
(c) Misleading viewers with fake videos (d) Composing music for movies

20. Which group proposed principles for ethical AI development ?
(a) OECD (b) OpenAI (c) DeepMind (d) Anthropic

21. What type of AI is used in RunwayML for creating images and videos ?
(a) Conventional AI (b) Generative AI
(c) Analytical AI (d) Reactive AI

22. Generative AI is different from conventional AI because it :
(a) Follows strict rules (b) Creates new content
(c) Only analyzes data (d) Does not learn from data

Fill in the Blanks

1. _____ is a type of artificial intelligence that creates new content.
2. Regular AI is good at classifying and predicting based on _____ data.
3. Generative AI learns by studying patterns from a large _____ of data.
4. Transformers are a type of model great at _____ text.
5. _____ are used to create photorealistic images and deepfakes.
6. Generative AI could potentially learn and amplify societal _____.
7. The _____ has proposed principles around ethical AI development.
8. Generative AI can explore ideas that humans may not _____ of.

9. Tools like DALL-E and Stable Diffusion use _____ to generate images.
10. There are concerns around the _____ property of AI-generated content.
11. These models learn the _____ behind what real data looks like.
12. _____ is a tool that allows users to modify images by adding or removing objects.
13. _____ uses AI to turn photos into artworks.
14. GANs consist of a Generator and a _____ .
15. VAE stands for _____ Autoencoder.
16. _____ is a tool that creates AI-generated videos with customizable avatars.
17. One limitation of Generative AI is potential _____ in data.
18. ChatGPT is an example of a _____ model.

True/False Questions

1. Generative AI can create entirely new data like images and text.
2. GANs have only one neural network.
3. Narrow AI can only perform specific, narrow tasks.
4. Limited memory AI cannot store knowledge or learn.
5. Generative AI creates new content by following fixed rules.
6. VAEs are used for image generation.
7. Generative AI can be used to enhance videos.
8. AIVA is used to generate realistic text.
9. GAN Paint can add or remove objects in images.
10. Ethical concerns of Generative AI include misinformation.
11. Generative AI always produces perfect content.
12. Generative AI only analyzes data without creating new content.
13. ChatGPT is a Transformer model used for text generation.
14. DeepArt is an AI tool for composing music.
15. One limitation of Generative AI is potential bias in data.
16. Generative AI is different from conventional AI because it follows strict rules.

PART B

Competency Based Questions

1. A company wants to create a series of training videos quickly without hiring actors. They want customizable avatars that can deliver training content effectively.

   (*i*) Which AI tool should they use ?

   (*a*) DeepArt (*b*) Synthesia (*c*) AI Dungeon (*d*) Amper Music

   (*ii*) Why is Synthesia a good choice for this task ?

   (*a*) It composes music

   (*b*) It creates AI-generated videos with customizable avatars

   (*c*) It generates human-like text

   (*d*) It turns photos into artworks

   (*iii*) If they have to use a tool other than Synthesia, which of the following tools can help them create videos ?

   (*a*) Claude AI (*b*) Boomy (*c*) Runway (*d*) Gemini

2. An artist wants to experiment with creating artwork in the style of famous painters using AI.
   (*i*) Which tool should the artist use ?
   (*a*) ChatGPT (*b*) DeepArt (*c*) GAN Paint (*d*) Synthesia
   (*ii*) What benefit does DeepArt offer to the artist ?
   (*a*) Generates text
   (*b*) Turns photos into artworks in the style of famous painters
   (*c*) Modifies images by adding objects (*d*) Composes music
   (*iii*) If the artist is exploring using AI to generate novel digital artworks. Which generative AI approach creates new images ?
   (*a*) Reinforcement learning
   (*b*) Classification model
   (*c*) Generative adversarial network (GAN)
   (*d*) Supervised learning

3. A writer wants to include some conversations for their novel using AI.
   (*i*) Which AI tool should the writer use ?
   (*a*) Claude AI (*b*) Synthesia (*c*) ChatGPT (*d*) Amper Music
   (*ii*) The tool(s) that you chose for the above question, what do they do?
   (*a*) generate images (*b*) create music tracks
   (*c*) generate human-like text conversations
   (*d*) create lots of text
   (*iii*) What type of AI system would it be it also includes language translation?
   (*a*) Narrow AI (*b*) Reactive machine
   (*c*) Limited memory AI (*d*) General AI

4. A company wants to automatically generate product descriptions for e-commerce website.
   (*i*) Which generative AI tool would be most suitable ?
   (*a*) DALL-E (*b*) ChatGPT (*c*) Boomy (*d*) Soundful
   (*ii*) What is a potential limitation of using generative AI like ChatGPT for generating product descriptions ?
   (*a*) The output text will always be factually accurate.
   (*b*) It cannot handle long-form content.
   (*c*) The descriptions may lack contextual understanding.
   (*d*) The AI will require too much computational power.
   (*iii*) What approach could the company take to improve the quality of AI-generated product descriptions ?
   (*a*) Use narrower training dataset
   (*b*) Combine the AI output with human post-editing
   (*c*) Train the model on competitor website data
   (*d*) Reduce the model's complexity

5. A biotech firm wants to accelerate drug discovery by generating new molecular structures.
   (*i*) Which generative AI model should they use ?
   (*a*) A diffusion model (*b*) A variational autoencoder (VAE)
   (*c*) GPT-3 (*d*) A convolutional neural network

(*ii*) What is a key challenge in using generative AI for drug discovery ?

(*a*) Models can propose molecules that violate chemical rules

(*b*) The compute costs are too high

(*c*) Models cannot learn 3D molecular structures

(*d*) Lack of large, high-quality training data

(*iii*) How can firms address some of the limitations in this application ?

(*a*) Use simpler machine learning models

(*b*) Combine generative AI with other methods like virtual screening

(*c*) Train the model only on failed drug candidates

(*d*) Reduce the complexity of the generative model

Assertion & Reasoning Questions

Directions

In the following questions, a statement of assertion (A) is followed by a statement of reason (R). Mark the correct choice as :

(*a*) Both A and R are true and R is the correct explanation of A.

(*b*) Both A and R are true but R is not the correct explanation of A.

(*c*) A is true but R is false (or partly true).

(*d*) A is false (or partly true) but R is true.

(*e*) Both A and R are false or not fully true.

PART B

1. Assertion (A). Generative AI can create new content such as images, text, music, and videos.
   Reason (R). Generative AI uses patterns learned from existing data to generate new content.
2. Assertion (A). GAN Paint can add objects to an image.
   Reason (R). GAN Paint is used to create human-like text.
3. Assertion (A). Deepfake technology is mainly used for composing music.
   Reason (R). Deepfake technology creates realistic videos where people appear to do or say things they never did.
4. Assertion (A). Generative AI enhances creativity.
   Reason (R). Generative AI tools provide new ways to create content quickly and efficiently.
5. Assertion (A). Generative AI can only be used for creating text.
   Reason (R). Generative AI tools include image, text, music, and video generation.
6. Assertion (A). Ethical concerns of Generative AI include bias in data.
   Reason (R). If AI is trained on biased data, it can produce biased content.
7. Assertion (A). GANs consist of just a single neural network
   Reason (R). GANs actually have two neural networks - a generator that creates output, and a discriminator that tries to classify it as real or fake.
8. Assertion (A). Using generative AI requires no human oversight.
   Reason (R). Generative AI models can produce biased, nonsensical or unintended outputs, so they require careful human monitoring and control.
9. Assertion (A). Generative AI's outcomes must be critically observed from an ethical lens.
   Reason (R). There are potential ethical issues with generative AI like privacy violations, bias, deepfakes, intellectual property, and malicious content generation.

LET US REVISE

- *AI belongs to one of these three main types: Narrow AI, Reactive Machines, and Limited Memory AI.*
- *Generative AI is a type of Narrow AI.*
- *Generative AI is a type of artificial intelligence that can create new content of different types, such as images, music, text, or even videos.*
- *Popular tools like ChatGPT and DALL-E are based on generative AI.*
- *Regular AI is good at classifying and predicting based on existing data, but can't create new data or things. However, generative AI is creative and can create new things.*
- *Generative AI works by following these steps : (i) Training with Data, (ii) Learning Patterns, (iii) Generating.*
- *Few main types of generative AI models are : Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Transformers, and Diffusion Models.*
- *Generative AI is used for many things, such as : Image Generation, Music Composition, Text Generation, Video Creation, etc.*
- *Although Generative AI tools offer numerous benefits, they also raise important ethical issues, such as: Misinformation, Bias, Privacy, Copyright Violations, etc.*
- *Before using the generative AI tools, one must follow and ensure ethical considerations.*

Solution Time

1. *What is Generative AI? What type AI is it ?*

   Ans. Generative AI is a type of artificial intelligence that creates new content such as images, text, music, and videos by learning patterns from existing data.

   Generative AI is a type of narrow AI that can only perform specific, narrow tasks like image recognition/ creation or language/text processing.

2. *What are the two main components of GANs ? What are their roles ?*

   Ans. The two main components of GANs are : the ***Generator*** and the ***Discriminator***.

   The *Generator* creates new content, and the *Discriminator* tries to guess if the content is real or fake. These two work in tandem by helping each other and improving the content being generated.

3. *How do text generator tools like Claude, AI, ChatGPT generate text ?*

   Ans. Text generator tools like *Claude, AI, ChatGPT* generate text by learning from large amounts of text data. From the bulk of data fed to them, they starting identifying patterns and their interconnections. Then using the learnt patterns, they can create human-like text.

4. *Name a tool that uses AI to turn photos into artworks.*

   Ans. DeepArt.

5. *What does VAE stand for ? What does it do ?*

   Ans. VAE stands for **Variational Autoencoder**.

   VAEs learn to compress data into smaller bits and then reconstruct it. This helps them understand the data's underlying patterns and generate new, similar content.

6. *What are some ethical concerns related to Generative AI ?*

   Ans. The main ethical concerns related to Generative AI are :

   - **Misinformation.** AI-generated content can be used to create fake news or misleading information.
   - **Bias.** If the AI is trained on biased data, it can produce biased content.

- **Privacy.** Using personal data to generate content can raise privacy concerns.
- **Copyright Violations.** AI-generated content might unintentionally infringe on existing copyrights.

7. *What are some benefits and limitations of Generative AI ?*

Ans. Generative AI offers the following benefits and has the following limitations.

S.No.	Benefits	Limitations
1.	It helps artists and musicians create new works.	Sometimes the generated content isn't perfect and needs human improvement.
2.	It comes up with ideas that humans might not think of.	If the training data has biases, the AI will also be biased.
3.	It generates content quickly, saving time and effort.	There are ethical concerns about using AI to create fake content.
4.	It can create synthetic data in bulk, which proves useful in situations where bulk data is required, such as training AI systems.	AI-generated content may not always be high quality or factual.

8. *How is generative AI different from Conventional AI ?*

Ans. The differences in Generative AI and conventional AI can be summarised as :

S.No.	Conventional AI	Generative AI
1.	Good at classifying and predicting based on existing data.	Good at creating new content of different types.
2.	Cannot create brand new data from scratch.	Can create brand new data from scratch.
3.	Follows rules to perform tasks.	Learns patterns from data and uses these patterns to generate new content.
4.	Limited to doing already planned jobs and tasks.	More creative and flexible and can create new things.
5.	Suitable for task-specific responses, *e.g.*, real-time response in automation.	Can create innovative and creative content and it can create a variety of things.
6.	Requires clear rules and standards.	Can create content without preset clear rules.

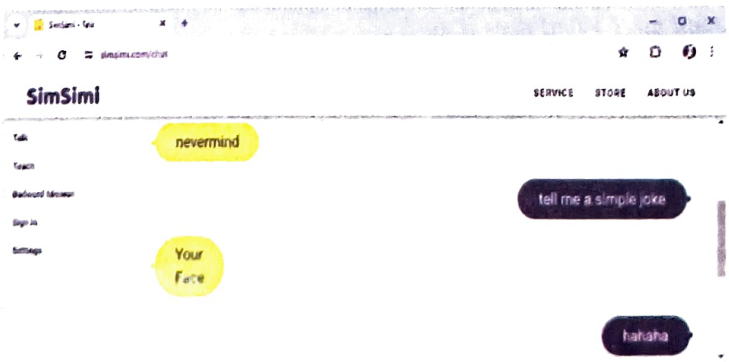
PART B

Practical Activity

1. *AI can generate human-like speech and vocals, imitating voices. Try interacting with an AI assistant that can have natural conversations. (you may use a fun chatbot like simsimi - url simsimi.com, or any other of your choice*

Solution.

1. Open URL *simsimi.com*
2. Choose preferred language and start chatting. Make sure to object if you do not like any response.



GLOSSARY

Generative AI a type of artificial intelligence that can create new content of different types.

Conventional AI a type of artificial intelligence that follows rules to perform tasks, such as classifying and predicting based on existing data.

Assignment

1. Generative AI can create new data like :
   (*a*) Images (*b*) Text (*c*) Audio (*d*) All of these
2. Which is NOT a type of generative AI model ?
   (*a*) Generative Adversarial Networks (GANs)
   (*b*) Variational Autoencoders (VAEs)
   (*c*) Convolutional Neural Networks (CNNs) (*d*) Transformers
3. How does generative AI learn to create new content ?
   (*a*) By studying a dataset of existing data and learning patterns
   (*b*) Through manual programming
   (*c*) It doesn't actually learn (*d*) None of these
4. The process of training Generative AI involves :
   (*a*) Only following fixed rules (*b*) Learning from examples
   (*c*) Ignoring data patterns (*d*) Generating content without prior data
5. What is the GAN in GAN Paint ?
   (*a*) Generative Adversarial Network (*b*) Graphics Animation Network
   (*c*) General Art Network (*d*) None of the above
6. Generative AI has potential benefits in which area ?
   (*a*) Automating content creation (*b*) Aiding scientific discovery
   (*c*) Creating synthetic training data (*d*) All of these
7. What is a key ethical risk of generative AI ?
   (*a*) It could automate too many jobs (*b*) It requires too much computing power
   (*c*) It could preserve societal biases (*d*) None of the above
8. Which statement about generative AI is (FALSE) ?
   (*a*) It learns patterns from data to create new outputs
   (*b*) Current systems have no limitations
   (*c*) It can generate images, text, audio etc. (*d*) It raises intellectual property concerns
9. Generative AI models need to be trained on :
   (*a*) A small dataset of examples (*b*) Handwritten rules and logic
   (*c*) A large corpus of data (*d*) Nothing, they create data spontaneously
10. What is a diffusion model good at generating ?
    (*a*) Deepfakes (*b*) Computer code
    (*c*) Photorealistic images (*d*) All of the above

11. Generative AI differs from conventional AI in that it can :
(a) Generate entirely new data
(b) Simply classify or predict on existing data
(c) Run on very limited hardware
(d) Both (b) and (c)

12. A language model like GPT-3 is an example of what type of generative AI ?
(a) GAN
(b) VAE
(c) Transformer
(d) Diffusion model

13. What does the abbreviation GAN stand for?
(a) Generative Adversarial Network
(b) Generative Art Network
(c) Geometric Animation Network
(d) None of the above

14. Generative AI cannot be used for which of these ?
(a) Generating new fashion designs
(b) Automating website coding
(c) Providing medical diagnoses
(d) Creating marketing videos

15. Generative AI helps in innovation by :
(a) Sticking to traditional methods
(b) Generating new ideas and content
(c) Slowing down the creative process
(d) Reducing the number of creative outputs

16. Which of the following is an advantage of using Generative AI?
(a) Always produces perfect content
(b) Can be biased
(c) Saves time and effort
(d) Ethical concerns

17. What does generative AI do?
(a) Classifies data
(b) Creates new data
(c) Analyzes existing data
(d) All of these

18. Generative AI learns patterns from:
(a) A small amount of data
(b) No data at all
(c) A large amount of existing data
(d) Human instructions

19. What does a GAN consist of ?
(a) One neural network
(b) Two competing neural networks
(c) Three neural networks
(d) No neural networks

20. What is a potential benefit of generative AI ?
(a) Increased cybersecurity risks
(b) Job displacement
(c) Aiding creativity and automating content creation
(d) More energy consumption

21. What is a limitation of generative AI ?
(a) Requires very little training data
(b) Always provides factual and high-quality output
(c) The output can be nonsensical or biased
(d) Eliminates human creativity

22. What is an ethical concern with generative AI ?
(a) Faster computing speeds
(b) Protecting individual privacy
(c) Lower hardware costs
(d) Reduced energy usage

23. Which type of AI just reacts to current stimuli ?
(a) Narrow AI
(b) Reactive machines
(c) Limited memory AI
(d) Generative AI

PART B

24. What type of generative AI is used for photorealistic image generation ?
(a) VAEs (b) GANs (c) Transformers (d) Diffusion models

25. Generative AI can be applied to which of these fields?
(a) Drug discover (b) Computer-aided design
(c) Creating synthetic training data (d) All of these

26. Generative AI essentially learns the _____ behind the training data.
(a) Rules & Patterns (b) Keywords
(c) Programming logic (d) Mathematical equations

27. Which is NOT a potential ethical issue with generative AI?
(a) Automated malicious code generation (b) Deepfakes and misinformation
(c) Increased energy efficiency (d) Privacy violations

28. Limited memory AI is commonly used in:
(a) Game playing (b) Self-driving cars (c) Computer vision (d) Both (a) and (b)

29. What type of generative AI can generate new text?
(a) VAEs (b) Transformers c) GANs (d) Diffusion models

30. An AI tool that learns from existing compositions to create new music is:
(a) AIVA (b) DeepArt (c) GAN Paint (d) AI Dungeon

31. Whose principles guide ethical AI development ?
(a) United Nations (b) OECD (c) World Bank (d) European Union

32. Generative AI requires _____ oversight.
(a) No (b) Limited (c) Extensive (d) Human

33. Generative AI learns _____ from a large amount of existing data. (patterns)

34. GANs consist of ____ (a number) competing neural networks. (two)

35. _____ are great at generating text. (Transformers)

36. _____ models learn data distributions to generate images. (Diffusion)

37. _____ is a tool that generates images from text prompts. (DALL-E or other similar tool)

38. _____ AI performs specific, narrow tasks. (Narrow)

39. _____ machines only react to current stimuli. (Reactive)

40. _____ AI can learn and make predictions. (Limited memory)

41. The _____ provides principles for ethical AI development. (OECD)

42. Generative AI requires _____ supervision. (human)

43. _____ AI is a type of AI that can create new data like images, text, and audio. (Generative)

44. Generative AI learns from data and identifies _____ . (patterns)

45. A Generative AI tool that helps filmmakers create original scores is _____ . (AIVA or Boomy or Soundful or any other music generator tool)

46. One ethical concern with Generative AI is the creation of _____ news. (fake)

47. Synthesia is useful for creating training _____ with virtual presenters. (videos)

48. A key limitation is that AI-generated output may not be fully _____ or factual. (accurate)

49. Generative AI classifies existing data. (F)

50. Generative AI learns from a small dataset. (F)

51. Generative AI essentially learns the rules of the data. (T)
52. GANs have two neural networks. (T)
53. Transformers generate images. (F)
54. A Generative AI limitation is biased or low-quality output. (T)
55. Narrow AI is very general intelligence. (F)
56. Reactive machines have memory. (F)
57. Limited memory AI makes predictions. (T)
58. Generative AI learns rules of the data. (T)
59. Chatbots use limited memory AI. (T)
60. VAEs learn surface patterns. (F)
61. Diffusion models generate from data distributions. (T)
62. Transformers can write human-like text. (T)
63. Generative AI raises no ethical issues. (F)
64. VAEs learn the underlying patterns in data. (T)
65. Transformers like GPT-3 can generate human-like text. (T)
66. What is generative AI ?
67. What are the three main types of Artificial intelligence? Which category does Generative AI belong to ?

PART B

68. Name three types of generative AI models.
69. What is a GAN ?
70. How do diffusion models work ?
71. What does a tool like DALL-E do ?
72. List a benefit and a limitation of generative AI.
73. What type of AI do self-driving cars use ?
74. Name some (*i*) image creation generative AI tools, (*ii*) text creation generative AI tools, (*iii*) music creation generative AI tools, (*iv*) video creation generative AI tools
75. How is generative AI different from conventional, regular AI ?
76. What are some ethical concerns with generative AI ?

PRACTICAL ASSIGNMENT

1. **Image Generation.** Tools like *DALL-E, Stable Diffusion, and Midjourney* use text prompts to generate photorealistic images and artworks.

   Use one of these tools to create poster for your school's annual day.
2. **Text Generation.** Large language models like GPT-3 can write human-like text on almost any topic.

   Use one of these tools to create text descriptions for various plays which are participating in the Annual School Theatre Festival.
3. **Video Synthesis.** Using generative AI tools that can generate entirely new video clips from text, create a short video of a "dog riding a skateboard".

ROADMAP AS IN UNIT V 2024 SYLLABUS

Unit V – Introduction to Python is same as the old Unit IV

Answers to Check Point

Unit I : AI Reflection, Project Cycle and Ethics

Session 9 : **Modelling, Evaluation, and Deployment** (Page S.12)

1. (*a*) 2. (*a*) 3. (*b*) 4. True 5. (*b*) 6. (*a*), (*d*)
7. (*b*), (*c*) 8. (*c*) 9. (*d*) 10. (*a*) 11. (*c*) 12. (*b*)
13. (*d*) 14. (*b*) 15. (*c*) 16. (*a*) 17. (*c*) 18. (*c*)
19. (*d*) 20. (*b*) 21. (*c*) 22. (*d*)

Unit II : Data Literacy

Session 1 : **Basics of Data Literacy** (Page S.25)

Multiple Choice Questions

1. (*b*) 2. (*d*) 3. (*c*) 4. (*d*) 5. (*b*) 6. (*d*)
7. (*d*) 8. (*b*) 9. (*c*) 10. (*b*) 11. (*d*) 12. (*a*)
13. (*d*) 14. (*a*) 15. (*c*) 16. (*d*) 17. (*d*) 18. (*a*)
19. (*c*) 20. (*b*) 21. (*d*) 22. (*a*) 23. (*a*)

Fill in the Blanks

1. understand, analyze, communicate
2. fact-check
3. Data Literacy Process
4. personal information
5. Data Security
6. informed
7. Explore & Visualize, Analyze & Communicate
8. data breaches
9. Security/Privacy
10. strong
11. software updates / security patches
12. unknown / untrusted
13. encryption
14. HTTPS, VPNs
15. cyber threats
16. awareness
17. data-driven

True/False Questions

1. F 2. F 3. T 4. F 5. F 6. F
7. T 8. F 9. F 10. F 11. F 12. F
13. F 14. T 15. F 16. T 17. F 18. T
19. F 20. T 21. F 22. T

Competency-Based Questions

1. (*c*) 2. (*d*) 3. (*c*) 4. (*c*)

Assertions and Reasons

1. (*c*) 2. (*d*) 3. (*c*) 4. (*d*) 5. (*a*) 6. (*b*)
7. (*c*)

Session 2 : **Acquiring Data, Processing, and Interpreting Data** (Page S.41)

Multiple Choice Questions

1. (*a*) 2. (*b*) 3. (*c*) 4. (*c*) 5. (*a*) 6. (*c*)
7. (*b*) 8. (*b*) 9. (*d*) 10. (*b*) 11. (*c*) 12. (*b*)
13. (*b*) 14. (*c*) 15. (*c*) 16. (*c*) 17. (*c*) 18. (*c*)
19. (*c*) 20. (*d*) 21. (*b*) 22. (*b*) 23. (*d*) 24. (*d*)
25. (*c*)

Fill in the Blanks

1. Qualitative 2. Quantitative 3. subjective 4. objective 5. preparing 6. preprocessing
7. acquisition 8. acquisition 9. data preprocessing 10. data cleaning
11. Data integration 12. Descriptive 13. Diagnostic
14. Predictive 15. Prescriptive 16. Diagnostic

True/False Questions

1. T 2. T 3. F 4. F 5. T 6. F
7. F 8. T 9. F 10. F 11. T 12. F
13. F 14. F 15. F 16. F 17. T 18. F
19. F 20. F 21. T 22. F 23. F

Competency-Based Questions

1. (*c*) 2. (*c*) 3. (*c*) 4. (*b*)

Assertions and Reasons

1. (*a*) 2. (*e*) 3. (*b*) 4. (*e*) 5. (*b*) 6. (*c*)
7. (*b*) 8. (*a*)

Unit III : Math for AI (Statistics & Probability)

Session 1 : **Importance of Math for AI** (Page S.68)

Multiple Choice Questions

1. (*b*) 2. (*c*) 3. (*c*) 4. (*b*) 5. (*b*) 6. (*a*)
7. (*b*) 8. (*a*) 9. (*a*) 10. (*a*) 11. (*c*) 12. (*c*)

Fill in the Blanks

1. data 2. predictions 3. Statistics 4. Linear algebra
5. Probability 6. multidimensional
7. gradients 8. trends 9. uncertain 10. statistics

True/False Questions

1. F 2. F 3. T 4. F 5. F 6. T 7. T

Competency-Based Questions

1. (*c*) 2. (*b*) 3. (*c*) 4. (*c*) 5. (*a*), (*c*)

Assertions and Reasons

1. (*a*) 2. (*a*) 3. (*b*) 4. (*a*) 5. (*c*)

Session 2 : **Statistics in Real Life** (Page S.81)

Multiple Choice Questions

1. (*a*) 2. (*c*) 3. (*a*) 4. (*c*) 5. (*a*) 6. (*a*) 7. (*c*)

Fill in the Blanks

1. data 2. game 3. weather 4. trends 5. analysis 6. disaster
7. risk 8. performance

True/False Questions

1. F 2. T 3. F 4. T 5. F 6. F 7. T

Competency-Based Questions

1. (*d*) 2. (*d*) 3. (*d*) 4. (*b*)

Assertions and Reasons

1. (*a*) 2. (*a*) 3. (*b*) 4. (*c*) 5. (*d*)

Session 3 : **Probability in Real Life** (Page S.94)

Multiple Choice Questions

1. (*d*) 2. (*a*) 3. (*c*) 4. (*b*) 5. (*a*) 6. (*b*)
7. (*a*) 8. (*b*) 9. (*b*) 10. (*c*) 11. (*a*) 12. (*c*)
13. (*b*) 14. (*a*) 15. (*b*) 16. (*d*)

Fill in the Blanks

1. Probability 2. 0.5 3. impossible 4. 1/3 5. 1 6. 1/3
7. high 8. players 9. travel 10. predictions 11. 0.7

True/False Questions

1. F 2. T 3. T 4. T 5. F 6. F
7. T 8. F 9. T 10. T 11. F 12. T
13. T 14. T 15. T 16. F 17. T 18. T
19. T 20. F

Competency-Based Questions

1. (*a*) 2. (*c*) 3. (*b*) 4. (*b*) 5. (*c*) 6. (*c*)

Assertions and Reasons

1. (*a*) 2. (*d*) 3. (*a*) 4. (*a*) 5. (*a*) 6. (*c*)

Unit IV : Introduction to Generative AI

Session 2 : **Introduction to Generative AI** (Page S.114)

Multiple Choice Questions

1. (*c*) 2. (*d*) 3. (*d*) 4. (*b*) 5. (*a*) 6. (*c*)
7. (*c*) 8. (*c*) 9. (*a*) 10. (*c*) 11. (*b*) 12. (*c*)
13. (*c*) 14. (*c*) 15. (*a*) 16. (*c*) 17. (*c*) 18. (*c*)
19. (*c*) 20. (*a*) 21. (*b*) 22. (*b*)

Fill in the Blanks

1. Generative AI 2. existing 3. dataset/corpus 4. generating
5. GANs 6. biases 7. OECD 8. think
9. text prompts 10. intellectual 11. patterns/rules 12. GAN Paint
13. Deep Art 14. Discriminator 15. Variational
16. Synthesia 17. bias 18. Transformer

True/False Questions

1. T 2. F 3. T 4. F 5. F 6. T
7. T 8. F 9. T 10. T 11. F 12. F
13. T 14. F 15. T 16. F

Competency-Based Questions

1. (*i*) – (*b*) ; (*ii*) – (*b*) ; (*iii*) – (*c*)
2. (*i*) – (*b*) ; (*ii*) – (*b*) ; (*iii*) – (*c*)
3. (*i*) – (*a*), (*c*) ; (*ii*) – (*c*) ; (*iii*) – (*a*)
4. (*i*) – (*b*) ; (*ii*) – (*c*) ; (*iii*) – (*b*)
5. (*i*) – (*b*) ; (*ii*) – (*d*) ; (*iii*) – (*b*)

Assertions and Reasons

1. (*a*) 2. (*c*) 3. (*d*) 4. (*a*) 5. (*d*) 6. (*a*)
7. (*c*) 8. (*c*) 9. (*a*)