

INDIVIDUAL TASK – MODULE 2

DATA ESSENTIALS ,TYPES , BIG DATA , PROCESSING , AND ETHICS

1.Explain “My Daily Inventory “,analyze a privacy policy of any popular app, explain one-real world example

1) My Daily Data Inventory

Daily we interact with many data sources like mobile apps, websites, sensors, social media, etc.

These data can be classified into:

Structured Data

Semi-Structured Data

Unstructured Data

A) Structured Data

Data stored in fixed format.

Easy to store in databases.

Organized in rows and columns (table format).

Can be easily searched and analyzed.

Examples:

Bank transaction details

Student marks list

Phone contacts list

Search history records

B) Semi-Structured Data

Data is not fully organized like tables but has tags or format.

It contains labels and markers to separate data elements.

Flexible structure compared to structured data.

Example formats:

JSON

XML

Logs

Examples:

Emails

Website cookies data

Social media profile details

Likes and comments

C) Unstructured Data

Data with no fixed format.

Difficult to store in table form.

Large amount of multimedia data.

Examples:

Photos & selfies

Videos on YouTube / Instagram

Voice recordings

Video content

2) Analyzing a Privacy Policy

A privacy policy tells how an app collects, stores, and uses user data.

Data Collected by WhatsApp:

Phone number

Contact list

Messages metadata

Device information

Location data

Ethical Considerations

User consent must be clear before collecting data.

Data sharing with third-party companies may affect privacy.

Tracking behavior may create ethical issues.

Users should be informed about how their data is used.

Data Governance Practices

Encryption used for message security.

Data stored on servers for backups and service improvement.

Policy explains how long data is stored.

Allows reporting and blocking of accounts for safety.

Ensures security measures to protect user information.

3) Big Data Around Me

YouTube is a real-world big data example because it processes millions of users daily.

Big Data is explained using 3Vs:

Volume

Huge amount of data is generated daily.

Billions of videos, likes, comments & watch history are stored.

Large-scale storage systems are required.

Velocity

Data is produced and updated very fast.

Recommendations change instantly based on what we watch.

Real-time processing is required.

Variety

YouTube collects different types of data like:

Video content (Unstructured)

Likes & comments (Semi-structured)

Search history (Structured)

User interests & viewing patterns

Device and location data

4) Data Preparation Thought Experiment

If data is messy and incomplete, we must prepare it before training AI models.

Steps to Prepare Data:

Step 1: Data Collection

Collect data from websites, sensors, apps, files.

Step 2: Remove Duplicate Data

Eliminate repeated records.

Step 3: Handle Missing Values

Fill missing data or remove incomplete records.

Step 4: Remove Noise & Errors

Correct wrong or inconsistent data.

Step 5: Data Formatting

Convert data into proper format.

Step 6: Data Transformation

Normalize or scale data.

Step 7: Feature Selection

Select important features for the model.

Step 8: Data Labeling

Assign correct labels to data.

Step 9: Split Data

Divide into training and testing data.

Step 10: Store Clean Data

Store processed data safely for model training.