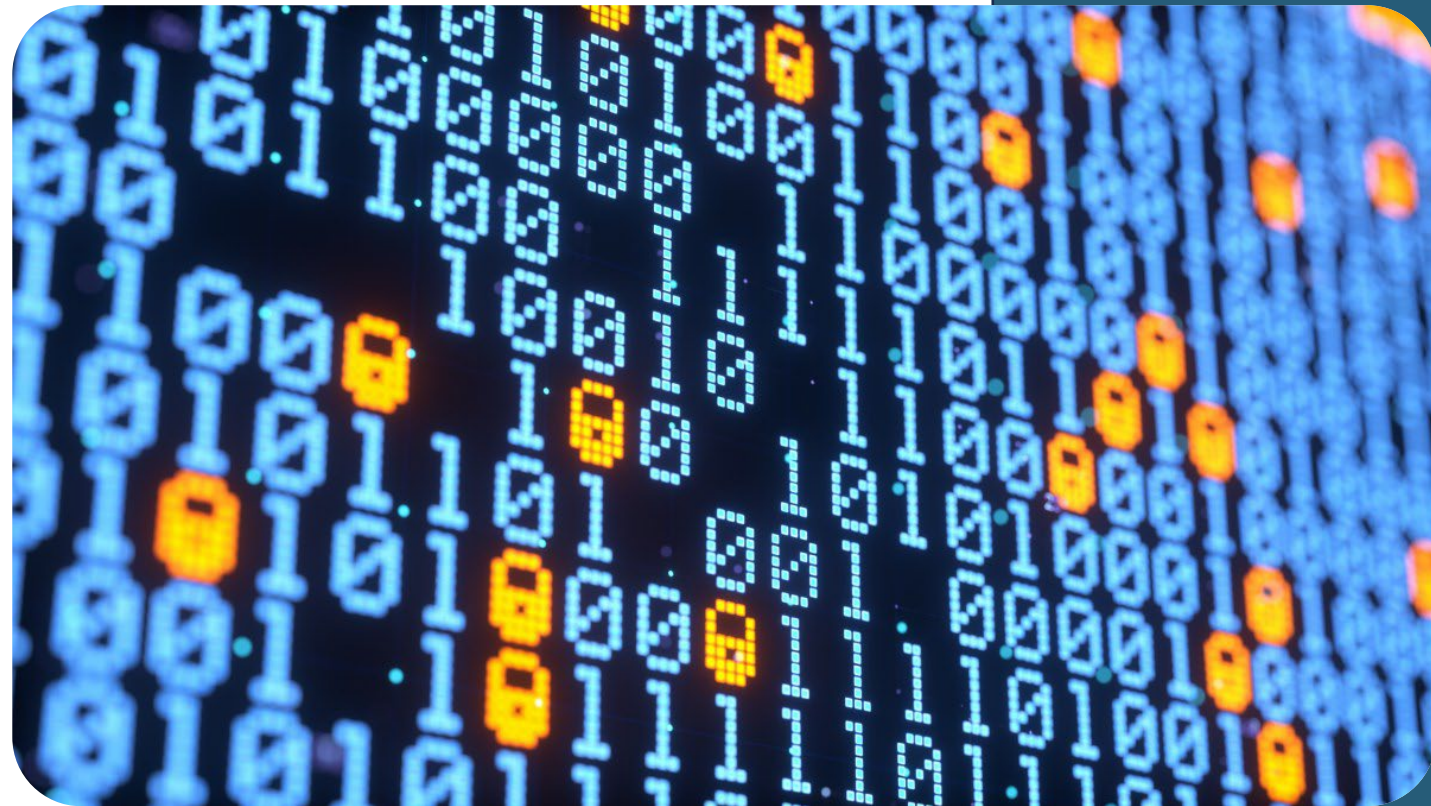# The Role of Data
in AI Predictive Models

Dr.Zeynab Barzegar

Introduction

❖Importance of Data in AI Models

❖Data Gathering

❖Key Steps in Data Preprocessing

❖Data in AI Models

Importance of Data in AI Models

❖Clean and structured data is critical for accurate predictions

❖Preprocessing enhances:

    ❖Model Accuracy: By reducing noise and irrelevant data

    ❖Generalization: Ensures models perform well on unseen data

    ❖Efficiency: Reduces computational costs and overfitting risks

Data Gathering

- ❖ **In-house Data Collection**
- ❖ **Automatic Data**
- ❖ **Manual Data Collection**
- ❖ **Off-the-shelf Datasets**
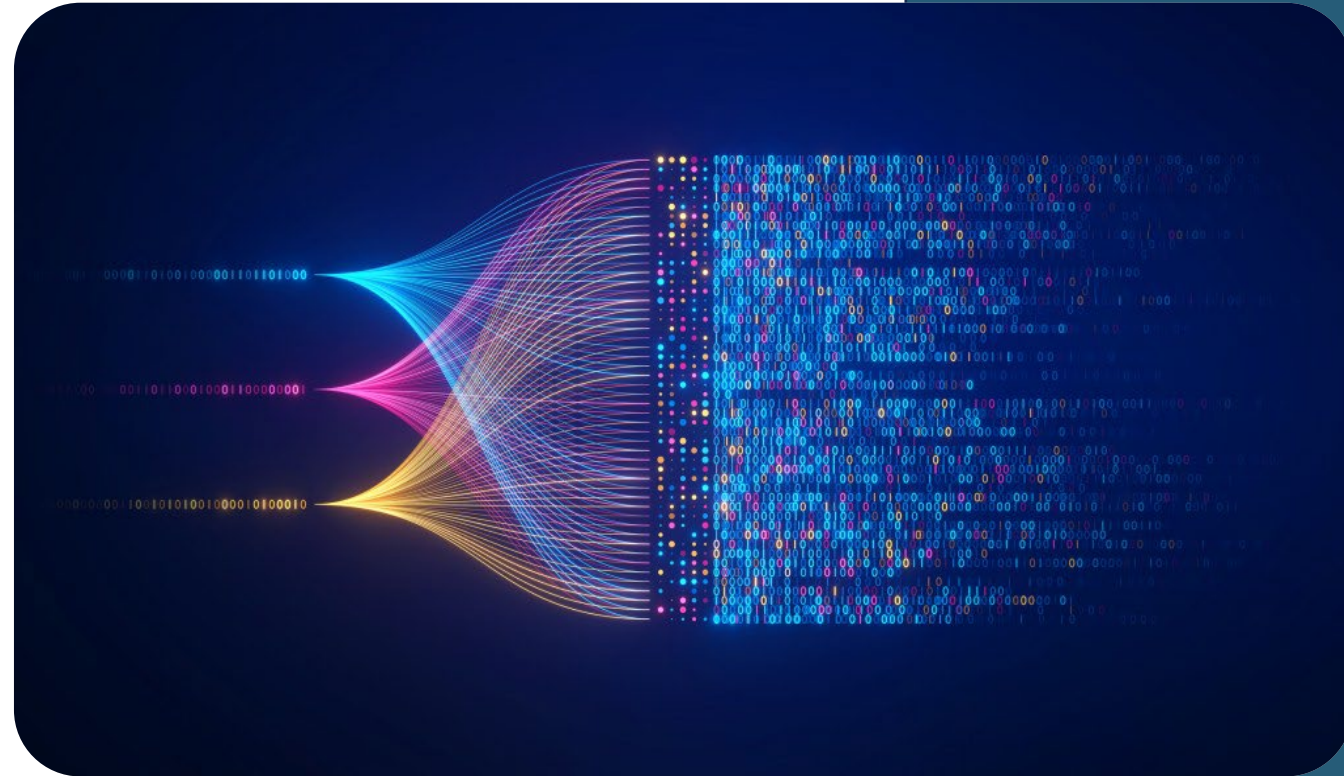
Data Preprocessing

Data Cleaning

❖ Definition: The process of identifying and correcting (or removing) inaccurate records from a dataset.

❖ Handling Missing Values:

➢ Deletion: Removing records with missing values

➢ Imputation: Filling in missing values using statistical methods (e.g., mean, median, mode)

➢ Predictive Modeling: Using algorithms to predict and fill missing values based on other data

❖ Outlier Detection and Removal:

➢ Identifying data points that deviate significantly from the rest of the dataset

➢ Statistical tests (e.g., Z-score, IQR)

➢ Visualization techniques (e.g., box plots)

Data Preprocessing

Data Transformation

❖**Definition: Modifying the format or structure of data to make it suitable for analysis**

❖**Normalization :**

➢**Scaling numerical features to a common range (e.g., or [-1, 1])**

➢**Common methods include Min-Max scaling and Z-score normalization**

❖**Encoding Categorical Variables**

➢**Converting categorical data into numerical format for model compatibility**

Feature Engineering

❖ **Definition: The process of using domain knowledge to create new input features from existing ones**

❖ **Creating New Features**

❖ **Polynomial Features**

❖ **Feature Selection**

Data in AI Models

❖ The quality, structure, and preprocessing of data directly impact the performance and accuracy of AI systems
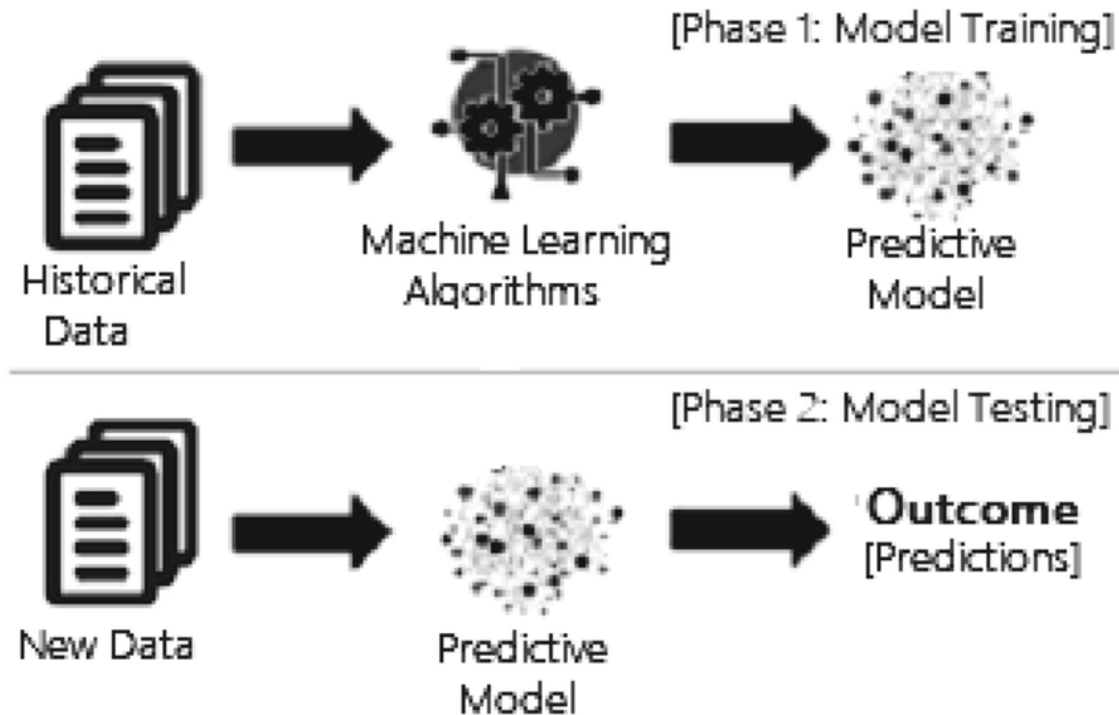
Role of Data in AI Models

❖ Foundation for Learning

❖ Foundation for Learning

❖ Decision-Making Support

❖ Continuous Improvement

Thank you!