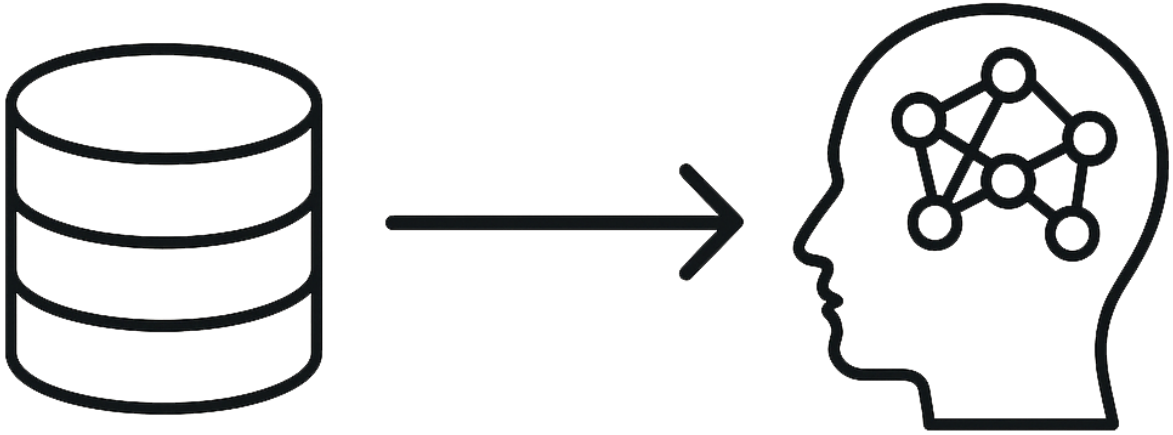
# Session 1: Data and AI

- Understanding why data is essential for AI models
- Exploring how AI learns from data
- Recognizing patterns
- Roadmap
- What is EDA?

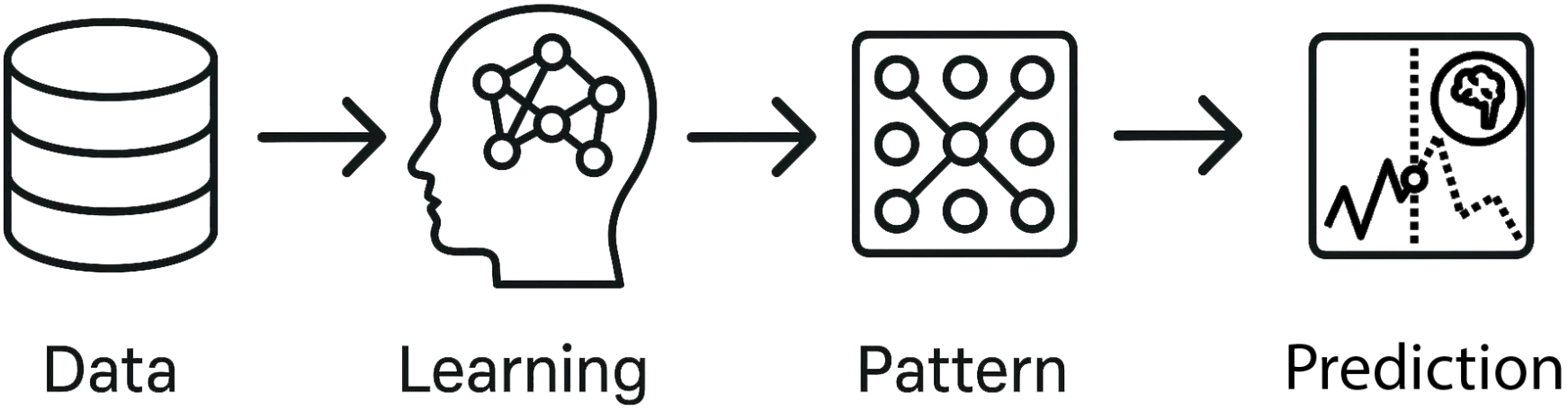
# Why Data is Required for AI Models

- Data is the foundation of AI models — without it, models cannot learn.
- AI learns patterns, relationships, and insights from examples.
- The more relevant and high-quality data we have, the better the AI performs.
- Garbage in, garbage out — poor data leads to poor AI predictions.



# How AI Models Work

- 1. Input Data → Feed data into the model.
  - numbers, text, images, or anything relevant.
- 2. Pattern Recognition → Model identifies trends and relationships.
  - It might notice that certain values in a blood test often appear in patients with a specific disease.
- 3. Learning → Adjusts internal parameters to improve accuracy.
  - think of these like ‘knobs’ it tunes to get better predictions.
- 4. Prediction → Produces results based on learned patterns.
  - AI can predict outcomes it hasn’t seen before, based on what it has learned from the data.



# Data and Pattern Recognition

- Data contains hidden patterns that AI can uncover.
  - A pattern is a relationship between inputs and outputs that repeats often enough to be useful.
- Example: In medical datasets, certain features may indicate higher disease risk.
- AI models find patterns humans might miss.
- Visualization tools help humans interpret these patterns.

```python
# Pattern-based decision
if bmi > 30 and activity_level == "low":
    recommendation = "High priority: Start regular exercise"
elif 25 <= bmi <= 30 and age > 40:
    recommendation = "Moderate priority: Increase activity"
elif bmi < 25 and activity_level == "high":
    recommendation = "Maintain current routine"
else:
    recommendation = "General health check recommended"
```

# Roadmap

- Introduction to Artificial Intelligence and Practical Tools
  - Overview of basic AI concepts, introduction to key Python libraries (NumPy, Pandas, Scikit-learn), and basic data preprocessing techniques.
- Supervised Learning
  - Learning fundamental algorithms such as Linear regression, Logistic Regression and Decision Trees with model performance evaluation.
- Unsupervised Learning
  - Exploring clustering algorithms like K-Means and their applications in analyzing unlabeled data.
- Deep Learning Basics
  - Introduction to Artificial Neural Networks, focusing on MLP architecture and training simple models with Keras.
- Natural Language Processing (NLP) Project
  - Conducting a sentiment analysis project using LSTM models and text data, including preprocessing, modeling, and performance evaluation.

# Data Science Workflow

1. Data Collection: From a database, CSV, or API
2. Exploratory Data Analysis (EDA): Visualizing, summarizing, and spotting patterns
3. Data Cleaning: Handling missing values, removing duplicates, etc.
4. Modeling: Classification, regression, clustering, etc.
5. Evaluation: Accuracy, precision, recall, etc.
6. Deployment: Use model in production or business setting

# What is EDA?

- EDA (Exploratory Data Analysis) is the process of exploring datasets to summarize their main characteristics.
- Purpose: Understand the data before building models.
- Analogy: Like checking ingredients before cooking — know what you have, how fresh it is, and if anything is missing.

# Why EDA is Important

- Detect errors or missing data early.
- Spot trends, patterns, and relationships.
- Choose the right features for modeling.
- Avoid bad data entering the model.

# Steps of EDA

- Understand the context — what does the dataset represent?
- Inspect the structure — rows, columns, data types.
- Check completeness — missing values, duplicates.
- Summarize statistics — mean, median, standard deviation.
- Spot patterns & outliers — box plots, scatter plots.
- Visualize relationships — correlations, group comparisons.

# Types of EDA

- Univariate Analysis — single variable (histograms, value counts).
- Bivariate Analysis — relationship between two variables (scatter plots, bar charts).
- Multivariate Analysis — multiple variables (heatmaps, pair plots).

# Common EDA Visualizations

- Histograms — distribution of values.
- Boxplots — outliers & spread.
- Scatter plots — relationships.
- Heatmaps — correlations.