

AI & ML – Core Concepts Document

This document provides a clear, structured explanation of essential AI and Machine Learning concepts, along with the complete ML workflow and modern AI practices used in real-world applications.

1. Understanding Data (Foundation of AI/ML)

Before building any model, the most important step is understanding the data.

Key Points:

- Understand what the dataset represents.
- Know the meaning of each column/feature.
- Identify the target/output variable.
- Understand the real-world problem being solved.

Example:

Problem: Predict customer churn.

- Features: Age, usage frequency, subscription type.
 - Target: Whether the customer will leave or stay.
-

2. Data Cleaning & Preprocessing

Raw data usually contains issues. Preprocessing makes it ready for modeling.

Tasks:

- Handle missing values.
- Remove duplicates.
- Fix or remove outliers.
- Encode categorical data.
- Normalize/scale numerical features.
- Split the dataset into train, validation, and test sets.

Why it matters:

Proper preprocessing often contributes to **70% of the final model accuracy**.

3. Understanding Model Types

Machine learning models fall into two broad categories:

A. Traditional Machine Learning

Used mainly for structured/tabular data.

Examples:

- Linear Regression
- Logistic Regression
- Random Forest
- Support Vector Machines (SVM)
- XGBoost

Use Cases:

- Predictions
- Classifications
- Recommendations

B. Deep Learning & Advanced AI (Neural Networks)

Used for unstructured data such as text, images, and audio.

Examples:

- Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Networks/LSTMs
- Transformers (BERT, GPT, LLaMA)

Use Cases:

- NLP
- Speech-to-text
- AI chatbots
- Image recognition

4. Model Training & Evaluation

This phase teaches the model to recognize patterns from data.

Steps:

- Train the model using training data.
- Test it using unseen data.
- Evaluate how well it performs.

Metrics:

- Accuracy
- Precision
- Recall
- F1 Score
- Confusion Matrix

Model training itself is often only **20%** of the total ML work. The remaining effort goes into data handling and evaluation.

5. Experimentation & Model Improvement

Machine learning requires constant iteration.

Typical Workflow:

- Try multiple algorithms.
- Tune hyperparameters.
- Compare evaluation metrics.
- Choose the best-performing model.

This process is called **experimentation**.

6. Modern AI Concepts (Practical & Applied)

Today, AI includes more than classical ML. Key modern concepts include:

A. RAG – Retrieval Augmented Generation

Allows AI models to respond using custom documents.

B. Fine-Tuning

Adapting large language models (LLMs) using custom data.

C. Embeddings

Converting text into vectors for:

- Search
- Document similarity
- Recommendations
- Chatbots

D. LLM Utilization

Using API-driven large language models like:

- OpenAI
- Claude
- Gemini
- Local models (LLaMA, Mistral)

E. MLOps Basics

Understanding:

- Deployment of ML models
 - Version control
 - Monitoring model performance
-