🧠 **Phase 1: Core AI & Machine Learning (2–4 months)**

📌 **Goal: Understand the fundamentals of machine learning, including supervised and unsupervised learning, and how to implement them using Python libraries.**

✅ **Step-by-Step Topics List (In Order)**

🔹 **1. Introduction to Machine Learning**

- What is Machine Learning?
- ML vs AI vs Deep Learning
- Types of ML: Supervised, Unsupervised, Reinforcement

📚 **Resources**:

- Andrew Ng Week 1 (Coursera)
- Kaggle: Intro to ML

🔹 **2. Python Libraries for ML**

Learn how to use key Python libraries:

- **NumPy** – Arrays, broadcasting, linear algebra
- **Pandas** – DataFrames, filtering, grouping
- **Matplotlib & Seaborn** – Data visualization
- **Scikit-learn** – ML algorithms

📚 **Resources**:

- FreeCodeCamp Pandas + NumPy
- Kaggle: Pandas, NumPy, Data Viz

🔹 **3. Data Preprocessing & Exploration**

- Handling missing values
- Feature scaling (StandardScaler, MinMaxScaler)
- Encoding categorical data (LabelEncoder, OneHotEncoder)
- Train-test split

📚 **Resources**:

- Sklearn docs: preprocessing
- Kaggle Datasets to practice

🔹 **4. Supervised Learning Algorithms**

📘 **a. Linear Regression**

- Concept & cost function
- Gradient descent
- Implement with scikit-learn

📚 Andrew Ng Week 2 & 3

📘 **b. Logistic Regression**

- Sigmoid function
- Binary classification
- Evaluation metrics: accuracy, precision, recall, F1-score

📘 **c. Decision Trees & Random Forest**

- Overfitting in Decision Trees
- Ensemble methods (Bagging, Random Forest)
- Feature importance

📘 **d. K-Nearest Neighbors (KNN)**

- Distance metrics (Euclidean)
- Choosing the right k

📘 **e. Naive Bayes**

- Bayes Theorem
- Text classification (spam detection)

📘 **f. Support Vector Machine (SVM)**

- Margins and hyperplanes
- Kernels (linear, RBF)

📚 **Resources for all supervised learning**:

- Scikit-learn tutorials
- Kaggle ML Course
- StatQuest on YouTube

🔹 **5. Model Evaluation**

- Confusion Matrix
- ROC-AUC Curve
- Cross-validation
- Bias-Variance Tradeoff
- Overfitting vs Underfitting

🔹 **6. Unsupervised Learning Algorithms**

📘 **a. K-Means Clustering**

- Clustering vs classification
- Elbow method to choose K

📘 **b. Hierarchical Clustering (optional)**

- Dendrograms

📚 Kaggle: Clustering

🔹 **7. Mini Projects / Practice**

- Iris dataset (classification)
- Titanic dataset (binary classification)

- House price prediction (regression)
- Customer segmentation (clustering)

📚 Datasets:

- Kaggle Datasets
- UCI ML Repository

🔹 **8. Extra Concepts (Optional)**

- Dimensionality Reduction (PCA)
- Feature engineering & selection
- Hyperparameter tuning (GridSearchCV)

📦 **Final Output of Phase 1**

✅ By the end of Phase 1, you should be able to:

- Preprocess and visualize real-world data
- Implement ML models using scikit-learn
- Evaluate and improve model performance
- Build end-to-end ML projects

🔁 **Suggested Learning Flow (Timeline)**

Week	Topics
1	ML Introduction, NumPy, Pandas, Data Viz
2	Data Preprocessing, Linear/Logistic Regression
3	Decision Trees, Random Forest, KNN
4	Naive Bayes, SVM
5	Model Evaluation, Unsupervised Learning (K-Means)
6	Mini Projects, Cross-validation, PCA
7–8	Portfolio Projects, Kaggle practice