I. GENERAL AI & ML CONCEPTS

Q1: What is the difference between AI, Machine Learning, and Deep Learning?
A: AI is the broad field of creating intelligent systems, ML is a subset where machines learn from data, and DL is a subset of ML using deep neural networks.

Q2: What are the main types of Machine Learning?
A: Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning.

Q3: Explain the difference between supervised and unsupervised learning.
A: Supervised uses labeled data to predict outcomes, while unsupervised uses unlabeled data to find patterns or clusters.

Q4: What is overfitting and underfitting?
A: Overfitting → model learns noise and fails to generalize. Underfitting → model is too simple and performs poorly.

Q5: How do you evaluate a machine learning model?
A: Using metrics like Accuracy, Precision, Recall, F1-score, ROC-AUC, RMSE, and Cross-validation.

Q6: What is the bias-variance trade-off?
A: High bias causes underfitting, high variance causes overfitting. The goal is to balance both.

Q7: Explain cross-validation and why it's important.
A: Splits data into folds for training/testing multiple times → ensures robust and reliable evaluation.

Q8: What are precision, recall, F1-score, and accuracy?
A: Accuracy = % correct predictions. Precision = correct positives out of predicted positives. Recall = correct positives out of actual positives. F1 = harmonic mean of precision & recall.

Q9: What is the difference between classification and regression?
A: Classification predicts categories, regression predicts continuous values.

Q10: What are some real-world applications of AI/ML?
A: Fraud detection, recommendations, self-driving cars, chatbots, healthcare diagnosis.

II. DATA PREPROCESSING & FEATURE ENGINEERING

Q11: How do you handle missing data in a dataset?
A: Drop rows/columns, fill with mean/median/mode, interpolation, or model-based imputation.

Q12: What is normalization and standardization?
A: Normalization scales values to [0,1]. Standardization transforms to mean=0, std=1.

Q13: What is one-hot encoding? When would you use it?
A: Converts categorical variables into binary vectors; used for nominal categories.

Q14: How do you handle categorical variables?
A: One-hot encoding, label encoding, or target encoding.

Q15: What is feature selection and why is it important?
A: Choosing the most relevant features → reduces overfitting, speeds training, improves accuracy.

Q16: What is dimensionality reduction? Explain PCA.
A: Reducing features while preserving variance. PCA projects data onto principal components.

Q17: What is the curse of dimensionality?
A: Higher dimensions cause sparse data, longer training, and weaker patterns.

Q18: What is feature scaling and when should it be applied?
A: Scaling features (normalization/standardization) so all contribute equally → needed for distance-based algorithms.

Q19: How do you deal with imbalanced datasets?
A: Resampling (SMOTE, undersampling), class weights, or anomaly detection methods.

Q20: Explain the role of EDA in ML.
A: Exploratory Data Analysis helps understand data, find patterns, outliers, and correlations → guides preprocessing.

III. MACHINE LEARNING ALGORITHMS

Q21: How does the Decision Tree algorithm work?
A: Splits data by feature conditions into branches until a prediction is made.

Q22: What is the difference between bagging and boosting?
A: Bagging trains models in parallel to reduce variance; boosting trains sequentially to reduce bias.

Q23: Explain how the K-Nearest Neighbors algorithm works.
A: Predicts by majority vote (classification) or average (regression) of nearest neighbors.

Q24: What is the intuition behind Support Vector Machines?
A: Finds the best hyperplane that maximizes margin between classes.

Q25: How does Naive Bayes classifier work?
A: Uses Bayes' theorem assuming feature independence to calculate class probabilities.

Q26: What is the difference between Random Forest and XGBoost?
A: Random Forest uses bagging of decision trees, while XGBoost uses boosting for faster and more accurate results.

Q27: How do gradient descent and stochastic gradient descent differ?
A: Gradient Descent uses the full dataset per step; SGD uses one sample or batch → faster but noisier.

Q28: Explain logistic regression and where it's used.
A: Linear model with sigmoid output, used for binary classification.

Q29: What are ensemble models?
A: Combine multiple models (bagging, boosting, stacking) to improve performance.

Q30: What is the difference between L1 and L2 regularization?
A: L1 (Lasso) shrinks some coefficients to zero (feature selection). L2 (Ridge) penalizes large coefficients smoothly.

IV. DEEP LEARNING & NEURAL NETWORKS

Q31: What is a perceptron?
A: A basic neural unit that calculates weighted sum + activation function.

Q32: How do activation functions like ReLU, Sigmoid, and Tanh work?
A: ReLU outputs positive values, Sigmoid maps to 0–1, Tanh maps to -1–1.

Q33: What are epochs, batch size, and learning rate?
A: Epoch = full pass over dataset, batch size = samples per update, learning rate = step size for weight updates.

Q34: What is the vanishing gradient problem?
A: Gradients shrink in deep nets, making training slow or ineffective.

Q35: What is the difference between CNN and RNN?
A: CNN works for spatial data (images), RNN for sequential data (text/time series).

Q36: How does an LSTM work and where is it used?
A: An RNN with memory cells to capture long-term dependencies, used in NLP and time series.

Q37: What are convolutional layers in CNN?
A: Layers that apply filters to extract local spatial patterns.

Q38: What is transfer learning?
A: Using pre-trained models on new tasks to save data and training time.

Q39: What is dropout and why is it used?
A: Randomly deactivates neurons during training to prevent overfitting.

Q40: How do you prevent overfitting in deep learning models?
A: Use dropout, regularization, early stopping, data augmentation.

V. NATURAL LANGUAGE PROCESSING

Q41: What is tokenization in NLP?
A: Splitting text into smaller units like words or sentences.

Q42: How do word embeddings like Word2Vec or GloVe work?
A: Represent words as dense vectors that capture semantic meaning.

Q43: What is the difference between stemming and lemmatization?
A: Stemming cuts words to base form, lemmatization uses rules/dictionaries for proper root words.

Q44: What is TF-IDF and why is it used?
A: Weighs word importance based on frequency in one document vs across documents.

Q45: What are Transformers in NLP?
A: Deep models using attention mechanism (e.g., BERT, GPT) for language tasks.

VI. MODEL DEPLOYMENT & MLOps

Q46: How do you save and load a trained model in Python?
A: Using `pickle`, `joblib`, or framework-specific methods like `model.save()`.

Q47: What is model drift and how do you monitor it?
A: When data changes reduce model performance; monitor with metrics and retrain as needed.

Q48: How do you deploy a machine learning model as an API?
A: Wrap model in Flask/FastAPI and expose prediction endpoints.

Q49: What are common tools for deploying ML models?
A: Flask, FastAPI, Streamlit, Docker, Kubernetes.

Q50: Explain the CI/CD pipeline in MLOps.
A: Automates code integration, testing, and deployment for reliable ML delivery.