1. What is a Markov Decision Process (MDP) and what are its components (states, actions, transition, reward) ?

2. How do you prevent overfitting in deep neural networks ?

3. What is the bias-variance tradeoff in machine learning, and how does it affect model performance ?

4. Explain the concept of a probability density function (PDF) vs. a cumulative distribution function (CDF) .

5. What is a prior distribution, and how does it affect Bayesian inference ?

6. How does a neural network learn? Describe the backpropagation algorithm and how gradients are used .

7. What is an ML feature store, and why might it be useful ?

8. How do ensemble methods like bagging and boosting work, and when might you use them ?

9. Explain Continuous Integration/Continuous Delivery (CI/CD) in MLOps. How do they apply to ML ?

10. What is reward shaping and why might it be used ?

11. What is dropout and why is it used? What is batch normalization and how does it help training ?

12. What are regularization techniques (e.g., L1, L2), and how do they prevent overfitting ?

13. How would you handle missing data in a probabilistic framework ?

14. How do you evaluate an NLP model such as a translation or summarization system (mention BLEU, ROUGE, etc.) ?

15. What are embeddings in deep learning (e.g., word embeddings, image embeddings), and why are they useful ?

16. Explain the vanishing/exploding gradient problem and how it affects training deep networks .

17. What is cross-validation, and why is it used in model evaluation ?

18. What is the difference between generative and discriminative models? Give examples of each .

19. What is word sense disambiguation and how might an NLP system handle it ?

20. What is algorithmic bias, and how can it arise in AI systems ?

21. What is an activation function? Compare common ones (ReLU, sigmoid, tanh) and explain why deep networks often use ReLU .

22. How would you mitigate bias in a dataset or model ?

23. What are some common image preprocessing techniques before feeding data into a neural network ?

24. Explain classification metrics such as precision, recall, and F1 score, and when you would prefer one over the others .

25. Describe how gradient descent variants (SGD, Adam, RMSprop) differ and when you might choose each .

26. What is the role of learning rate in training a neural network? How do you choose or adapt it ?

27. Describe RNNS, LSTMs, and GRUS. Why are LSTMs/GRUs used for sequential data ?

28. What is a Bayesian network (belief network)? How does it represent dependencies between variables ?

29. How do you evaluate a model's performance? Describe metrics for classification (e.g., accuracy, precision, recall, ROC-AUC) and regression (e.g., MSE, R^2) .

30. Explain underfitting and overfitting. How can you detect and address each ?

31. What is the exploration vs exploitation dilemma in RL, and how can it be addressed (e.g., ε-greedy, UCB) ?

32. What is transfer learning in deep learning, and when would you use it ?

33. Explain the difference between bag-of-words and word embeddings (e.g., word2vec, GloVe). Why are embeddings powerful ?

34. Explain actor-critic algorithms at a high level .

35. How does gradient descent work? What are some challenges associated with it (e.g., local minima, learning rate choice) ?

36. What is a graphical model? How does it help in probabilistic reasoning ?

37. What are some challenges of RL (e.g., sample inefficiency, sparse rewards) ?

38. How does an object detection model (e.g., YOLO, Faster R-CNN) differ from an image classification model ?

39. What is a language model? How do n-gram models differ from modern transformer-based models ?

40. What is the No Free Lunch Theorem in machine learning, and what does it imply for model selection ?

41. What is explainability or interpretability in AI, and why is it important ?

42. How does the TF-IDF statistic work, and when might you use it ?

43. What is model governance, and why is it important in ML workflows ?

44. Explain different notions of fairness in machine learning (e.g., equalized odds, demographic parity) and potential trade-offs .

45. What strategies exist to make AI systems more transparent or explainable (e.g., model-agnostic methods like LIME/SHAP) ?

46. What is an autoencoder? How can autoencoders be used for tasks like dimensionality reduction or denoising ?

47. What is semantic segmentation versus instance segmentation ?

48. Why are GPU/TPU accelerators especially beneficial for computer vision tasks ?

49. What are Intersection over Union (IoU) and mean Average Precision (mAP), and how are they used in object detection evaluation ?

50. Describe the assumptions of linear regression and how you would check them .

51. What is data augmentation in computer vision, and why is it helpful ?

52. What is tokenization in NLP, and why is it important for text processing ?

53. What does the term "ethical AI" mean to you, and what principles are important (e.g., transparency, accountability, justice) ?

54. Explain feature selection and feature engineering. What methods can be used for selecting features ?

55. How can you automate retraining and redeployment of models when new data arrives ?

56. Explain how containerization (e.g., Docker) and orchestration (e.g., Kubernetes) are used in ML deployment .

57. Explain how sampling methods (e.g., Monte Carlo, Gibbs sampling) are used in probabilistic models .

58. Explain the difference between on-policy and off-policy learning (e.g., SARSA vs Q-learning) .

59. Define ROC curve and AUC. What do they tell you about a binary classifier ?

60. What is a convolution operation in a CNN, and how do filters (kernels) work ?

61. What is the curse of dimensionality, and how can it impact machine learning algorithms ?

62. Explain the attention mechanism and how it is used in Transformer architectures (e.g., BERT, GPT) .

63. What is model drift (data drift or concept drift), and how can you detect and mitigate it ?

64. Describe a Gaussian Mixture Model and its use cases .

65. What is the Bellman equation for a value function ?

66. How does temporal difference (TD) learning work ?

67. Describe Generative Adversarial Networks (GANs). How do the generator and discriminator work together ?

68. Explain a Markov chain and a Hidden Markov Model (HMM) in broad terms .

69. What is the role of a feature extractor vs a classifier in CNN-based pipelines ?

70. How do non-max suppression algorithms work for selecting bounding boxes in detection ?

71. Why is monitoring of deployed ML models important, and which metrics would you monitor ?

72. How do transformers (like BERT or GPT) differ from RNNs in processing text ?

73. How do you version control datasets and models in an ML project ?

74. How does Deep Q-Network (DQN) differ from classical Q-learning ?

75. What is informed consent in data collection, and why is it important for AI ethics ?

76. What are the main steps of the ML development lifecycle (data, training, deployment, monitoring) ?

77. Explain pooling layers (max pooling vs average pooling) and their effect on feature maps .

78. What are some common NLP preprocessing steps (e.g., stop-word removal, stemming, lemmatization) and why are they used ?

79. How do privacy regulations like GDPR impact AI data collection and model building ?

80. Describe the Expectation-Maximization (EM) algorithm and when you would use it .

81. Describe policy gradient methods and how they differ from value-based methods .

82. What is the difference between a generative model and a discriminative model? Give examples .

83. Explain transfer learning in the context of CV (e.g., using pre-trained CNNs on ImageNet for new tasks) .

84. Describe canary deployments or A/B testing for ML models in production .

85. What are function approximators (e.g., neural networks) used for in RL ?

86. State Bayes' theorem and explain each term (prior, likelihood, posterior). Examples.

87. What is maximum likelihood estimation (MLE)? How does it differ from maximum a posteriori (MAP) estimation ?

88. What is the difference between batch inference and real-time inference, and when would you use each ?

89. Explain self-supervised learning in NLP (e.g., masked language modeling) and why it's useful .

90. Describe a popular CNN architecture (e.g., ResNet, VGG) and its key ideas (like residual connections) .

91. Discuss the trade-off between model performance and fairness. Can they conflict ?

92. Give an example of an Al application that raised ethical concerns, and how those concerns were addressed .

93. What is PCA (Principal Component Analysis) and when would you use it？

94. How might an Al model inadvertently cause harm (e.g., reinforcement of stereotypes, safety issues)？

95. Compare Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in terms of architecture and applications .

96. How can AI systems discriminate against protected groups, and how would you test for it？

97. Explain discounted reward and the role of the discount factor in RL .

98. What is the attention mechanism, and how has it improved NLP models？

99. What are some challenges unique to deploying ML models compared to traditional software？

100. Explain the architecture of a sequence-to-sequence (seq2seq) model with attention. What tasks does it solve？