**Q1: In Large Language Models (LLMs), what is the primary benefit of using Mixture of Experts (MoE) architectures?**

A) They improve training stability by using batch normalization  
B) They reduce the number of layers required for deep learning  
C) They eliminate the need for self-attention mechanisms  
D) They improve efficiency by activating only a subset of model parameters per inference

**Q2: What is the main advantage of Flask over Django in terms of Web development in Python?**

A) It has minimal setup requirements and easy to learn comparing with Django

B) Offers user login authentication system

C) Compatible with common Web development technologies like HTML

D) It is more flexible than Django for dynamic and multi-page web development

**Q3: Which of the following is a key advantage of using Python for AI development?**

A) Python’s built-in AI capabilities eliminate the need for external libraries or frameworks

B) Python offers extensive AI and machine learning libraries, a simple syntax, and strong community support, making development more efficient

C) Python is the only programming language capable of implementing AI models due to its exclusive AI-focused features.

D) Python’s execution speed is significantly faster than C++ and Java, making it the best choice for high-performance AI applications.

**Q4: Which of the following is a crucial decision criterion when selecting a computer vision model for an AI application?**

A) Programming Language Used – Prioritizing models developed exclusively in Python over other languages

B) Random Model Selection – Choosing any pre-trained model without considering dataset relevance or deployment constraints

C) Ignoring Hardware Limitations – Developing a model without considering GPU/CPU capabilities or memory constraints

D) Accuracy vs. Speed Trade-off – Balancing model precision with real-time processing requirements based on application needs

**Q5: Which of the following tasks cannot be performed directly with the NLTK library in Python?**

A) Tokenizing text into words or sentences

B) Performing part-of-speech tagging and named entity recognition

C) Performing sentiment analysis using pre-trained models

D) Building and training custom machine learning models for text classification

**Q6: Which optimization algorithm is commonly used for training deep neural networks and helps mitigate vanishing gradients?**

A) Stochastic Gradient Descent (SGD)

B) Adam

C) Newton’s Method

D) K-Nearest Neighbors (KNN)

**Q7: In the Transformer architecture, what is the primary role of the self-attention mechanism?**

A) Reduce computational complexity

B) Capture long-range dependencies between words

C) Perform dimensionality reduction

D) Convert words into one-hot vectors

**Q8: Which of the following is NOT a common way to prevent overfitting in deep learning?**

A) Dropout

B) L2 Regularization

C) Increasing model complexity

D) Data augmentation

**Q9: What does the term "LLM alignment" refer to in AI safety?**

A) Fine-tuning a model for higher accuracy

B) Ensuring the model behaves according to human intentions and ethics

C) Using reinforcement learning to improve training

D) Optimizing hyperparameters for language tasks

**Q10: In Bayesian machine learning, what does the prior distribution represent?**

A) The distribution of data points after training

B) The initial belief about parameters before observing data

C) The likelihood of new data given a trained model

D) The final probability distribution after applying Bayes' theorem

**Q11: In a high-dimensional dataset, what problem does the Curse of Dimensionality refer to?**

A) More features always improve performance

B) Increasing dimensions makes distances between points less meaningful

C) Deep learning models perform better in high-dimensional spaces

D) The need for more training epochs as dimensions increase

**Q12: What is a key difference between Reinforcement Learning (RL) and Supervised Learning?**

A) RL does not require a labeled dataset

B) RL minimizes loss functions like cross-entropy

C) RL uses static datasets while Supervised Learning continuously updates

D) RL only works for robotic control tasks

**Q13: What is the primary advantage of using LoRA (Low-Rank Adaptation) in fine-tuning Large Language Models?**

A) It allows fine-tuning with significantly fewer trainable parameters

B) It completely replaces the transformer layers

C) It increases inference speed by reducing model size

D) It improves zero-shot learning performance

**Q14: What is the key reason behind the efficiency of FlashAttention in LLMs?**

A) It reduces memory access overhead by optimizing matrix multiplications

B) It replaces attention heads with simpler linear layers

C) It increases token throughput by compressing input embeddings

D) It converts all transformer layers into convolutional layers

**Q15: In contrastive learning for self-supervised models, what is the main idea behind negative samples?**

A) They are additional labels assigned during supervised training

B) They help the model differentiate between similar and dissimilar representations

C) They improve generalization by removing redundant features

D) They are used only in generative models like GANs

**Q16: Question 1: When using K-Fold Cross-Validation, what is the main advantage compared to a simple train-test split?**

A. It eliminates the need for a separate validation set

B. It provides a more robust estimate of model performance by using multiple train-test splits

C. It guarantees that the model will never overfit to the training data

D. It automatically tunes hyperparameters without additional computation

**Q17: Which of the following statements about L1 (Lasso) and L2 (Ridge) regularization is TRUE?**

A. L1 regularization tends to produce sparse models while L2 regularization does not

B. L2 regularization is always superior to L1 for preventing overfitting

C. L1 regularization adds the square of the weights to the loss function

D. Both L1 and L2 regularization increase model complexity

**Q18: In a neural network using backpropagation, what problem does the vanishing gradient issue cause?**

A. The model learns too quickly and overshoots the optimal weights

B. Earlier layers learn more slowly than later layers because gradients become extremely small

C. The model requires more GPU memory during training

D. The model produces inconsistent predictions for similar inputs

**Q19: Which of the following techniques is not typically used to handle imbalanced datasets?**

A. Oversampling the minority class

B. Undersampling the majority class

C. Normalizing feature values to be between 0 and 1

D. Using class weights in the loss function

**Q20: Which of the following ensemble methods combines multiple models by having them vote on the final prediction, with each model trained on a random subset of the training data?**

A. Gradient Boosting

B. AdaBoost

C. Random Forest

D. Stacking