

A company has developed an ML model for image classification. The company wants to deploy the model to production so that a web application can use the model. The company needs to implement a solution to host the model and serve predictions without managing any of the underlying infrastructure. Which solution will meet these requirements?

- ☒ A. Use Amazon SageMaker Serverless Inference to deploy the model.
- ☐ B. Use Amazon CloudFront to deploy the model.
- ☐ C. Use Amazon API Gateway to host the model and serve predictions.
- ☐ D. Use AWS Batch to host the model and serve predictions.

A company is building a chatbot to answer customer questions by using a foundation model (FM). The company wants to confirm that the model is unbiased and fair. Which solution will assess the fairness of the model?

- ☐ A. Fine-tuning a custom model on Amazon Bedrock
- ☐ B. User surveys about chatbot satisfaction
- ☐ C. Automatic model evaluation tasks
- ☒ D. Regular accuracy testing of chatbot responses

A company is building an ML model to analyze archived data. The company must perform inference on large datasets that are multiple GBs in size. The company does not need to access the model predictions immediately.

Which Amazon SageMaker inference option will meet these requirements?

- ☒ A. Batch transform
- ☐ B. Real-time inference
- ☐ C. Serverless inference
- ☐ D. Asynchronous inference

A company has petabytes of unlabeled customer data to use for an advertisement campaign. The company wants to classify its customers into tiers to advertise and promote the company's products. Which methodology should the company use to meet these requirements?

- ☐ A. Supervised learning
- ☒ B. Unsupervised learning
- ☐ C. Reinforcement learning
- ☐ D. Reinforcement learning from human feedback (RLHF)

Which strategy evaluates the accuracy of a foundation model (FM) that is used in image classification tasks?

- ☐ A. Calculate the total cost of resources used by the model.
- ☒ B. Measure the model's accuracy against a predefined benchmark dataset.
- ☐ C. Count the number of layers in the neural network.
- ☐ D. Assess the color accuracy of images processed by the model.

Which part of a prompt needs the user to provide information to set up the situation for the model?

- ☐ A. Output indicator
- ☐ B. Input
- ☒ C. Instruction
- ☐ D. Context