Microsoft.AI-300.v2026-11-26.q76

Exam Code:	AI-300
Exam Name:	Operationalizing Machine Learning and Generative AI Solutions
Certification Provider:	Microsoft
Free Question Number:	76
Version:	v2026-11-26
# of views:	102
# of Questions views:	760
https://www.freecram.net/torrent/Microsoft.AI-300.v2026-11-26.q76.html	

NEW QUESTION: 1
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear on the review screen.
You manage an Azure Machine Learning workspace. The Python script named script.py reads an argument named training_data. The training_data argument specifies the path to the training data in a file named dataset1.csv.
You plan to run the script.py Python script as a command job that trains a machine learning model.
You need to provide the command to pass the path for the dataset as a parameter value when you submit the script as a training job.
Solution: python train.py --training_data training_data
Does the solution meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)
Correct:
* python script.py --training_data ${{inputs.training_data}}
The scipt is named script.py.
For the parameter use ${{inputs.training_data}}
Incorrect:
* python script.py --training_data dataset1.csv
* python script.py dataset1.csv
* python train.py --training_data training_data
Note: Read a TabularDataset, Example
In the Input object, specify the type as AssetTypes.MLTABLE, and mode as InputOutputModes.DIRECT:
* Details omitted*
job = command(
code="./src", # Local path where the code is stored
*-> command="python train.py --inputs ${{inputs.input_data}}",

inputs=my_job_inputs,
environment="<environment_name>:<version>",
compute="cpu-cluster",
)
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-read-write-data-v2

NEW QUESTION: 2
A team deploys a model to a real-time endpoint in Azure Machine Learning. You deploy some updates to the endpoint.
The endpoint returns errors after the new deployment is released.
You need to restore the service as quickly as possible.
What should you do first?
A. Roll back traffic to the previous deployment.
B. Delete the endpoint and immediately redeploy it.
C. Change the authentication type to Azure Machine Learning token-based authentication.
D. Increase the compute size.
Answer: (SHOW ANSWER)
To restore the service as quickly as possible, you can roll back traffic to the previous deployment by updating the traffic allocation settings of your Azure Machine Learning online endpoint.
Azure Machine Learning managed online endpoints support multiple deployments under a single endpoint, allowing for blue-green deployment strategies where you can shift traffic between versions instantly.
Key Benefits of This Approach
Instant Recovery: Traffic shifting is a routing change and does not require redeploying the previous model's code or environment, making it the fastest recovery method.
No Downtime: Because the previous deployment remains "warm" (active but receiving no traffic), the switch happens without interrupting the service.
Isolation for Debugging: You can keep the failing deployment at 0% traffic to inspect its logs using az ml online-deployment get-logs without affecting end users
How to Roll Back Traffic If your new deployment (e.g., "green") is returning errors, you can reallocate 100% of the traffic back to the known stable deployment (e.g., "blue") using the following methods:
* Azure CLI: Use the az ml online-endpoint update command to set the traffic percentage:
az ml online-endpoint update --name <your-endpoint-name> --traffic "blue=100 green=0"
* Azure Machine Learning Studio:
Navigate to Endpoints in the left menu.
Select your specific real-time endpoint.
Go to the Details or Live Traffic tab.
Adjust the traffic percentages so the previous deployment receives 100% and the failing deployment receives 0%.
Select Update or Save to apply the changes immediately
Reference:
https://learn.microsoft.com/en-us/answers/questions/1275110/azure-ml-v2-yaml-code-for-live-traffic-allocation

NEW QUESTION: 3

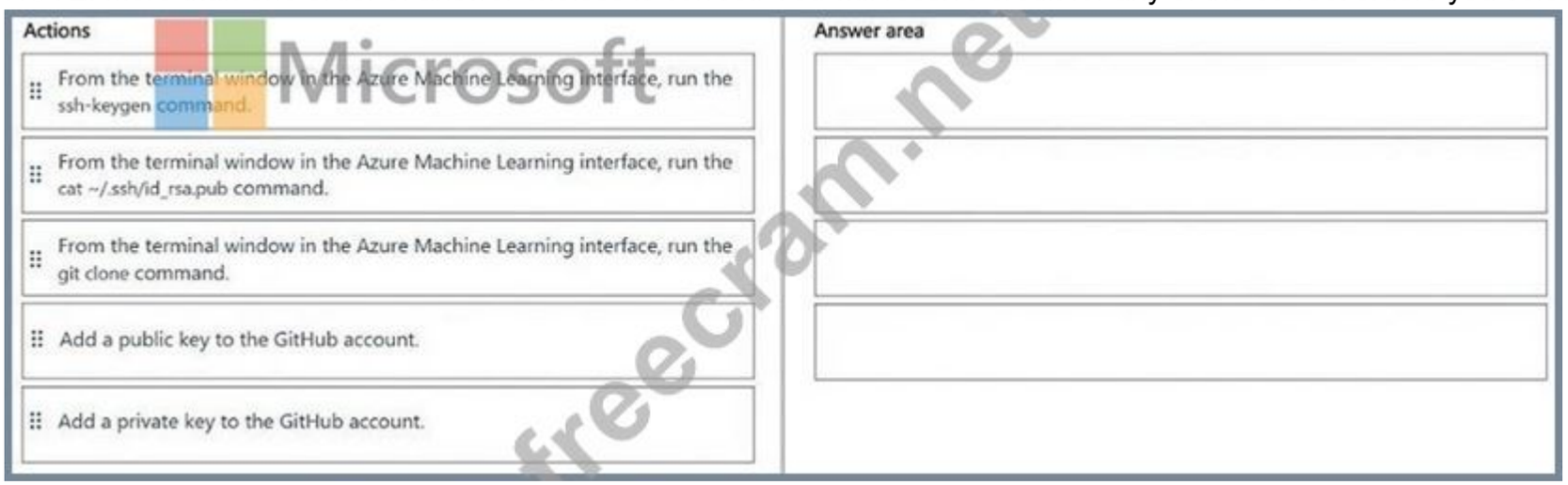
You have an Azure subscription named Sub1 that contains an Azure Machine Learning workspace named Workspace1. Workspace1 contains the following assets:
- a registered MLflow model named Model1
- an online endpoint named Endpoint1
Outbound network connectivity from Endpoint1 is blocked.
You need to deploy Model1 to Endpoint1.
What should you do first?
A. In Sub1, create an Azure Machine Learning registry.
B. In Workspace1, create a linked service.
C. In Workspace1, create a package.
D. In Sub1, create a private endpoint.
Answer: C (LEAVE A REPLY)
To successfully deploy the registered MLflow model to an online endpoint that lacks outbound internet connectivity, you must use model packaging to build a deployment package before deploying.
By default, Azure Machine Learning deploys MLflow models using a "no-code deployment" approach, which requires an outbound internet connection during container runtime to dynamically download and install Python dependencies listed in the model's conda.yaml file.
Because outbound connectivity is blocked, this process fails. Packaging the model bundles the model binaries, dependencies, and environment ahead of time, entirely removing the requirement for runtime internet access.
Reference:
https://docs.azure.cn/en-us/machine-learning/how-to-deploy-mlflow-models-online-endpoints

NEW QUESTION: 4
You train models on GPU-enabled clusters but deploy them on CPU-based endpoints. Recently, inference failures occur due to incompatible dependencies. What should you do to ensure consistency?
A. Use same compute for training and inference
B. Define and reuse environment configurations
C. Increase endpoint compute size
D. Use batch endpoints
Answer: (SHOW ANSWER)
Defining and reusing environment configurations ensures that dependencies remain consistent between training and inference. This prevents runtime errors caused by mismatched libraries.
Using identical compute resources is unnecessary and inefficient, as consistency depends on environment configuration rather than hardware.

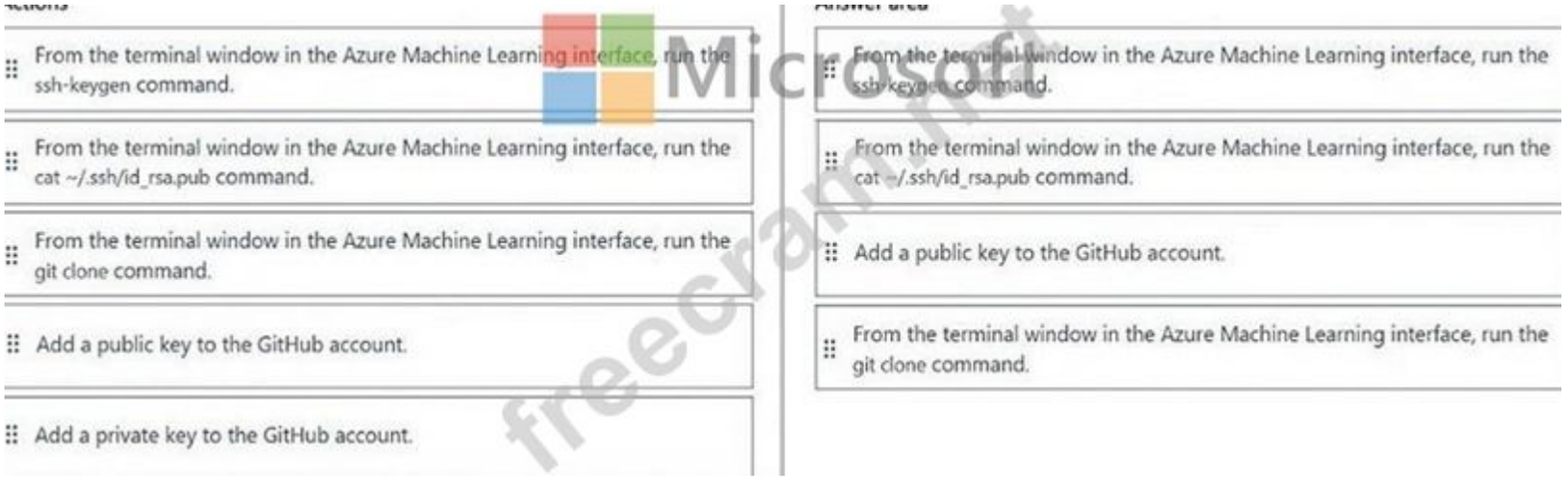
NEW QUESTION: 5
Multiple teams need access to approved models with version tracking, lineage, and governance controls. Models must be discoverable and reusable across projects. What Azure ML feature should you use?
A. Blob storage containers
B. Model registry
C. Git repositories
D. Data lake
Answer: (SHOW ANSWER)

A model registry provides centralized management of models, including versioning, lineage tracking, and governance. This enables teams to discover, share, and reuse models efficiently.
Simple storage solutions like blob storage lack these advanced capabilities and do not support proper lifecycle management.

NEW QUESTION: 6
Drag and Drop Question
You have an existing GitHub repository containing Azure Machine Learning project files.
You need to clone the repository to your Azure Machine Learning shared workspace file system.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

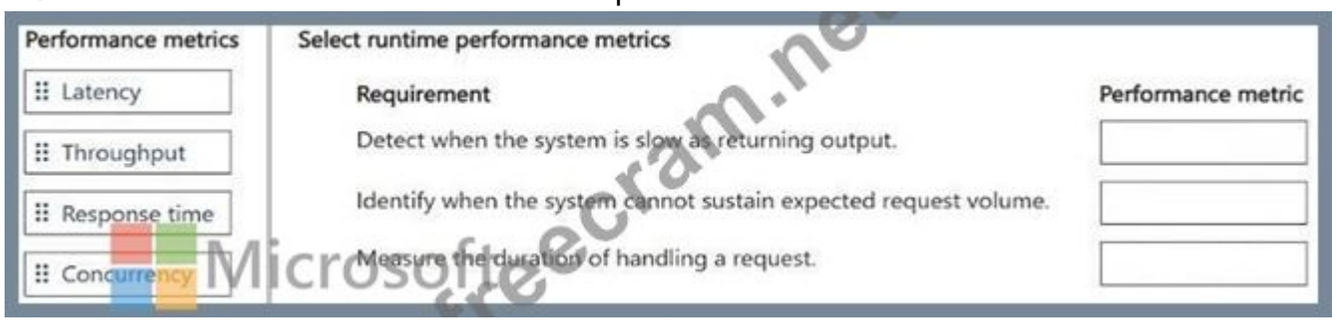


Answer:

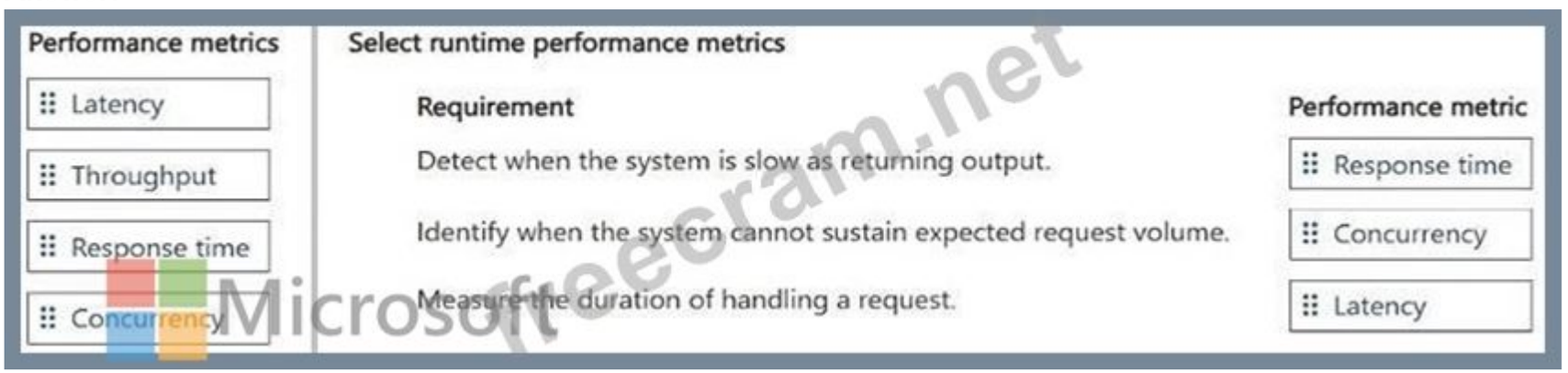


NEW QUESTION: 7
Drag and Drop Question
A team operates a generative AI-powered customer support assistant built on Microsoft Foundry.
The application serves users globally and supports both real-time chat interactions and batch summarization jobs.
The team must ensure that the application continues to meet defined service-level objectives (SLO) as usage increases.
The team requires visibility into runtime behavior to identify performance regressions that affect the user experience and system capacity.
You need to select the performance metrics that meet the requirements.

Which performance metric should you monitor for each requirement? To answer, move the appropriate performance metrics to the correct requirements. You may use each performance metric once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 8

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear on the review screen.
You manage an Azure Machine Learning workspace. The Python script named script.py reads an argument named training_data. The training_data argument specifies the path to the training data in a file named dataset1.csv.
You plan to run the script.py Python script as a command job that trains a machine learning model.
You need to provide the command to pass the path for the dataset as a parameter value when you submit the script as a training job.
Solution: python script.py dataset1.csv
Does the solution meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)
Correct:
* python script.py --training_data ${{inputs.training_data}}
The scipt is named script.py.
For the parameter use ${{inputs.training_data}}
Incorrect:
* python script.py --training_data dataset1.csv
* python script.py dataset1.csv
* python train.py --training_data training_data

Note: Read a TabularDataset, Example
In the Input object, specify the type as AssetTypes.MLTABLE, and mode as InputOutputModes.DIRECT:
* Details omitted*
job = command(
code="./src", # Local path where the code is stored
*-> command="python train.py --inputs ${{inputs.input_data}}",
inputs=my_job_inputs,
environment="<environment_name>:<version>",
compute="cpu-cluster",
)
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-read-write-data-v2

NEW QUESTION: 9
A team is validating a generative AI assistant for a company. The assistant generates responses by using internal knowledge sources.
The company requires assurance that responses are accurate, supported by sources, and related to the user prompts before enabling production access.
You need to implement quality metrics that confirm the assistant produces reliable and meaningful responses.
Which two evaluation metrics should you use? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. Groundedness
B. Relevance
C. Harmfulness
D. Tone
E. Fairness
Answer: (SHOW ANSWER)
To ensure your Azure Machine Learning generative AI assistant (which utilizes a Retrieval- Augmented Generation or RAG architecture) produces reliable and meaningful responses before going live, you should use the RAG Triad of built-in quality evaluation metrics.
These primary automated metrics-Groundedness, Relevance, and Response Completeness- directly measure accuracy, connection to internal knowledge sources, and alignment with user prompts.
The Primary Metrics (The RAG Triad)
These metrics are evaluated on a 1-to-5 scale using Azure Machine Learning's built-in, AI- assisted "LLM-as-a-judge" evaluators:
Groundedness: Measures how well the assistant's generated answer aligns only with the information retrieved from your internal knowledge sources. Even if a response is factually correct in the real world, it is penalized if the information cannot be verified inside the retrieved context document. This is your primary defense against hallucinations.
Relevance: Assesses how pertinently the model's generated response directly addresses the user's specific prompt. This checks whether the system understood the user's intent or if it provided an off-topic or distracted response.
Response Completeness: Focuses on the "recall" aspect of the assistant. It measures whether the generated text effectively answers all parts of the user prompt using the ground truth data, ensuring no critical insights or data points are omitted.
Reference:
https://medium.com/thedeephub/a-deep-dive-into-evaluation-in-azure-prompt-flow-dd898ebb158c

NEW QUESTION: 10

A team schedules weekly retraining of a model using Azure ML pipelines. They also want retraining triggered automatically when production data significantly deviates from training data distribution, without duplicating pipeline logic. What should they implement?

A. Two independent pipelines with shared scripts

B. One pipeline triggered by schedule and data drift alerts

C. Azure Function to retrain model manually

D. Notebook-based retraining process

Answer: (SHOW ANSWER)

Using a single pipeline triggered by both a schedule and data drift alerts ensures consistent retraining logic and avoids duplication. This approach minimizes operational overhead and maintenance complexity. Creating multiple pipelines can lead to inconsistencies, duplicated code, and increased effort when updating retraining logic or dependencies.

NEW QUESTION: 11

Hotspot Question

A regression model is deployed to a production online endpoint and is actively serving predictions.

Business stakeholders report that the model's prediction quality has changed.

The endpoint remains available and responsive.

You need to determine which metric to examine to validate each quality concern.

Which metric should you review for each concern? To answer, select the appropriate options in the answer area.

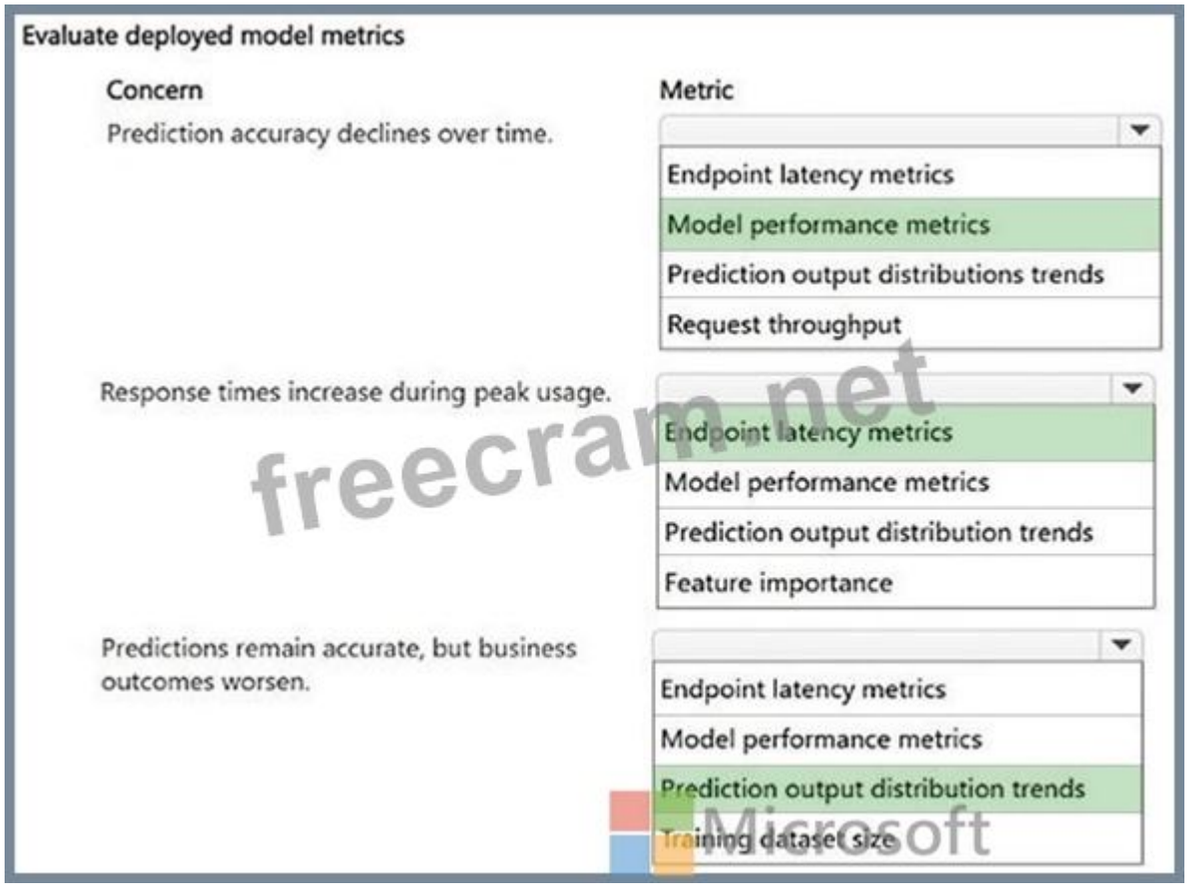
NOTE: Each correct selection is worth one point.

:valuate deployed model metrics

Concern	Metric
Prediction accuracy declines over time.	Endpoint latency metrics Model performance metrics Prediction output distributions trends Request throughput
Response times increase during peak usage.	Endpoint latency metrics Model performance metrics Prediction output distribution trends Feature importance
Predictions remain accurate, but business outcomes worsen.	Endpoint latency metrics Model performance metrics Prediction output distribution trends Training dataset size

Freecram.net
Microsoft
freecram.net

Answer:



NEW QUESTION: 12

Drag and Drop Question

A team iterates prompts used by a generative AI agent. The team must support internal review before releasing changes.

The team must:

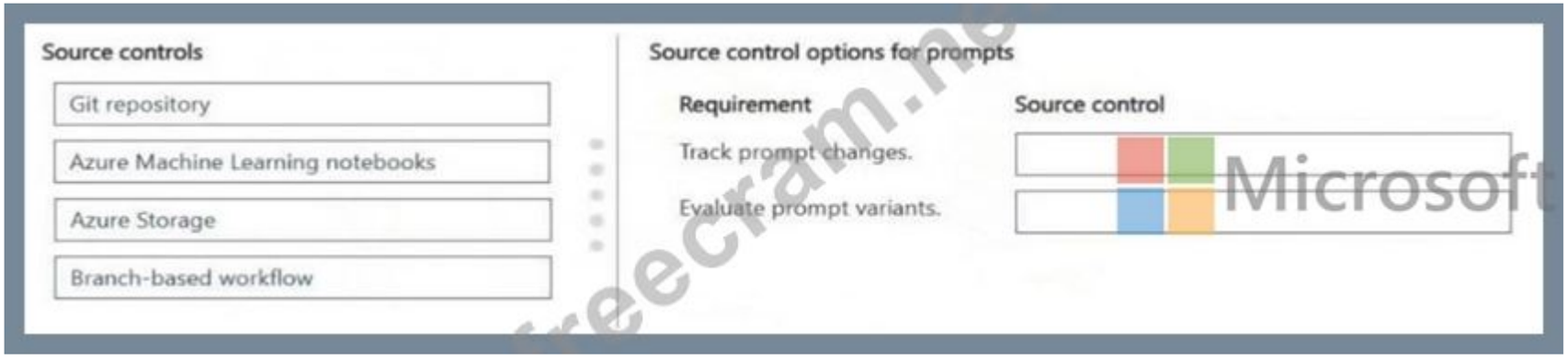
- Track prompt changes with a clear history for audit and rollback.
- Compare prompt variants in parallel without affecting the prompt used in the production environment.

You need to select the appropriate source control approach for each requirement.

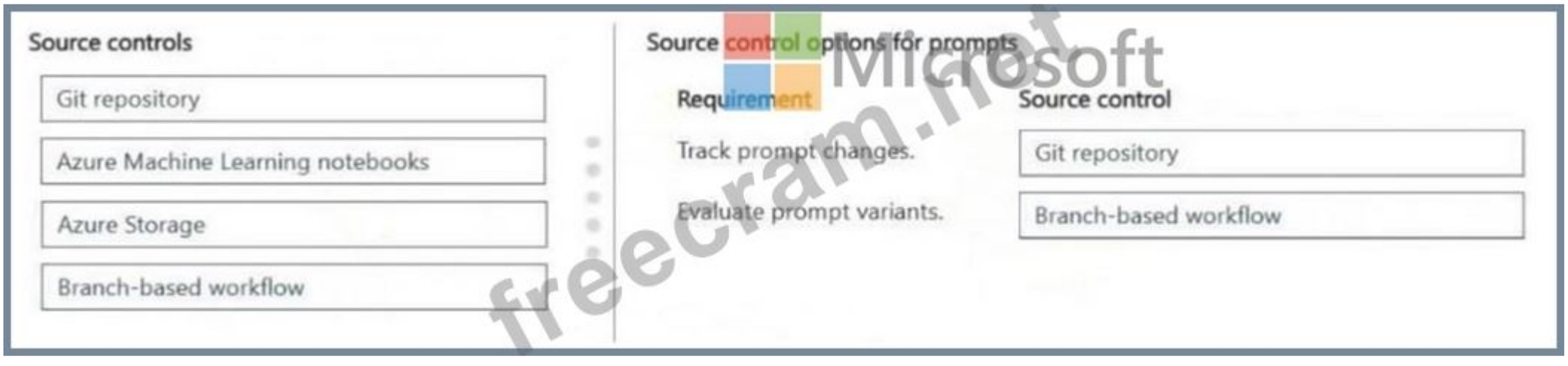
What should you use for each requirement? To answer, move the appropriate source controls to the correct requirements. You may use each source control once, more than once, or not at all.

You may need to move the split bar between panes or scroll to view content.

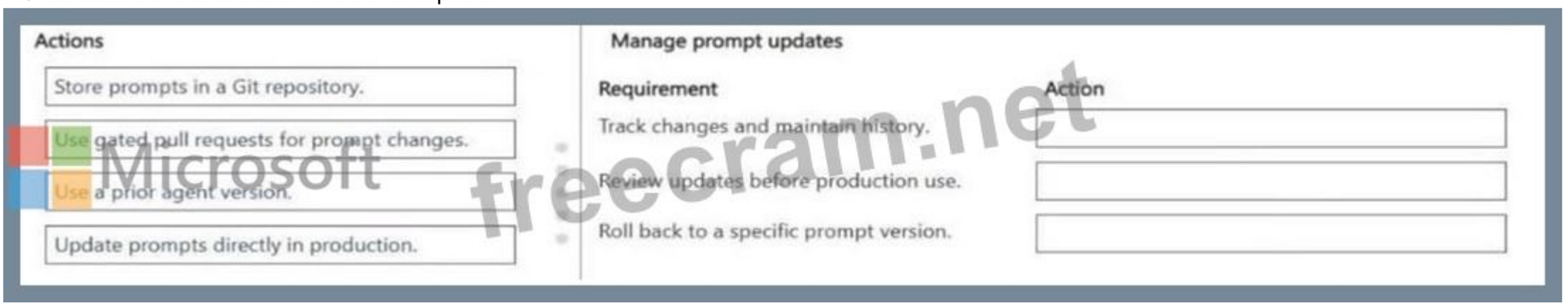
NOTE: Each correct selection is worth one point.



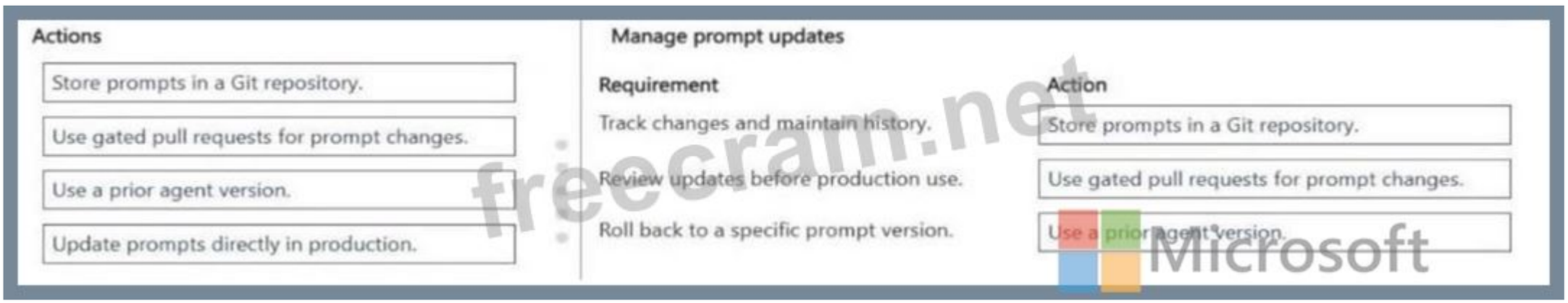
Answer:



NEW QUESTION: 13
Drag and Drop Question
A team manages prompts that are used by a generative AI application built on Microsoft Foundry.
Multiple developers contribute prompt updates, and changes must be reviewed and tracked over time.
The team requires that:
- Prompt changes are reviewed before being applied to the version in
production.
- Previous prompt versions can be restored if issues occur.
- Prompt updates follow the same governance practices as the
application code.
You need to implement a controlled process for managing and updating prompts in production.
How should you manage prompt updates to meet the requirements? To answer, move the appropriate actions to the correct requirements. You may use each action once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:

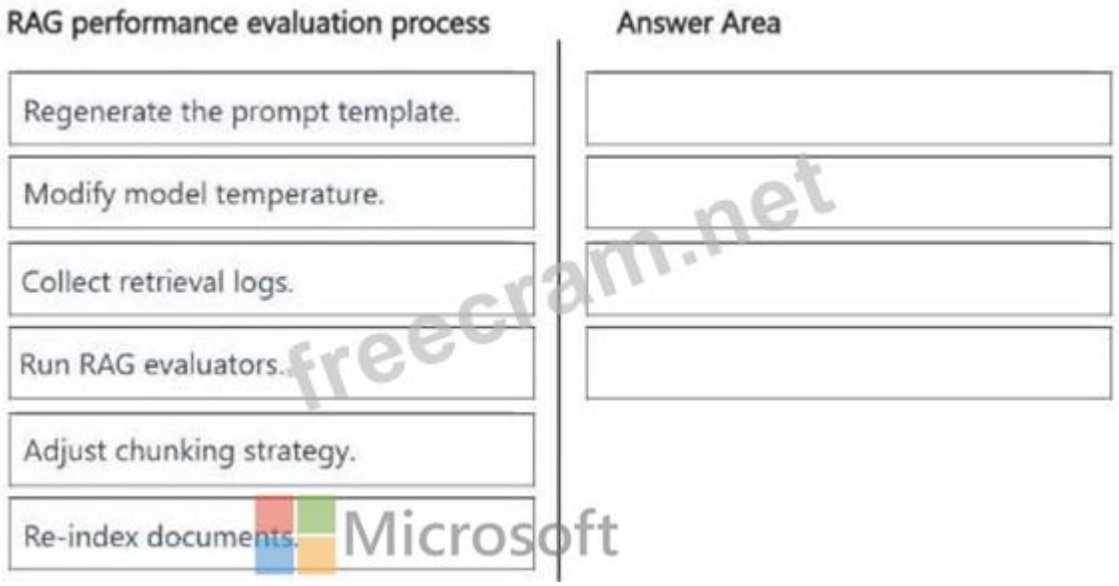


NEW QUESTION: 14
You create an Azure Machine Learning workspace. You train an MLflow-formatted regression model by using tabular structured data.
You must use a Responsible AI dashboard to assess the model.
You need to use the Azure Machine Learning studio UI to generate the Responsible AI dashboard.
What should you do first?
A. Register the model with the workspace.
B. Create the model explanations.
C. Convert the model from the MLflow format to a custom format.
D. Deploy the model to a managed online endpoint.
Answer: (SHOW ANSWER)
The first step you must take is to register the model with the workspace.
To access the no-code, guided wizard for generating a Responsible AI dashboard directly within the Azure Machine Learning studio UI, the trained model must first exist as a recognized asset inside your workspace's model registry.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-responsible-ai-dashboard

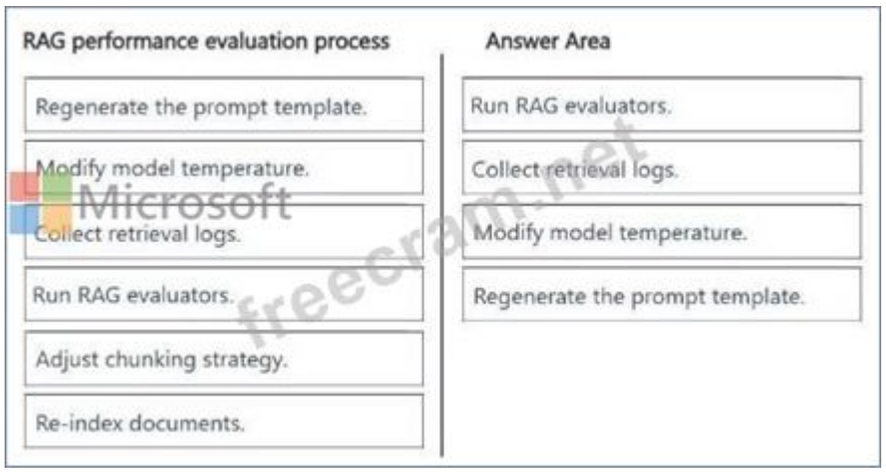
NEW QUESTION: 15
You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters.
In previous model training and tuning runs, many models showed similar performance.
You need to select an early termination policy that meets the following requirements:
- accounts for the performance of all previous runs when evaluating the current run
- avoids comparing the current run with only the best performing run to date Which two early termination policies should you use? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. Bandit
B. Default
C. Median stopping
D. Truncation selection
Answer: (SHOW ANSWER)
The Median Stopping Policy and the Truncation Selection Policy are the most appropriate early termination policies for this scenario. Both evaluate runs based on the performance of all previous runs instead of strictly benchmarking against the single best run.

Median Stopping Policy: This policy calculates the running averages of the primary metric across all historical and current training runs at each evaluation interval. It terminates any ongoing run if its performance is worse than the median of those averages. This directly accounts for all previous run performances rather than just the single best run.
Truncation Selection Policy: This policy evaluates all active runs at each interval and terminates a specified bottom percentage (X%) of the lowest-performing runs. Because it aggregates and compares the entire cohort of runs, it avoids making narrow comparisons against only the best- performing run.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION: 16
Drag and Drop Question
A team is developing a Retrieval-Augmented Generation (RAG) system.
The team requires improvements to the system's retrieval quality to ensure accurate, grounded responses.
You need to assess RAG performance before you can suggest an improvement strategy.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



Explanation:
To properly assess your Azure Retrieval-Augmented Generation (RAG) system's performance before implementing an improvement strategy, you should include the following four steps: Run RAG evaluators, Collect retrieval logs, Modify model temperature, and Regenerate the prompt template.

Step 1: Run RAG evaluators
Run RAG evaluators is the primary method to objectively measure system performance.
Evaluators calculate data-driven metrics like groundedness, relevance, and retrieval precision using tools like Azure AI Studio Evaluators.
Step 2: Collect retrieval logs
Collect retrieval logs provides the raw operational data needed for assessment. Analyzing these logs helps you identify exactly which documents were retrieved, their relevance scores, and where the retrieval pipeline failed to fetch the correct context.
Step 3: Modify model temperature
Modify model temperature: Adjusting the temperature during assessment helps isolate whether poor responses are caused by bad retrieval or by the LLM being too creative (high temperature) or too rigid (low temperature). Testing variations helps establish a performance baseline.
Step 4: Regenerate the prompt template
Regenerate the prompt template: Evaluating how different prompt variations alter the output allows you to assess if the current template is effectively forcing the model to rely only on the retrieved context, which is critical for identifying grounding issues.
Incorrect:
Adjust the chunking strategy
This is an improvement action, not an assessment step. You would perform this optimization strategy after your assessment reveals that information is being cut off or poorly contextualized.
Re-index documents: This is a heavy remediation step. Re-indexing is a time- and resource- consuming strategy used to fix issues once the assessment phase has already proven that the current index or embedding model is faulty.
Reference:
https://flytoleisure.medium.com/guideline-for-building-a-practical-and-effective-rag-retrieval-augmented-generation-application-f6cf50676e37

Valid AI-300 Dumps shared by ExamDiscuss.com for Helping Passing AI-300 Exam! ExamDiscuss.com now offer the **newest AI-300 exam dumps**, the ExamDiscuss.com AI-300 exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com AI-300 dumps with Test Engine here: https://www.examdiscuss.com/Microsoft/exam/AI-300/premium/ (**232** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 17
Hotspot Question
You have an Azure Machine Learning workspace.
You plan to use Azure Machine Learning Python SDK v2 to define a pipeline component that trains an image classification model. The execution logic of the component is contained in the train() function in the file named model_train.py.
You write code to import all required libraries and store it as train_component.py in the same folder that contains model_train.py.
You need to complete the remaining code in train_component.py.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Azure Machine Learning Python SDK v2 code

```
@[ tool | pipeline | command_component ] (
    name="train_image_classification",
    version="1",
    display_name="Train Image Classification",
    description="train image classification",
    environment=dict(
        conda_file=Path(__file__).parent / "conda.yaml",
        image="mcr.microsoft.com/azureml/openmpi4.1.0-ubuntu20.04",
    ),
)
def train_component(
    input_data: Input(type="uri_folder"),
    output_model: Output(type="uri_folder"),
    epochs=10,
):
    from [ train | model_train | train_component ] import train

    train(input_data, output_model, epochs)
```

Microsoft
freecram.net

Answer:

Azure Machine Learning Python SDK v2 code

```
@ [tool | pipeline | command_component] (
    name="train_image_classification",
    version="1",
    display_name="Train Image Classification",
    description="train image classification",
    environment=dict(
        conda_file=Path(__file__).parent / "conda.yaml",
        image="mcr.microsoft.com/azureml/openmpi4.1.0-ubuntu20.04",
    ),
)
def train_component(
    input_data: Input(type="uri_folder"),
    output_model: Output(type="uri_folder"),
    epochs=10,
):
    from [train | model_train | train_component] import train

    train(input_data, output_model, epochs)
```

freecram.net

Explanation:

Box 1: command_component

The @command_component decorator transforms a standard Python function into a reusable pipeline component within Azure ML SDK v2.from model_ Box 2: model_train model_train import train: Because model_train.py resides in the same directory as train_component.py, you import the file directly by its module name (model_train) to access its execution logic inside the component function.

Reference:

https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-component-pipeline-python

NEW QUESTION: 18

You deploy a new model version to a managed online endpoint. You must test it with 10% traffic and automatically roll back if latency or error rate increases beyond threshold. What should you configure?

A. Separate endpoint for testing

B. Traffic splitting with monitoring alerts

C. Batch endpoint validation

D. Manual testing workflow

Answer: (SHOW ANSWER)

Traffic splitting enables controlled rollout of a new model version by directing a percentage of requests to it. Combined with monitoring alerts, it supports automated rollback when performance degrades. Separate endpoints lack built-in traffic management and do not provide seamless or automated rollback capabilities.

NEW QUESTION: 19
Drag and Drop Question
An organization is adopting Microsoft Foundry to support multiple generative AI projects across different teams. Microsoft Foundry platform administrators require centralized governance.
Development teams need isolated environments for experimentation and deployment.
Shared policies must be enforced consistently while allowing teams to work independently.
You need to configure the Microsoft Foundry environment to meet the requirements.
Which Microsoft Foundry components should you configure? To answer, move the appropriate configurations to the correct responsibilities. You may use each configuration once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:



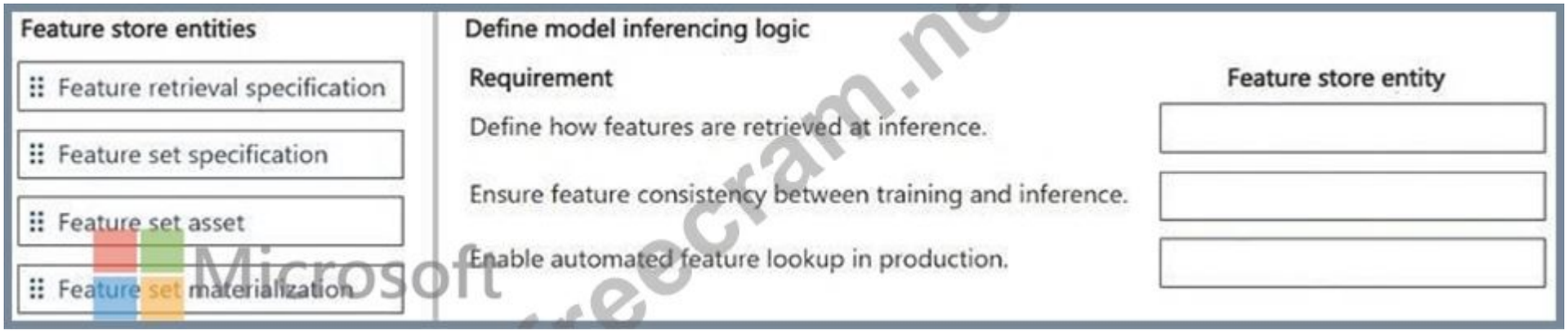
NEW QUESTION: 20
You are preparing training data for a fine-tuning job in Microsoft Foundry.
Real production conversations cannot be used due to compliance requirements.
You need to generate synthetic interaction data that can be used for fine-tuning a generative model.
What should you do?
A. Export model evaluation logs and use them directly as training data.
B. Use a simulator to generate prompt-response interaction data that matches the target task.
C. Enable A/B testing and capture live user traffic for data generation.
D. Run a simulator to produce telemetry logs and trace data from user interactions.
Answer: (SHOW ANSWER)

You can use a simulator or an LLM-as-a-judge pipeline to generate synthetic interaction data for fine-tuning. This technique is standard practice for maintaining strict data privacy while training models on specific business tasks.
Here is how to effectively structure and execute a synthetic data generation pipeline for Microsoft Azure AI Foundry (formerly Azure AI Studio).
Generation Methods
Persona-Driven Simulation: Prompt one LLM to act as a customer and another as your support agent to generate multi-turn dialogues.
Seed Data Expansion: Feed 10-20 hand-written, compliant examples into an LLM and instruct it to generate hundreds of diverse variations.
Schema-Based Evolution: Define variables (e.g., product types, user intents, sentiment levels) and programmatically combine them into prompt templates for an LLM to flesh out.
Reference:
https://www.digitaldividedata.com/blog/synthetic-data-generation-in-gen-ai

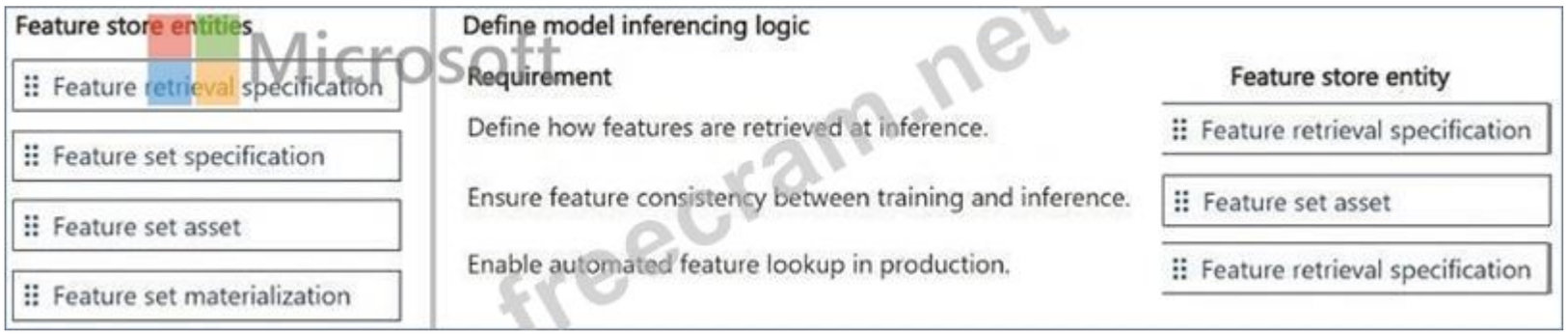
NEW QUESTION: 21
A team is experimenting with traditional models for a classification workflow in Azure Machine Learning.
The team requires a consistent way to manage assets that are created during experimentation.
You need to ensure that artifacts can be reused and governed across projects.
Which asset should you register?
A. Model
B. Component
C. Environment
D. Pipeline
Answer: (SHOW ANSWER)
In an Azure Machine Learning classification workflow, you should register Models.
Registration creates a versioned asset in your workspace or a centralized registry, which is essential for ensuring that artifacts are reusable, governed, and trackable across different projects and environments.
Key Assets for Reuse and Governance
To maintain a consistent and governed workflow, you should focus on registering these specific assets:
Models: The primary artifact. Registering a model allows you to track its lineage (which experiment created it), version it, and deploy it consistently across environments.
Components: These are self-contained pieces of code that perform specific steps in a pipeline (e.g., data cleaning, training). Registering them allows different teams to reuse the same
"traditional" classification logic without rewriting code.
Environments: Encapsulates the software dependencies (Python packages, Docker images) required for your model to run. Registering these ensures reproducibility across different compute targets.
Data Assets: Registering your training and testing datasets as versioned assets ensures that you can always audit exactly what data was used to train a specific model version.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/concept-azure-machine-learning-v2

NEW QUESTION: 22
Drag and Drop Question
A data science team trains a model that depends on features that are stored in a managed feature store.

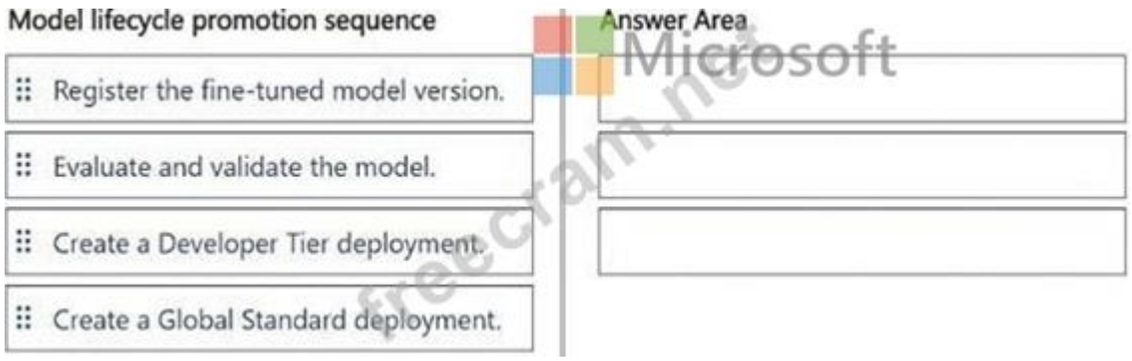
The model is registered in Azure Machine Learning and will be deployed to a real-time endpoint.
After deployment, the model must:
- Retrieve feature values dynamically at inference time.
- Use the same feature definitions that were used during training.
- Run without manual configuration changes across environments.
You need to define feature store entities so that feature retrieval behaves as expected when the model is deployed.
Which feature store entity should you select for each requirement? To answer, move the appropriate feature store entities to the correct requirements. You may use each feature store entity once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



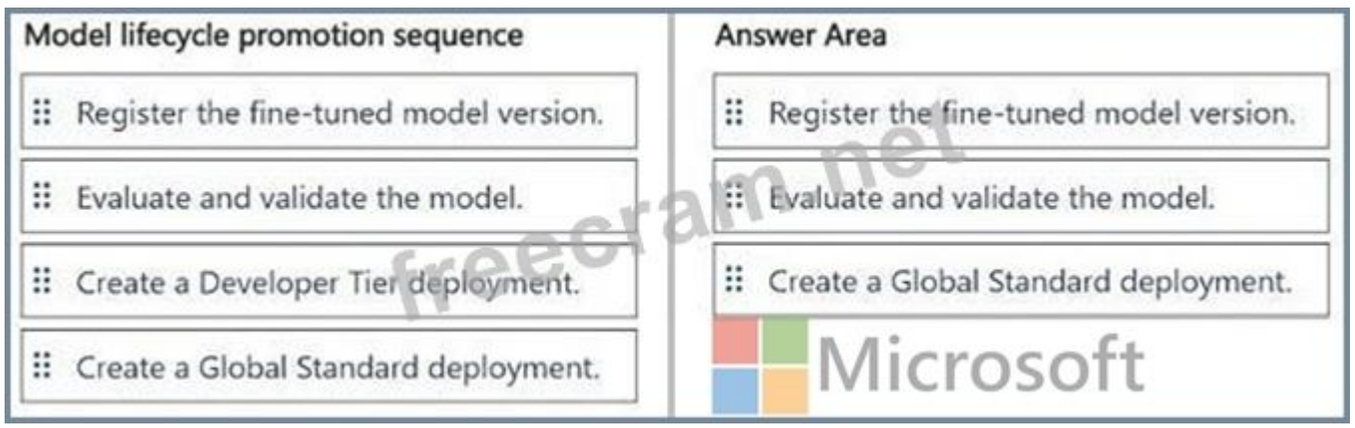
Answer:



NEW QUESTION: 23
Drag and Drop Question
You complete the fine-tuning of a generative model in Microsoft Foundry. The fine-tuned model now appears as a new model variant in your development environment.
The deployment process must ensure that proper validation and control is maintained.
You need to promote the fine-tuned model from development to production.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



NEW QUESTION: 24

Hotspot Question

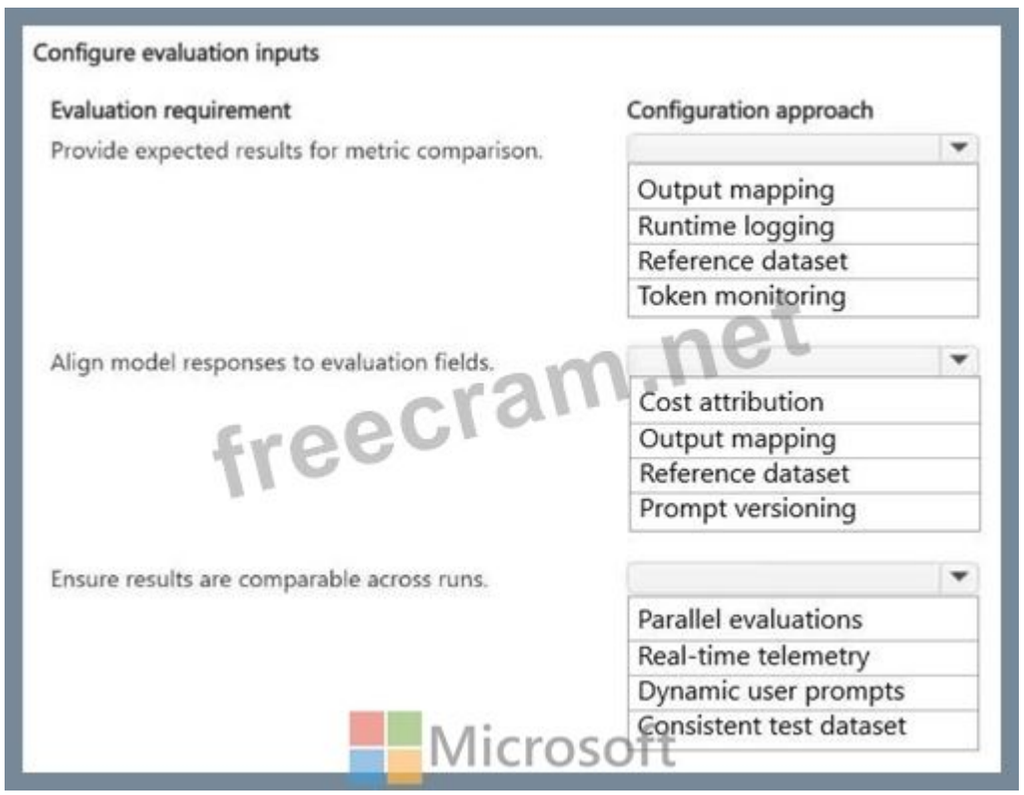
A team is preparing a generative AI application for production deployment. The application generates structured responses that must be evaluated for quality before each release.

The organization requires repeatable evaluation results that can be compared across builds and environments.

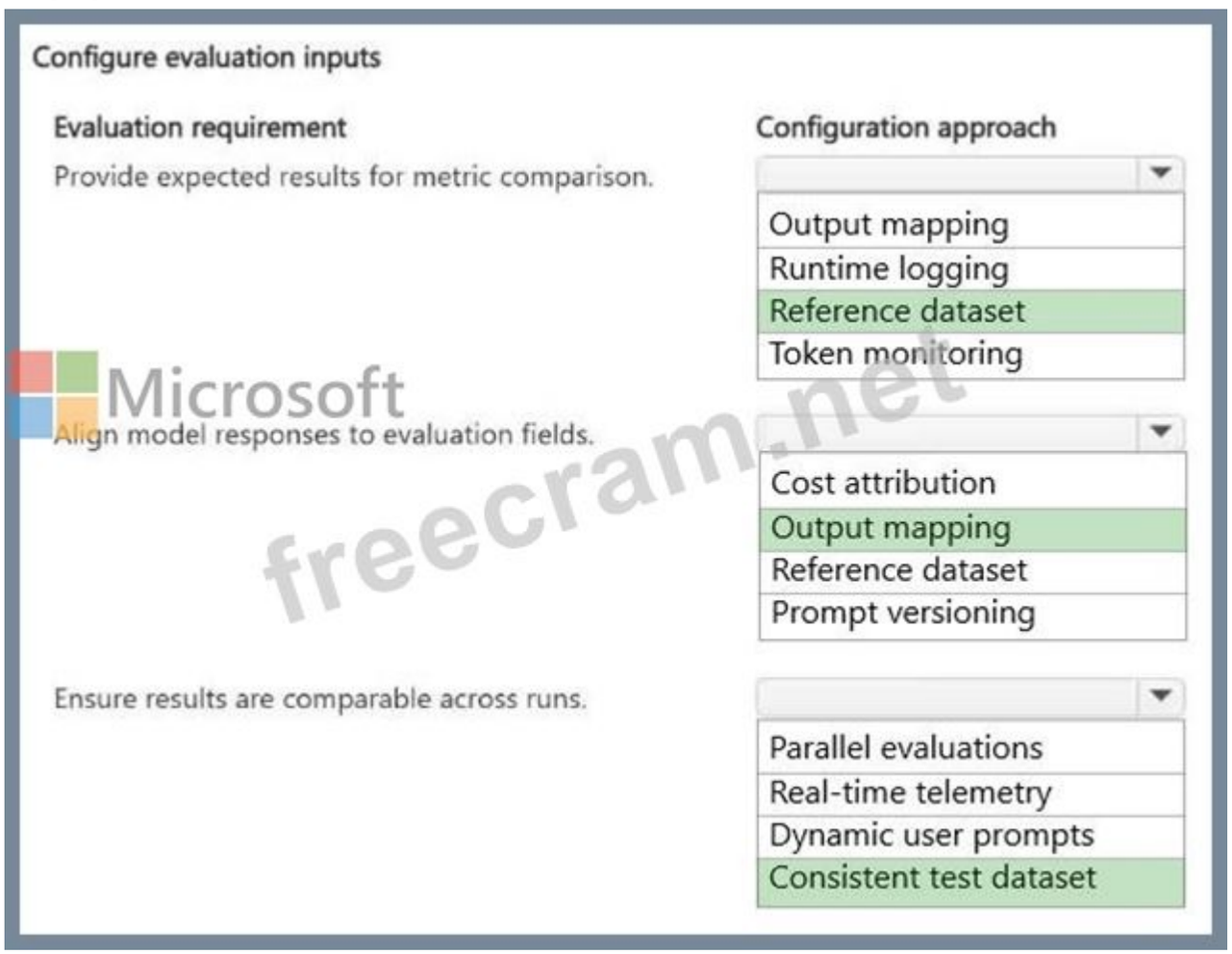
You need to configure evaluation inputs so quality metrics can be reliably calculated across test runs.

How should you prepare the evaluation inputs? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:



Explanation:

Box 1: Reference dataset

Provide expected results for metric comparison.

A reference dataset contains predefined input prompts matched with ground-truth "expected results" (or golden outputs). To calculate quality metrics reliably and compare them across builds and environments, the evaluation framework needs this baseline data to score the generated outputs against a consistent benchmark.
Box 2: Output mapping
Align model responses to evaluation fields.
Output mapping is the process of aligning generative model responses to specific evaluation fields. For structured outputs, your evaluation dataset needs to know exactly which part of the generated response corresponds to the ground truth or target metrics. Configuring output mapping ensures that data fields are aligned consistently across different test runs, making the evaluation repeatable and comparable across builds and environments.
Box 3: Consistent test dataset
Ensure results are comparable across runs.
A consistent test dataset is used to reliably compare evaluation metrics across different builds and environments, you must test the application using the exact same inputs each time. A fixed, high-quality test dataset ensures that changes in metrics reflect changes in the model or application logic rather than variations in the test data itself.
Reference:
https://testquality.com/llm-evaluation-metrics-testing-strategies/
https://www.freecodecamp.org/news/how-to-evaluate-and-select-the-right-llm-for-your-genai-application/

NEW QUESTION: 25
Case Study 1 - Fabrikam Inc.
Background
Fabrikam Inc. is a mid-sized healthcare analytics company that provides population health dashboards and predictive insights to regional hospital systems across the United States.
Fabrikam Inc. customers rely on near real time analytics to monitor patient flow, staffing needs, and readmission risks. They use multiple traditional forecasting machine learning models for predictions.
Fabrikam Inc. has an established Microsoft Azure footprint. The company uses Jupyter Notebooks that run on a local server as the primary development environment. The data science team is experiencing scalability, asset management and code management issues with the current development platform.
Fabrikam Inc. plans to migrate to a cloud-based development environment to mitigate the issues.
Additionally, the company plans to implement a Retrieval-Augmented Generation (RAG)-based chat application for client support. Leadership requires the application to be developed and deployed with a low operational risk.
Current Environment
Fabrikam Inc. operates a single Azure subscription that has the following components:
* Azure Data Lake Storage Gen2 that contains de-identified clinical and operational datasets
* Azure AI Search indexing curated analytical documents and reference materials
* A small set of Python-based training scripts maintained by data scientists
* Azure OpenAI Service with deployed foundational models
* A Microsoft Foundry resource for building a RAG-based solution
Evaluation data has manually defined expected responses.
The current challenges faced by the data science team include the following:
* Model training jobs are run manually from notebooks.
* Experiment tracking is inconsistent
* Model versions are registered without standardized metadata.

* Deployment is performed manually by data scientists, with limited rollback capability.
* The team has no standardized evaluation process for generative AI outputs.

The environment currently allows public network access. Authentication relies on user accounts rather than managed identities. Compute targets are manually created and shared across experiments. This has led to resource contention during peak usage.

Business Requirements

Fabrikam Inc. has the following business requirements for the modernization initiative:

* Provide a conversational interface that answers analytics questions by using internal documents and datasets.
* Ensure that sensitive healthcare-related data is not exposed outside the Fabrikam Inc. Azure tenant.
* Enable repeatable and auditable model training and deployment processes.
* Support experimentation to compare prompt strategies and fine-tuned models.
* Align the model with the ranked preferences and optimize behavior for the long term.
* Minimize disruption to existing analytics workloads during rollout.

Technical Requirements

To support the business goals, Fabrikam Inc. identifies these technical requirements:

* Use Azure Machine Learning workspaces to centrally manage data assets, models, and environments.
* Implement experiment tracking and model versioning for all training jobs.
* Orchestrate training and evaluation by using pipelines rather than manually running notebooks.
* Deploy traditional machine learning models with support for staged rollout and rollback.
* Improve RAG-based solution output quality.
* Use the existing evaluation datasets that are based on real data with input-output pairs.
* Apply advanced fine-tuning techniques only when prompt engineering is insufficient Issues and Constraints Fabrikam Inc. must comply with internal security policies that require the company to restrict network access and avoid long-lived secrets. The data science team has limited Azure DevOps experience, so solutions must favor managed services and automation over custom infrastructure.

Cost predictability is important. Leadership prefers serverless or managed compute options where possible but is willing to approve dedicated compute for stable production workloads.

Problem Statement

Fabrikam Inc. must design and implement an Azure-based AI operations solution that enables reliable training, evaluation, deployment, and iteration of generative AI models. The solution must support experimentation and gradual rollout while ensuring governance, security, and operational stability. The data science and platform teams must collaborate to deliver this solution by using Azure Machine Learning and Microsoft Foundry capabilities.

You need to improve a GPT-5 model performance based on Fabrikam Inc.'s technical requirements. Which action should you perform first?

A. Deploy the model to production to gather real-world feedback.

B. Evaluate the model output.

C. Fine-tune the model to improve accuracy.

D. Generate synthetic interaction data.

Answer: (SHOW ANSWER)

The recommended first action is to evaluate the model output.

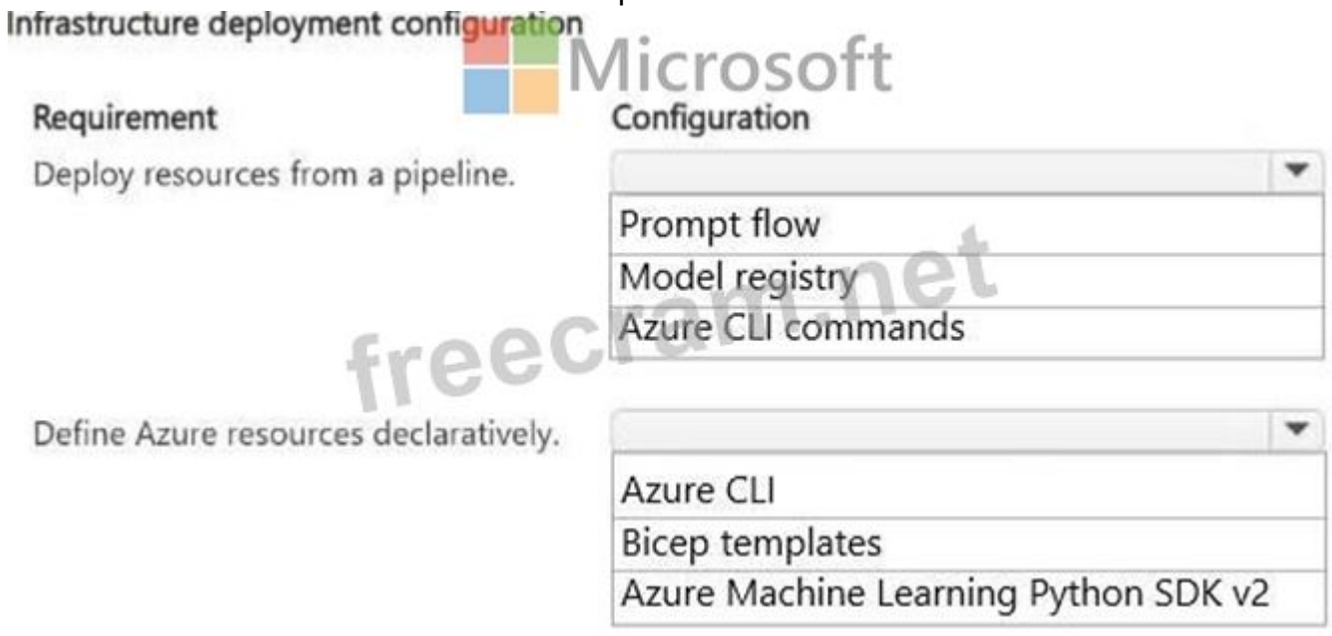
Why This Is the Correct First Step

According to OpenAI's Model Optimization Framework, you must measure your baseline performance before moving to advanced optimization techniques like fine-tuning. You cannot objectively determine whether prompt engineering is "insufficient" without an established evaluation suite to quantify where the model diverges from expectations.

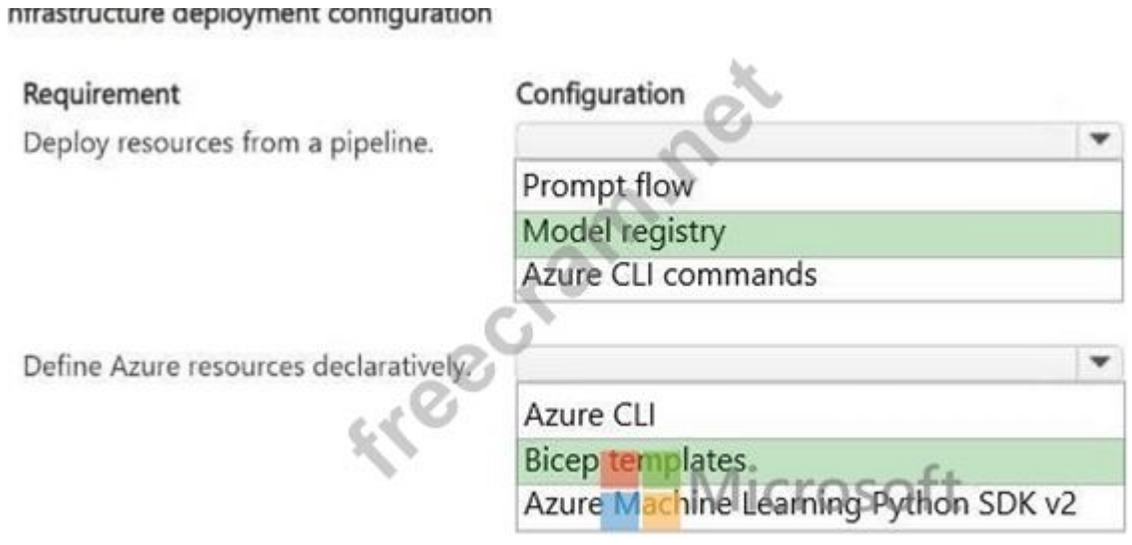
Scenario: To support the business goals, Fabrikam Inc. identifies these technical requirements:

Deploy traditional machine learning models with support for staged rollout and rollback.
Improve RAG-based solution output quality.
Use the existing evaluation datasets that are based on real data with input-output pairs.
-> Apply advanced fine-tuning techniques only when prompt engineering is insufficient Reference:
https://developers.openai.com/api/docs/guides/model-optimization

NEW QUESTION: 26
Hotspot Question
A team is standardizing MLOps practices by using automated deployments.
The team requires infrastructure to be defined declaratively and deployed through automation pipelines.
You need to configure infrastructure deployment.
What should you configure for each requirement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



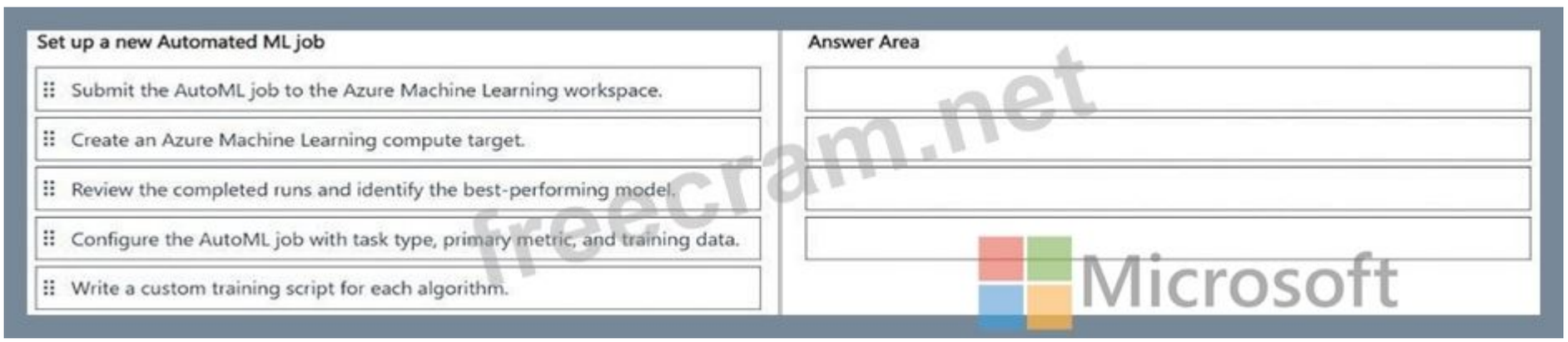
Answer:



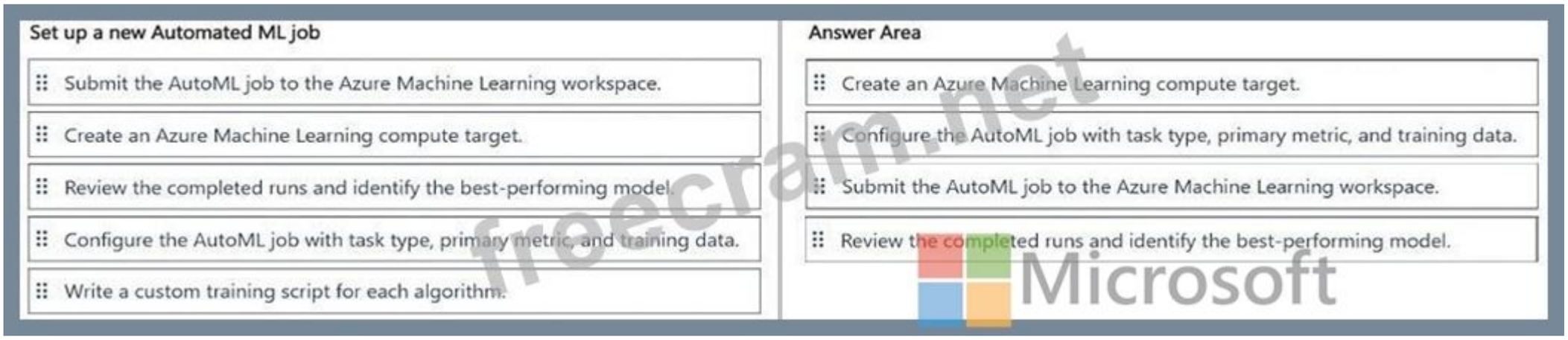
Explanation:
Box 1: Model Registry
Deploy resources from a pipeline.

To standardize MLOps practices with declarative infrastructure and automated pipelines, the best choice is to use a Model Registry via Infrastructure as Code (IaC) for your core setup, supplemented by Azure CLI inside your pipelines for operational tasks. Prompt Flow is an application development tool and should not be used to deploy infrastructure.
A Model Registry (like Azure Machine Learning registry or MLflow) is the standard architectural pattern for managing ML artifacts declaratively.
Declarative Tracking: It stores model versions, lineages, and environments as code-like configurations.
Pipeline Integration: CI/CD pipelines can natively fetch from or push to a registry using configuration files.
Environment Isolation: It allows you to promote the exact same model artifact across Dev, Staging, and Production environments without rebuilding.
Box 2: Bicep templates
Define Azure resources declaratively.
Bicep templates are an excellent choice for defining Azure resources declaratively in an automated MLOps pipeline.
Native Azure Integration: Bicep supports all Azure Machine Learning (Azure ML) resources immediately upon release.
Declarative Syntax: You define the desired end-state of your infrastructure without writing complex deployment scripts.
No State Management: Azure manages the state automatically, unlike Terraform which requires a remote state file.
Tooling Support: Deep integration with Azure Pipelines and GitHub Actions allows easy deployment validation.
Reference:
https://azure.microsoft.com/en-us/blog/new-azure-capabilities-to-simplify-deployment-and-management/

NEW QUESTION: 27
Drag and Drop Question
A data science team plans to evaluate multiple algorithms and hyperparameters for a classification problem without manually authoring separate training scripts.
The team must run an automated machine learning (AutoML) job that compares models and identifies the best-performing configuration.
You need to configure and run an AutoML training job.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



freecram.net
Microsoft

NEW QUESTION: 28

A product team is building a customer support assistant that must respond consistently across multiple channels.

Early testing shows that small wording changes in prompts cause large differences in tone and factual accuracy.

The team needs prompts that are reliable, reusable, and adaptable across multiple use cases without retraining the underlying model.

You need to design prompts that improve response quality while remaining flexible for future changes.

Which two actions should you perform? Each correct answer presents part of the solution.

(Choose two.)

NOTE: Each correct selection is worth one point.

A. Fine-tune the model for each conversational variation.

B. Apply prompt transformations to separate system instructions from user input.

C. Use the system prompt to establish the role, tone, and style.

D. Increase the temperature setting to encourage creativity.

E. Repeat the instructions at the end of the system prompt.

Answer: (SHOW ANSWER)

To achieve reliable, reusable, and adaptable prompts without retraining the model, you should implement Prompt Flow, use System Messages for tone/role definition, and enforce Output Parsing. These techniques standardize how the model behaves across all channels.

Here is the most effective approach to stabilize your customer support assistant:

[B]

1. Implement Prompt Flow

Instead of hardcoding raw prompts in your code, use Azure Machine Learning Prompt Flow.

Visual Workflows: Create executable flows that link LLMs, prompts, and Python tools.

Versioning: Easily track, iterate, and roll back prompt versions without touching application code.

Evaluation: Test prompt variants systematically against a set of baseline queries to objectively measure factual accuracy and tone before pushing to production.

[C]

2. Standardize System Messages

A well-defined system message serves as the "guardrails" for your assistant. It should dictate the persona, scope of knowledge, and safety rules.

Define the Persona: Instruct the model exactly how to behave (e.g., "You are a helpful, empathetic, and strictly technical support assistant for [Company Name].").Strict Guidelines:

Provide instructions on how to handle out-of-scope queries (e.g., "If you do not know the answer, do not guess. Apologize and route the user to [Support Email].").

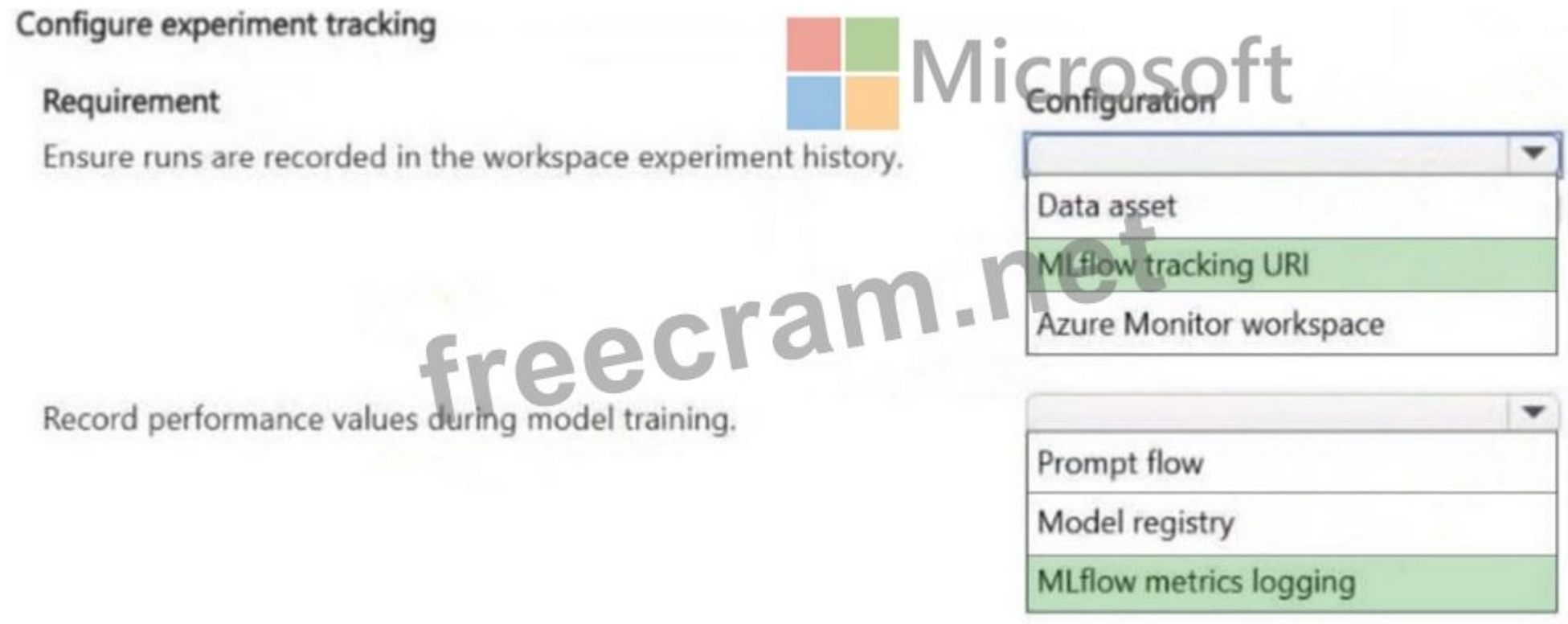
Reference:
https://whitebeardstrategies.com/blog/5-best-practices-for-efficient-language-model-prompting/

NEW QUESTION: 29
A team manages an Azure Machine Learning workspace and deploys a model to an endpoint.
A deployed online endpoint shows inconsistent response times during periods of high traffic.
You need to identify potential performance degradation.
Which three metrics should you monitor? Each correct answer presents part of the solution.
Choose three.
NOTE: Each correct selection is worth one point.
A. Feature count
B. Requests per minute
C. Connections active
D. Dataset size
E. Request latency
Answer: (SHOW ANSWER)
To locate potential performance degradation in an Azure Machine Learning online endpoint during high traffic, you should monitor these three metrics:
Requests per minute: This metric tracks the volume of incoming traffic and helps identify if spikes in load correlate with slower response times.
Connections active: This monitors the total number of concurrent TCP connections from clients, which can indicate if the endpoint is reaching its capacity limits during peak periods.
Request latency: This directly measures the time taken to respond to requests, allowing you to observe exactly when and by how much performance is degrading.
Reference:
https://oneuptime.com/blog/post/2026-02-16-how-to-deploy-a-machine-learning-model-as-a-real-time-endpoint-in-azure-machine-learning/view

NEW QUESTION: 30
Hotspot Question
You use Azure Machine Learning to train models across multiple experiments by using the same workspace.
You must record training runs in a centralized location to compare results from different jobs.
During training, performance values must be captured so they appear in the experiment run history.
You need to configure experiment tracking.
What should you configure for each requirement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



freecram.net

NEW QUESTION: 31

Hotspot Question

A biomedical research company plans to enroll people in an experimental medical treatment trial.

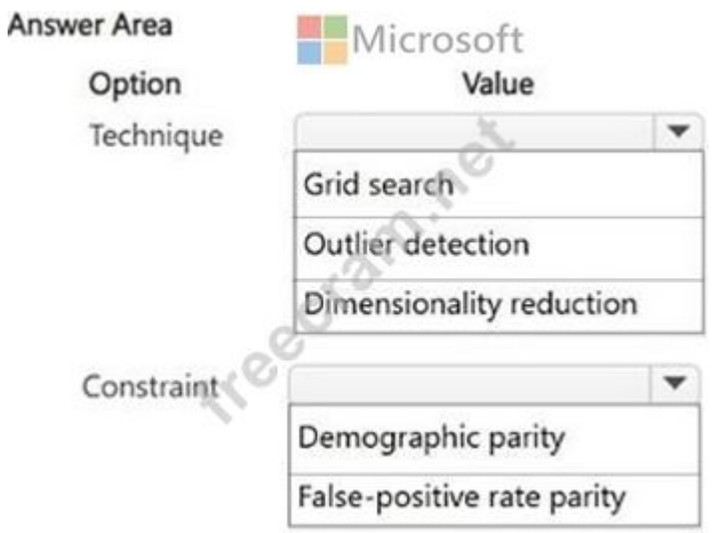
You create and train a binary classification model to support selection and admission of patients to the trial. The model includes the following features: Age, Gender, and Ethnicity.

The model returns different performance metrics for people from different ethnic groups.

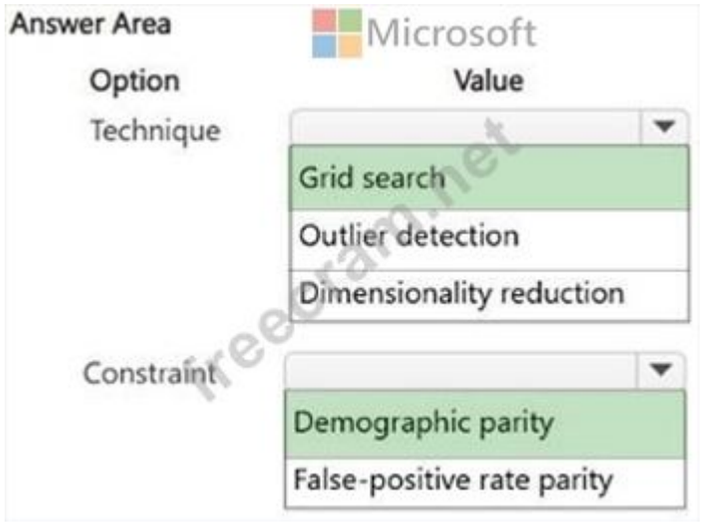
You need to use Fairlearn to mitigate and minimize disparities for each category in the Ethnicity feature.

Which technique and constraint should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:



Valid AI-300 Dumps shared by ExamDiscuss.com for Helping Passing AI-300 Exam! ExamDiscuss.com now offer the **newest AI-300 exam dumps**, the ExamDiscuss.com AI-300 exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com AI-300 dumps with Test Engine here: https://www.examdiscuss.com/Microsoft/exam/AI-300/premium/ (**232** Q&As Dumps, **35%OFF** **Special Discount Code: freecram**)

NEW QUESTION: 32

Case Study 1 - Fabrikam Inc.

Background

Fabrikam Inc. is a mid-sized healthcare analytics company that provides population health dashboards and predictive insights to regional hospital systems across the United States.

Fabrikam Inc. customers rely on near real time analytics to monitor patient flow, staffing needs, and readmission risks. They use multiple traditional forecasting machine learning models for predictions.

Fabrikam Inc. has an established Microsoft Azure footprint. The company uses Jupyter Notebooks that run on a local server as the primary development environment. The data science team is experiencing scalability, asset management and code management issues with the current development platform.

Fabrikam Inc. plans to migrate to a cloud-based development environment to mitigate the issues.

Additionally, the company plans to implement a Retrieval-Augmented Generation (RAG)-based chat application for client support. Leadership requires the application to be developed and deployed with a low operational risk.

Current Environment

Fabrikam Inc. operates a single Azure subscription that has the following components:

* Azure Data Lake Storage Gen2 that contains de-identified clinical and operational datasets
* Azure AI Search indexing curated analytical documents and reference materials
* A small set of Python-based training scripts maintained by data scientists
* Azure OpenAI Service with deployed foundational models
* A Microsoft Foundry resource for building a RAG-based solution

Evaluation data has manually defined expected responses.

The current challenges faced by the data science team include the following:

* Model training jobs are run manually from notebooks.
* Experiment tracking is inconsistent
* Model versions are registered without standardized metadata.
* Deployment is performed manually by data scientists, with limited rollback capability.
* The team has no standardized evaluation process for generative AI outputs.

The environment currently allows public network access. Authentication relies on user accounts rather than managed identities. Compute targets are manually created and shared across experiments. This has led to resource contention during peak usage.

Business Requirements

Fabrikam Inc. has the following business requirements for the modernization initiative:

* Provide a conversational interface that answers analytics questions by using internal documents and datasets.
* Ensure that sensitive healthcare-related data is not exposed outside the Fabrikam Inc. Azure tenant.
* Enable repeatable and auditable model training and deployment processes.
* Support experimentation to compare prompt strategies and fine-tuned models.
* Align the model with the ranked preferences and optimize behavior for the long term.
* Minimize disruption to existing analytics workloads during rollout.

Technical Requirements

To support the business goals, Fabrikam Inc. identifies these technical requirements:

* Use Azure Machine Learning workspaces to centrally manage data assets, models, and environments.
* Implement experiment tracking and model versioning for all training jobs.
* Orchestrate training and evaluation by using pipelines rather than manually running notebooks.
* Deploy traditional machine learning models with support for staged rollout and rollback.
* Improve RAG-based solution output quality.
* Use the existing evaluation datasets that are based on real data with input-output pairs.
* Apply advanced fine-tuning techniques only when prompt engineering is insufficient Issues and Constraints Fabrikam Inc. must comply with internal security policies that require the company to restrict network access and avoid long-lived secrets. The data science team has limited Azure DevOps experience, so solutions must favor managed services and automation over custom infrastructure.

Cost predictability is important. Leadership prefers serverless or managed compute options where possible but is willing to approve dedicated compute for stable production workloads.

Problem Statement

Fabrikam Inc. must design and implement an Azure-based AI operations solution that enables reliable training, evaluation, deployment, and iteration of generative AI models. The solution must support experimentation and gradual rollout while ensuring governance, security, and operational stability. The data science and platform teams must collaborate to deliver this solution by using Azure Machine Learning and Microsoft Foundry capabilities.
You need to refine a GPT-5 model so that its performance and behavior align with the technical and business requirements of Fabrikam Inc.
Which two Foundry strategies should you apply? Each correct answer presents a complete solution. Choose two.
NOTE: Each correct selection is worth one point.
A. Supervised fine-tuning
B. Evaluations
C. Guardrails
D. Synthetic data generation
Answer: (SHOW ANSWER)
Scenario, Technical Requirements:
Improve RAG-based solution output quality.
Apply advanced fine-tuning techniques only when prompt engineering is insufficient Business Requirements:
Provide a conversational interface that answers analytics questions by using internal documents and datasets.
The two Foundry strategies that would be most useful in this scenario are Evaluations and Supervised fine-tuning.
Evaluations: This strategy is critical for improving RAG-based solution output quality. It provides systemic measurement to test how changes to your data, prompts, or retrieval chunks impact the accuracy, relevance, and groundedness of the model's answers before moving to more complex methods.
Supervised fine-tuning: This strategy directly satisfies the requirement to apply advanced fine- tuning techniques only when prompt engineering is insufficient. It allows you to deeply customize the model's behavior and conversational tone to fit your specific analytics interface requirements.
Incorrect:
Guardrails: While valuable for safety and compliance, guardrails focus on enforcing hard operational boundaries (e.g., blocking toxic content or preventing data leaks). They do not optimize RAG output quality or provide advanced behavioral updates when prompt engineering fails.
Synthetic data generation: While helpful for bootstrapping training datasets when real-world data is scarce, it is an data-preparation step rather than a core model refinement strategy designed to iteratively solve prompt engineering limitations or build conversational systems.
Reference:
https://medium.com/@hugoparreao/context-engineering-what-really-improves-the-performance-of-llms-0c0e3ed45c98

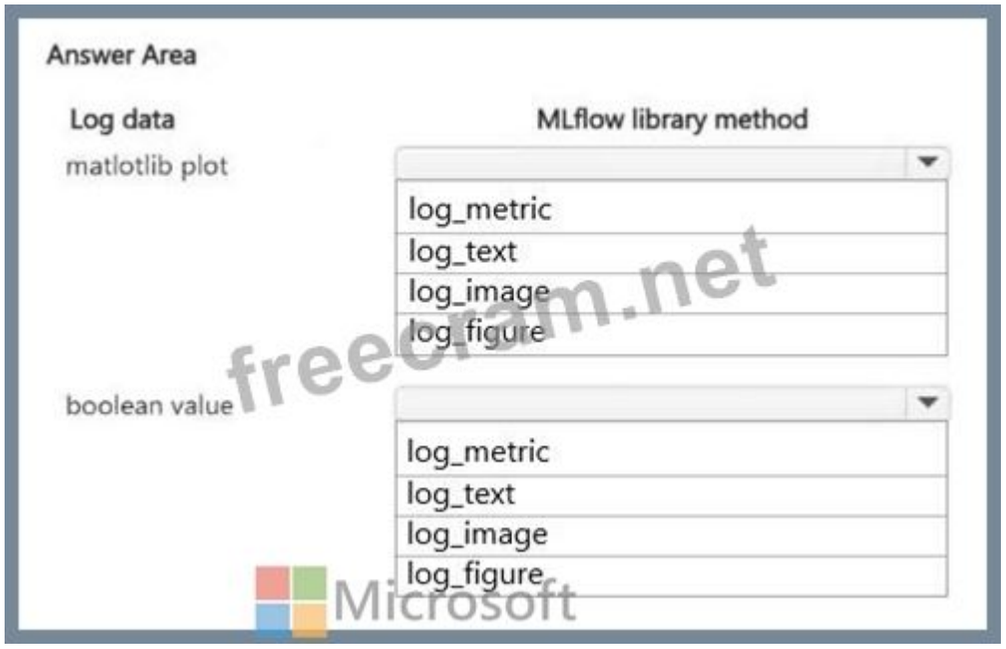
NEW QUESTION: 33
You create a workspace by using Azure Machine Learning Studio.
You must run a Python SDK v2 notebook in the workspace by using Azure Machine Learning Studio.
You need to reset the state of the notebook.
Which three actions should you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.
A. Navigate to another section of the workspace.
B. stop the current kernel.
C. Reset the compute.
D. Change the current kernel.
E. Change the compute.
Answer: (SHOW ANSWER)
To reset the state of a Python SDK v2 notebook in Azure Machine Learning Studio, you can perform any of the following three individual actions:

Stop the current kernel
Reset the compute instance
Change the current kernel
Each of these steps acts as a standalone solution to clear the current runtime memory, variables, and overall execution state of your notebook.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-run-jupyter-notebooks

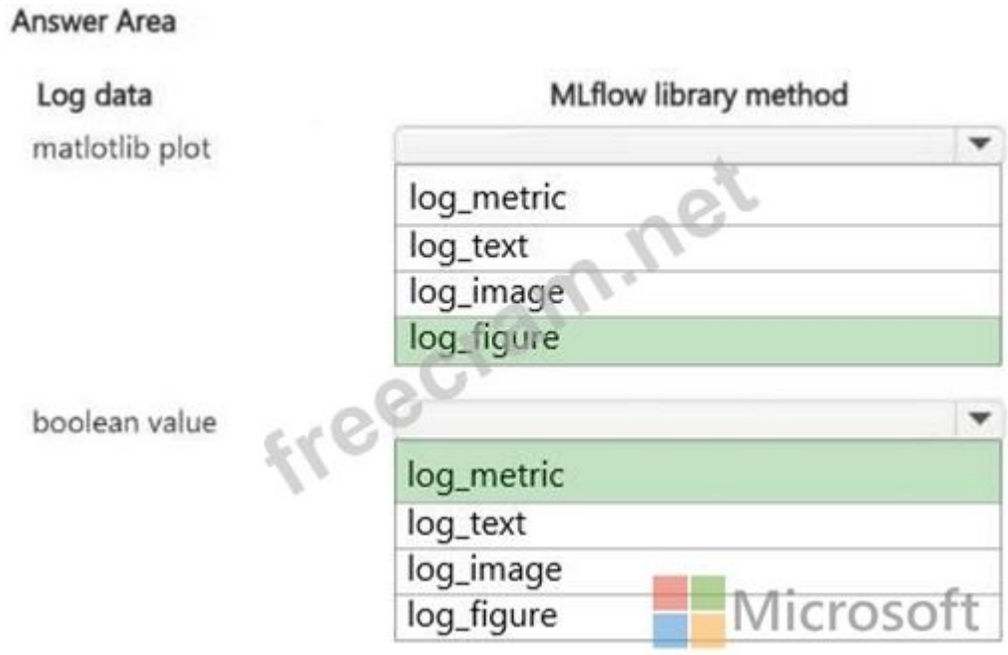
NEW QUESTION: 34
A team is building a Retrieval-Augmented Generation (RAG) system.
The team observes that the retrieved documents are often irrelevant or incomplete.
You need to improve retrieval accuracy.
What should you adjust?
A. Chunk size and overlap
B. Temperature parameter
C. Token limits
D. Embedding strategy
Answer: (SHOW ANSWER)
To best improve retrieval accuracy when documents are irrelevant or incomplete, you should adjust Chunk size and overlap.
Fixes Irrelevant Retrievals: If your chunks are too large, they pack in multiple unrelated topics.
This adds noise, dilutes the semantic vector, and causes the system to pull in massive blocks of text where only a single sentence was actually relevant. Tuning the size ensures chunks stay highly specific.
Fixes Incomplete Retrievals: If your chunks are too small or lack sufficient overlap, vital context gets completely cut off at the boundaries. Adding an optimal overlap (typically 10%-20%) ensures that a sentence split across a boundary preserves its surrounding meaning, keeping the context whole.
Incorrect:
[Not B]
Temperature parameter: This is a generative LLM setting that dictates the creativity or randomness of the output text. It has absolutely zero control over the vector search database or the specific documents picked during the initial retrieval phase.
[Not C]¨
Token limits: This caps how much context you can pass into the prompt window or how much text the model can generate. While a higher limit lets you pass more documents, it does not magically fix the underlying data quality or prevent the retrieval algorithm from pulling in wrong files.
[Not D]
Embedding strategy: While choosing a stronger model or fine-tuning embeddings changes how text is numerically represented, it cannot overcome poorly structured data. Even the best embedding model cannot successfully capture semantic meaning if your data ingestion pipeline is feeding it randomly sliced, incomplete paragraphs Reference:
https://pmc.ncbi.nlm.nih.gov/articles/PMC12649634/

NEW QUESTION: 35
Hotspot Question
You create an Azure Machine Learning workspace and install the MLflow library.
You need to log different types of data by using the MLflow library.
Which method should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:



Explanation:

Box 1: log_figure

matlotlib plot

To log a Matplotlib plot data using the MLflow library, you should use the mlflow.log_figure() method.

Alternatively, if you save the Matplotlib plot as an image object (such as a PIL image) or a local image file first, you can use mlflow.log_image() or mlflow.log_artifact(). However, mlflow.log_figure() is specifically built to accept a Matplotlib figure object directly.

Box 2: log_metric

Boolean value

To log a boolean value data using the MLflow library in an Azure Machine Learning workspace, you should use the mlflow.log_param() method.

Alternatively, if you want to track it as a numerical metric over time, you can map the boolean to an integer (1 for True, 0 for False) and log it using mlflow.log_metric().

Depending on your precise objective for tracking the boolean data, choose from the two main use cases:

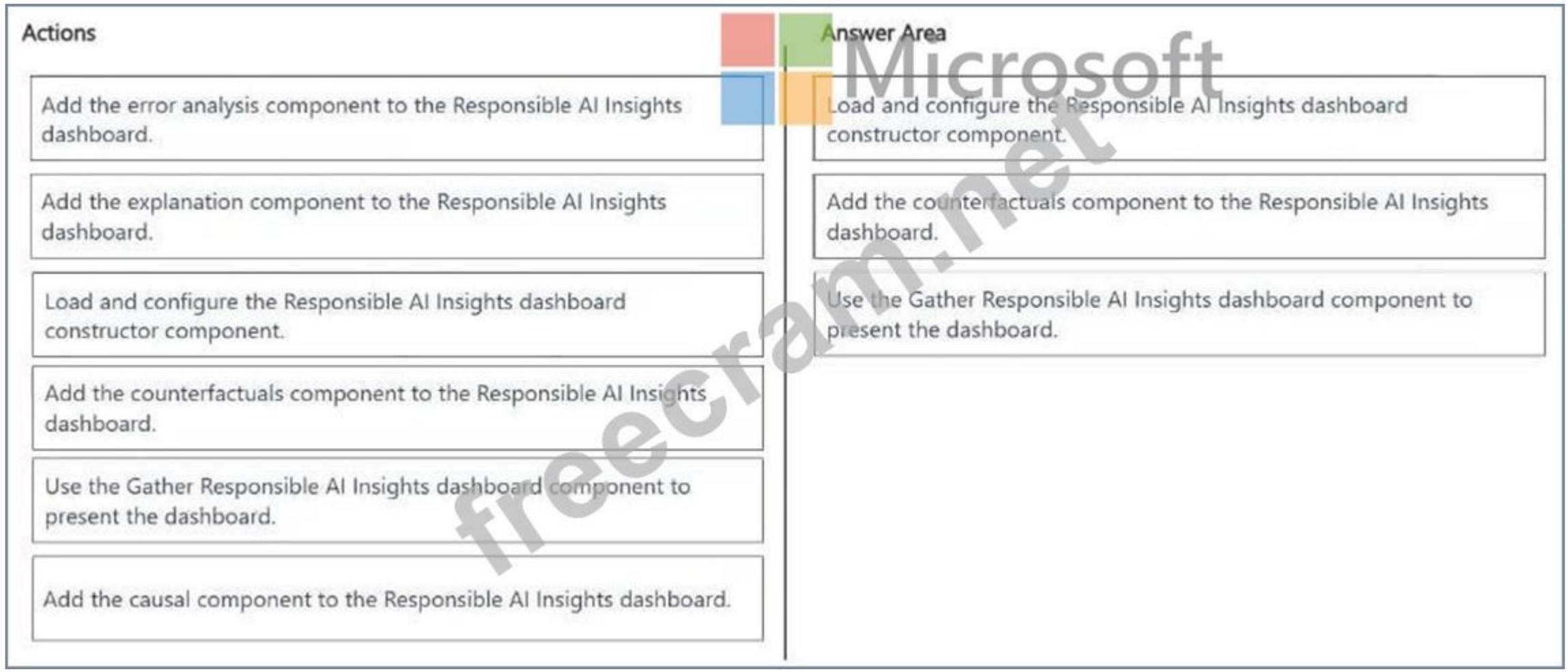
mlflow.log_param(key, value): Best for static configuration data, flags, or hyperparameter switches (e.g., is_training=True). Parameters accept string representations of boolean values.
mlflow.log_metric(key, value): Best if the boolean represents a status or binary occurrence that shifts or updates throughout execution. MLflow metrics exclusively accept numeric values, meaning you must cast the boolean into a 1 or 0.
Reference:
https://docs.azure.cn/en-us/machine-learning/how-to-log-view-metrics

NEW QUESTION: 36
Drag and Drop Question
You manage an Azure Machine Learning workspace. You train a model named model1.
You must identify the features to modify for a differing model prediction result.
You need to configure the Responsible AI (RAI) dashboard for model1.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Microsoft
freecram.net

Answer:



Microsoft
freecram.net

Explanation:

To setup the Responsible AI (RAI) dashboard in Azure Machine Learning and specifically analyze minimal feature modifications needed to change a prediction result, you must use Counterfactual Analysis.

Here are the specific sequential steps you need to take:

Step 1: Load and configure the Responsible AI Insights dashboard constructor component.

Initialize the RAI Insights Dashboard Constructor

1. Create the root pipeline component that acts as the container for your tools.
2. Call the RAI Insights dashboard constructor component.
3. Pass your registered model and your test dataset as the mandatory inputs to this component.

Step 2: Add the Counterfactuals to Responsible AI Insights dashboard

Configure Counterfactual Analysis

1. Add the specific component required to identify what features to modify for a differing prediction.
2. Call the Add Counterfactuals to RAI Insights dashboard component.
3. Link its input to the output of the constructor component initialized in Step 1 above.
4. Configure the parameter variables, including the number of counterfactual examples you want to generate per data point.

Step 3: Use the Gather Responsible AI Insights dashboard component to present the dashboard.

Assemble and Submit the Dashboard Pipeline

1. Gather the components into a cohesive Azure ML pipeline job to execute them.

-> 2. Call the Gather RAI Insights dashboard component to aggregate the constructor and the counterfactual tool outputs.

3. Submit the pipeline job to your Azure Machine Learning workspace compute cluster for processing.

Reference:
https://oneuptime.com/blog/post/2026-02-16-how-to-implement-responsible-ai-dashboards-in-azure-machine-learning/view

NEW QUESTION: 37
You want to ensure ML pipelines produce identical results when executed in different regions or workspaces. Which factor is MOST critical to control for reproducibility?
A. Same compute SKU
B. Same dataset name
C. Versioned datasets and environments
D. Same user identity
Answer: (SHOW ANSWER)
Reproducibility across environments depends on consistent versioning of datasets and environments. Differences in data or dependencies can lead to inconsistent results even if compute resources are identical. Proper asset versioning ensures experiments can be reliably reproduced in any workspace or region.

NEW QUESTION: 38
Hotspot Question
You have an Azure Machine Learning workspace and a collection of image files stored in two Azure Blob Storage accounts.
You need to configure data asset properties.
Which values should you use in your configuration? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

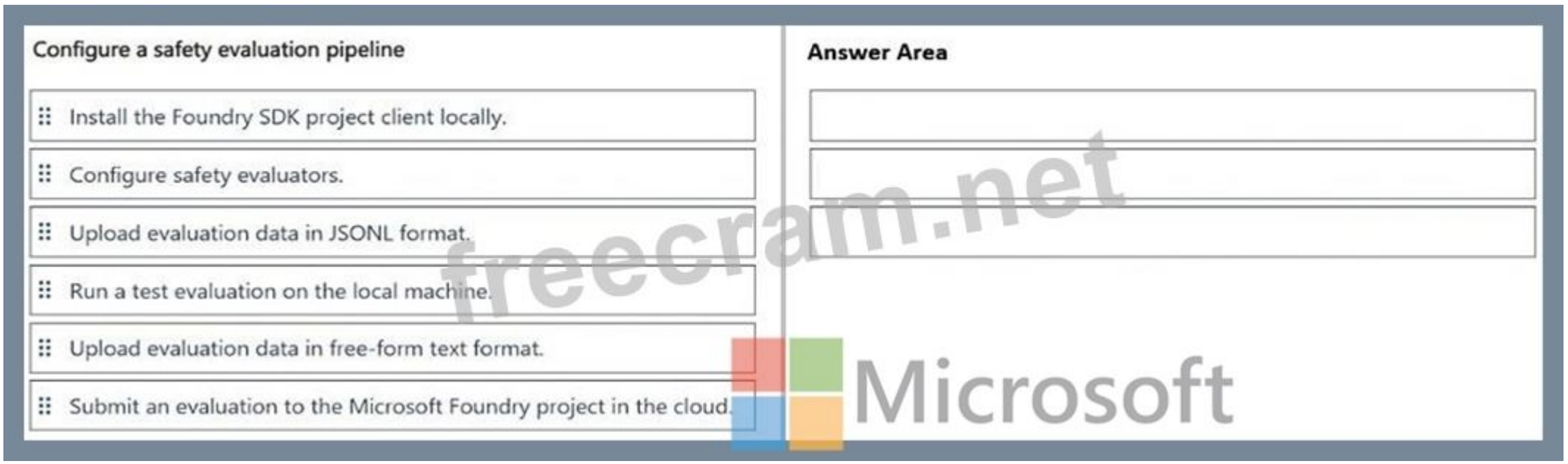


Answer:

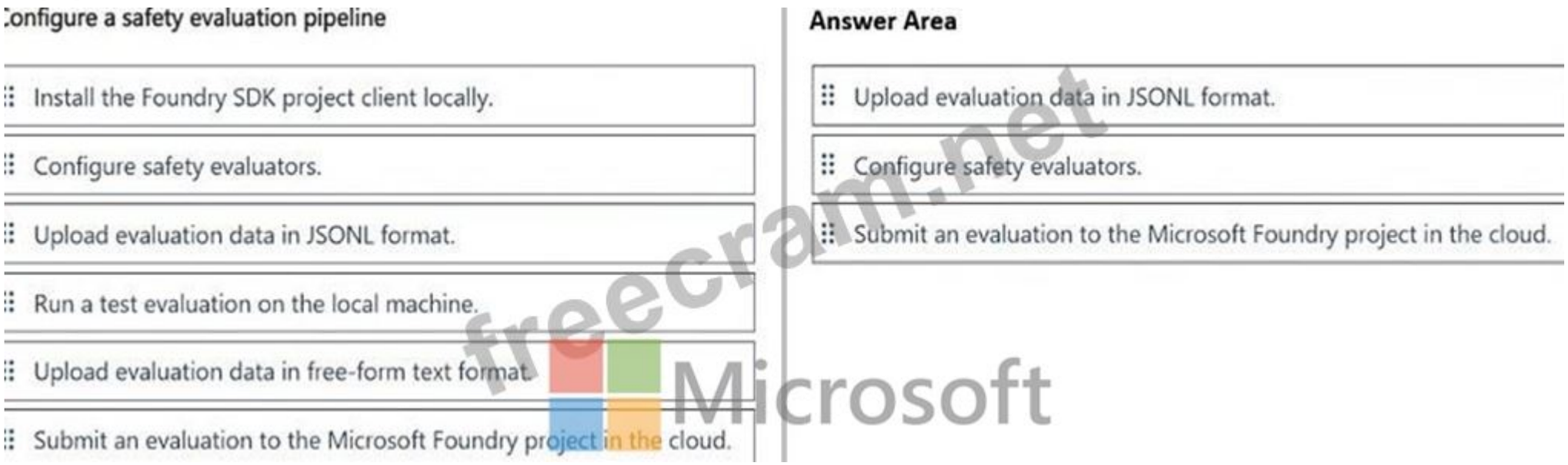


Explanation:
Box 1: uri_folder
The best data asset type for this scenario is a File data asset (uri_file or uri_folder).
In Azure Machine Learning, image files used for training computer vision models are best managed as file-based data assets because machine learning frameworks read individual images directly from storage paths rather than tabular rows.
Preserves Formats: Keeps images in their native formats (PNG, JPEG).Direct Access: Allows training scripts to mount or download files easily.
*-> Folder Mapping: A uri_folder references the entire directory containing your images.
Efficiency: Avoids the overhead of parsing unstructured binary data into a table.
Box 2: azureml
To point a uri_folder data asset to your Azure Blob Storage locations in Azure Machine Learning, you should use the azureml:// URI scheme.
This is the recommended, modern standard that leverages Azure ML datastores for secure tokenless access.
Alternatively, you can use direct Azure Storage URI schemes depending on your configuration Recommended Scheme: Azure ML Datastore This scheme abstracts the storage credentials by referencing an Azure ML Datastore.
Format: azureml://datastores/<datastore_name>/paths/<path_to_folder>/
Example: azureml://datastores/myblobdatastore/paths/images/training_set/ Incorrect:
[not wasbs]
Azure Blob Storage (wasbs)
Used when connecting via the legacy Windows Azure Storage Blob driver.
Format:
wasbs://<container_name>@<storage_account_name>.blob.core.windows.net/<path_to_folder>/
[Not abfss]
Azure Data Lake Storage Gen2 (abfss)
Used if your Blob Storage accounts have the hierarchical namespace enabled.
Format:
abfss://<container_name>@<storage_account_name>.dfs.core.windows.net/<path_to_folder>/ Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-data-assets

NEW QUESTION: 39
Drag and Drop Question
A team validates a generative AI application that produces free-form text responses by using Microsoft Foundry SDK.
The evaluation dataset is registered in the Microsoft Foundry environment.
You need to configure a safety evaluation pipeline that reliably evaluates model outputs for harmful content.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



NEW QUESTION: 40

Drag and Drop Question

A team deploys a machine learning model to a managed online endpoint. The team monitors model performance and data quality metrics in production.

When monitoring thresholds are exceeded, the team requires an automated operational response that notifies downstream systems.

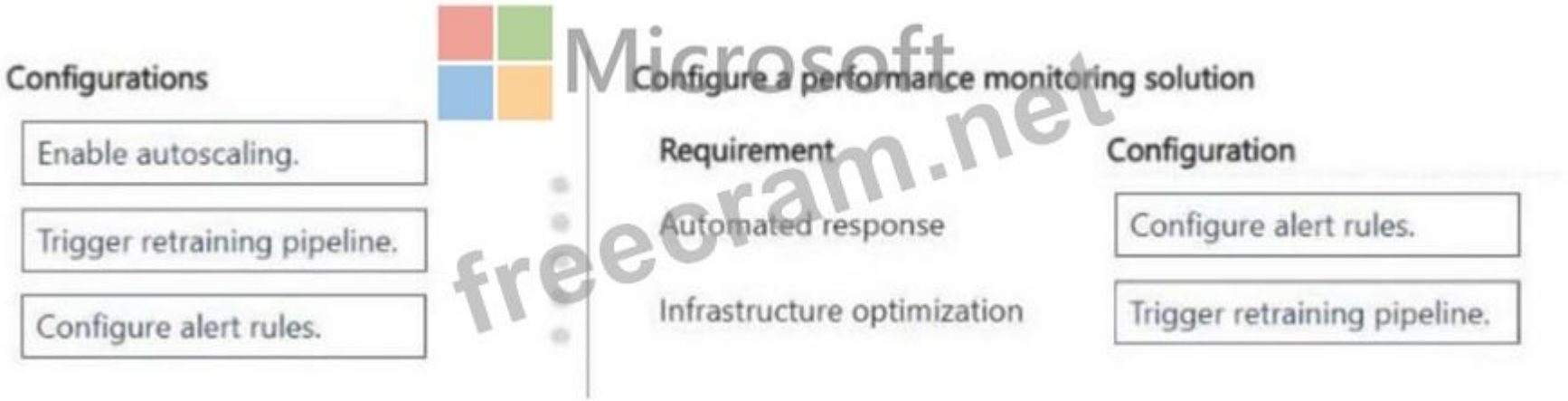
You need to configure the monitoring solution to meet the requirements.

Which configuration should you associate with each requirement as a first step? To answer, move the appropriate configurations to the correct requirements. You may use each configuration once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 41

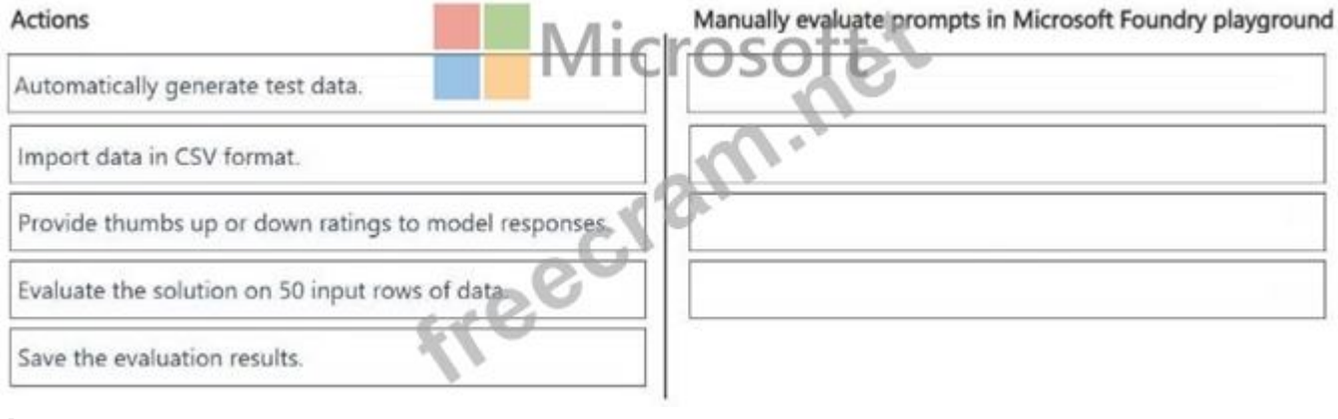
Drag and Drop Question

You manage an Microsoft Foundry project.

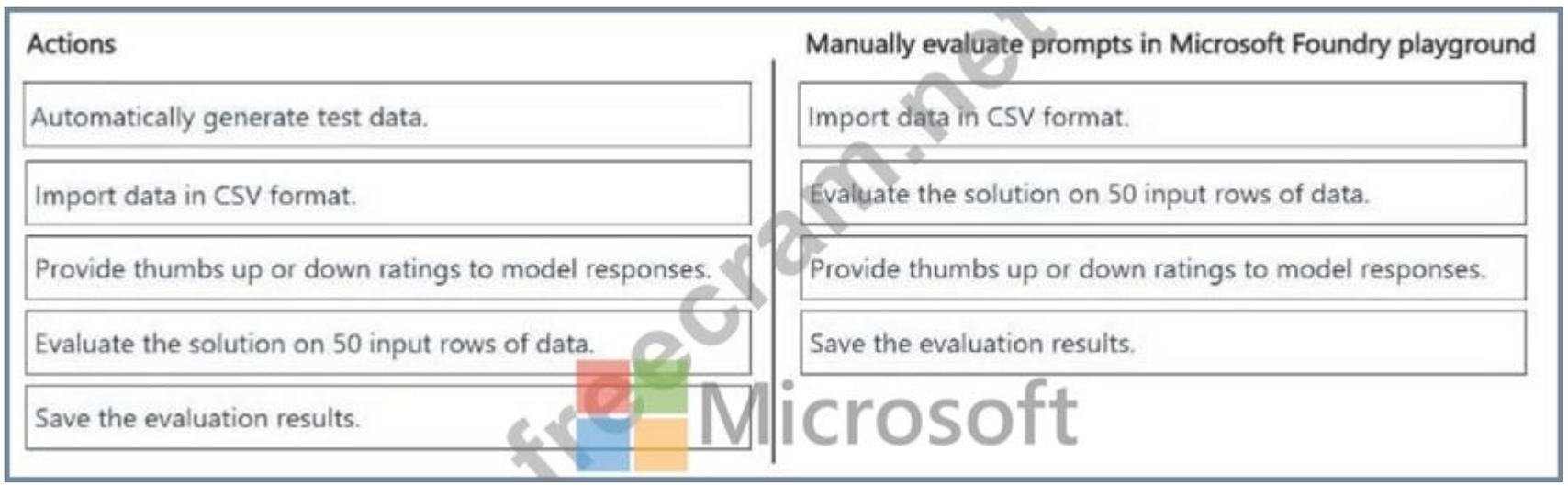
You deploy a large language model from the model catalog.

You need to manually evaluate the model, collect the statistics, and be able to review the results later.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



freecram.net
Microsoft

Explanation:

Step 1: Import Data in CSV Format

Supported formats: Azure AI Studio natively accepts .csv and .jsonl files for evaluation datasets.

Requirement: Your file must contain the input columns (e.g., user prompts) that you want to test against the model.

Step 2: Evaluate the Solution on 50 input Rows

Sample size: 50 rows is an excellent size for a manual, qualitative test baseline.

Execution: You will upload this dataset into the Evaluation blade of your project and map your data fields to the model's inputs.

Step 3: Provide thumbs up or down ratings to model responses

Manual UI: The platform features a manual review interface (often called human-in-the-loop evaluation).

Feedback: You can view the model's generated response for each of the 50 rows side-by-side with the input and log your binary (thumbs up/down) or detailed feedback.

Step 4: Save the evaluation results

Persistence: Once completed, the session is saved to your project's run history.

Review: You can return to the dashboard later to view aggregate statistics, check pass/fail rates, and export the annotated data for your records.

Reference:

https://oneuptime.com/blog/post/2026-02-16-how-to-perform-azure-migrate-at-scale-using-csv-import-for-large-datacenter-inventories/view

NEW QUESTION: 42

Hotspot Question

A team is building a generative AI agent by using Retrieval-Augmented Generation (RAG) in Microsoft Foundry.

The team frequently updates prompt content. The team must be able to track changes across contributors while avoiding full application redeployments.

You need to enable rapid prompt iteration with traceability. Applications consuming the agent must be able to use updated prompts without requiring redeployment.

What should you configure for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Prompt management configuration

Requirement	Configuration
Track prompt changes across teams.	Git repository / Azure Blob Storage / Azure Machine Learning Datastores
Apply prompt updates without redeployment.	CI/CD pipeline / Environment variables / Agent's latest published version reference

freecram.net
Microsoft

Answer:

Prompt management configuration

Requirement	Configuration
Track prompt changes across teams.	**Git repository** / Azure Blob Storage / Azure Machine Learning Datastores
Apply prompt updates without redeployment.	**CI/CD pipeline** / Environment variables / Agent's latest published version reference

freecram.net
Microsoft

NEW QUESTION: 43

Hotspot Question

You manage a Retrieval-Augmented Generation (RAG) system that retrieves internal policy documents from a vector index.

Recent analysis shows that:

- Retrieved results frequently include duplicated content from the same document.
- Retrieved chunks sometimes span unrelated policy sections.

You review the following retrieval and ingestion configurations:

Configuration 1

Parameter	Setting
Chunk size	1,200
Chunk overlap	400
Similarity threshold	0.80
Retrieval strategy	Vector
Top k	8

Configuration 2

Parameter	Setting
Chunk size	300
Chunk overlap	40
Similarity threshold	0.90
Retrieval strategy	Vector
Top k	8

You need to reduce duplicated retrieval results and improve chunk relevance across policy sections. For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

RAG retrieval and chunking configurations

Statement	Yes	No
Configuration 1 produces duplicated content across retrieved chunks.	○	○
Configuration 2 enforces stricter similarity filtering that can exclude relevant chunks.	○	○
Lowering the similarity threshold is required to reduce semantic dilution.	○	○
Switching to keyword-only search will eliminate semantic dilution while preserving semantic relevance.	○	○

Answer:

freecram.net
Microsoft

RAG retrieval and chunking configurations

Statement	Yes	No
Configuration 1 produces duplicated content across retrieved chunks.	○	○
Configuration 2 enforces stricter similarity filtering that can exclude relevant chunks.	○	○
Lowering the similarity threshold is required to reduce semantic dilution.	○	○
Switching to keyword-only search will eliminate semantic dilution while preserving semantic relevance.	○	○

freecram.net Microsoft

NEW QUESTION: 44

Hotspot Question

You manage a Retrieval-Augmented Generation (RAG) system that uses Azure AI Search to retrieve documents from an indexed knowledge base.

The system must support the following retrieval requirements:

- Queries that include exact policy identifiers must return matching documents even when semantic similarity is low.
- Natural-language questions must prioritize semantically relevant documents even when keywords are not an exact match.

You need to configure the retrieval approach to meet the requirements.

How should you configure the retrieval behavior for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hybrid retrieval configuration

Retrieval requirement	Configuration
Exact identifier matching	Semantic ranking only Keyword-based retrieval disabled Vector similarity without keyword scoring Keyword-based retrieval with lexical scoring
Natural-language relevance	Exact-match filtering only Keyword-only search without embeddings Keyword-based retrieval with lexical scoring Semantic ranking over vector search results

freecram.net Microsoft

Answer:

Hybrid retrieval configuration

Retrieval requirement	Configuration
Exact identifier matching	Keyword-based retrieval with lexical scoring
Natural-language relevance	Semantic ranking over vector search results

freecram.net Microsoft

NEW QUESTION: 45

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear on the review screen.

You manage a Retrieval-Augmented Generation (RAG) application built on Microsoft Foundry.

The application retrieves documents from an indexed knowledge base and generates answers for internal users.

Recent feedback indicates that answers are fluent but sometimes include information that is not supported by the documents that were retrieved.

You need to evaluate whether a proposed change improves RAG answer quality by using supported and measurable techniques.

Solution: Review user feedback comments collected after deployment to determine whether answers appear more accurate.

Does the solution meet the goal?

A. Yes

B. No

Answer: (SHOW ANSWER)

Correct:

* Review user feedback comments collected after deployment to determine whether answers appear more accurate.

Relying solely on user feedback comments is not the proper immediate action for evaluating this specific issue but it is still the best choice.

Evaluate whether your proposed change improves the Retrieval-Augmented Generation (RAG) system by reducing hallucinations (unsupported information), you need to measure faithfulness and context relevance using automated, quantifiable metrics.

The proper course of action is to implement an LLM-as-a-Judge framework using an open-source evaluation library like Ragas or TruLens.

Recommended Evaluation Plan
Establish a baseline: Run your current RAG pipeline through a test dataset of 50-100 representative user queries.
Capture the outputs: Save the user query, the exact retrieved document snippets, and the generated response for every test.
Apply the change: Deploy your proposed modification (e.g., altered prompt, different temperature, or re-ranking algorithm).Run the evaluation: Pass the test dataset through the updated pipeline to generate a new set of responses.
Compare the metrics: Use the framework to score both sets of data and mathematically verify if the change reduced unsupported claims.
Incorrect:
* Compare embedding vector dimensions used by the retrieval pipeline before and after the change.
Comparing embedding vector dimensions is not a valid or effective method for measuring RAG answer quality. Vector dimensions (e.g., 1536 or 3072) are static architectural properties of your embedding model. They do not reflect factual accuracy, semantic grounding, or the rate of hallucinations in your text generation.
* Measure token throughput and average response latency before and after applying the proposed change.
Measuring token throughput and latency is not the correct action to solve this specific problem.
Reference:
https://www.getmaxim.ai/articles/how-to-evaluate-your-rag-system/

NEW QUESTION: 46
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have an Azure Machine Learning workspace. You connect to a terminal session from the Notebooks page in Azure Machine Learning studio.
You plan to add a new Jupyter kernel that will be accessible from the same terminal session.
You need to perform the task that must be completed before you can add the new kernel.
Solution: Delete the Python 3.6 - AzureML kernel.
Does the solution meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)
Correct:
* Create an environment.
Incorrect:
* Delete the Python 3.6 - AzureML kernel.
* Delete the Python 3.8 - AzureML kernel.
Note:
Before you can add a new Jupyter kernel on an Azure Machine Learning compute instance terminal, you must create a Conda environment.
Required Workflow
To officially provision and expose the new kernel to your Azure Machine Learning studio Notebooks, you need to execute the following full process from your terminal session:
Create the environment: Provision a new isolated environment (e.g., using conda create -n newenv python=3.10).
Activate the environment: Run conda activate newenv.
Install dependencies: Add the required ipykernel package using conda install ipykernel or pip install ipykernel.
Register the kernel: Bind the new environment configuration to the global Jupyter directory by running:

python -m ipykernel install --user --name newenv --display-name "My New Kernel" Reference:
https://docs.azure.cn/en-us/machine-learning/how-to-access-terminal

Valid AI-300 Dumps shared by ExamDiscuss.com for Helping Passing AI-300 Exam! ExamDiscuss.com now offer the **newest AI-300 exam dumps**, the ExamDiscuss.com AI-300 exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com AI-300 dumps with Test Engine here: https://www.examdiscuss.com/Microsoft/exam/AI-300/premium/ (**232** Q&As Dumps, **35%OFF** **Special Discount Code: freecram**)

NEW QUESTION: 47
Case Study 1 - Fabrikam Inc.
Background
Fabrikam Inc. is a mid-sized healthcare analytics company that provides population health dashboards and predictive insights to regional hospital systems across the United States.
Fabrikam Inc. customers rely on near real time analytics to monitor patient flow, staffing needs, and readmission risks. They use multiple traditional forecasting machine learning models for predictions.
Fabrikam Inc. has an established Microsoft Azure footprint. The company uses Jupyter Notebooks that run on a local server as the primary development environment. The data science team is experiencing scalability, asset management and code management issues with the current development platform.
Fabrikam Inc. plans to migrate to a cloud-based development environment to mitigate the issues.
Additionally, the company plans to implement a Retrieval-Augmented Generation (RAG)-based chat application for client support. Leadership requires the application to be developed and deployed with a low operational risk.
Current Environment
Fabrikam Inc. operates a single Azure subscription that has the following components:
* Azure Data Lake Storage Gen2 that contains de-identified clinical and operational datasets
* Azure AI Search indexing curated analytical documents and reference materials
* A small set of Python-based training scripts maintained by data scientists
* Azure OpenAI Service with deployed foundational models
* A Microsoft Foundry resource for building a RAG-based solution
Evaluation data has manually defined expected responses.
The current challenges faced by the data science team include the following:
* Model training jobs are run manually from notebooks.
* Experiment tracking is inconsistent
* Model versions are registered without standardized metadata.
* Deployment is performed manually by data scientists, with limited rollback capability.
* The team has no standardized evaluation process for generative AI outputs.
The environment currently allows public network access. Authentication relies on user accounts rather than managed identities. Compute targets are manually created and shared across experiments. This has led to resource contention during peak usage.
Business Requirements
Fabrikam Inc. has the following business requirements for the modernization initiative:
* Provide a conversational interface that answers analytics questions by using internal documents and datasets.
* Ensure that sensitive healthcare-related data is not exposed outside the Fabrikam Inc. Azure tenant.

* Enable repeatable and auditable model training and deployment processes.
* Support experimentation to compare prompt strategies and fine-tuned models.
* Align the model with the ranked preferences and optimize behavior for the long term.
* Minimize disruption to existing analytics workloads during rollout.

Technical Requirements

To support the business goals, Fabrikam Inc. identifies these technical requirements:

* Use Azure Machine Learning workspaces to centrally manage data assets, models, and environments.
* Implement experiment tracking and model versioning for all training jobs.
* Orchestrate training and evaluation by using pipelines rather than manually running notebooks.
* Deploy traditional machine learning models with support for staged rollout and rollback.
* Improve RAG-based solution output quality.
* Use the existing evaluation datasets that are based on real data with input-output pairs.
* Apply advanced fine-tuning techniques only when prompt engineering is insufficient Issues and Constraints Fabrikam Inc. must comply with internal security policies that require the company to restrict network access and avoid long-lived secrets. The data science team has limited Azure DevOps experience, so solutions must favor managed services and automation over custom infrastructure.

Cost predictability is important. Leadership prefers serverless or managed compute options where possible but is willing to approve dedicated compute for stable production workloads.

Problem Statement

Fabrikam Inc. must design and implement an Azure-based AI operations solution that enables reliable training, evaluation, deployment, and iteration of generative AI models. The solution must support experimentation and gradual rollout while ensuring governance, security, and operational stability. The data science and platform teams must collaborate to deliver this solution by using Azure Machine Learning and Microsoft Foundry capabilities.

You need to recommend an experiment-tracking strategy that ensures consistent experiment results. What should you recommend?

A. Azure Machine Learning job output logs

B. MLflow experiment tracking

C. Application Insights logs

D. Azure Monitor alerts

Answer: (SHOW ANSWER)

Scenario:

The current challenges faced by the data science team include the following: Experiment tracking is inconsistent To support the business goals, Fabrikam Inc. identifies these technical requirements: Implement experiment tracking and model versioning for all training jobs.

In Azure-based AI operations, integrating MLflow with Azure Machine Learning (Azure ML) provides a unified interface to track experiments, version models, and manage the lifecycle of both traditional ML and Generative AI workloads.

Direct Implementation Strategy

To ensure consistent experiment results and comparison of prompt strategies versus fine-tuned models, use the following architectural approach:

Centralized Tracking: Configure the MLflow tracking URI to point to your Azure ML Workspace.

This allows all logs (from local notebooks, remote training jobs, or Prompt Flow) to aggregate in a single "Experiments" dashboard.

Prompt Strategy Comparison: Use Azure ML Prompt Flow to develop and test prompt variants.

Prompt Flow automatically logs metrics (like groundedness and relevance) which can be viewed alongside fine-tuned model metrics in the Azure ML Studio.

Model Versioning: Use the MLflow Model Registry hosted within Azure ML. Each successful training or fine-tuning run should be registered as a new version of a named model, providing a clear lineage from data to deployment.

Reference:

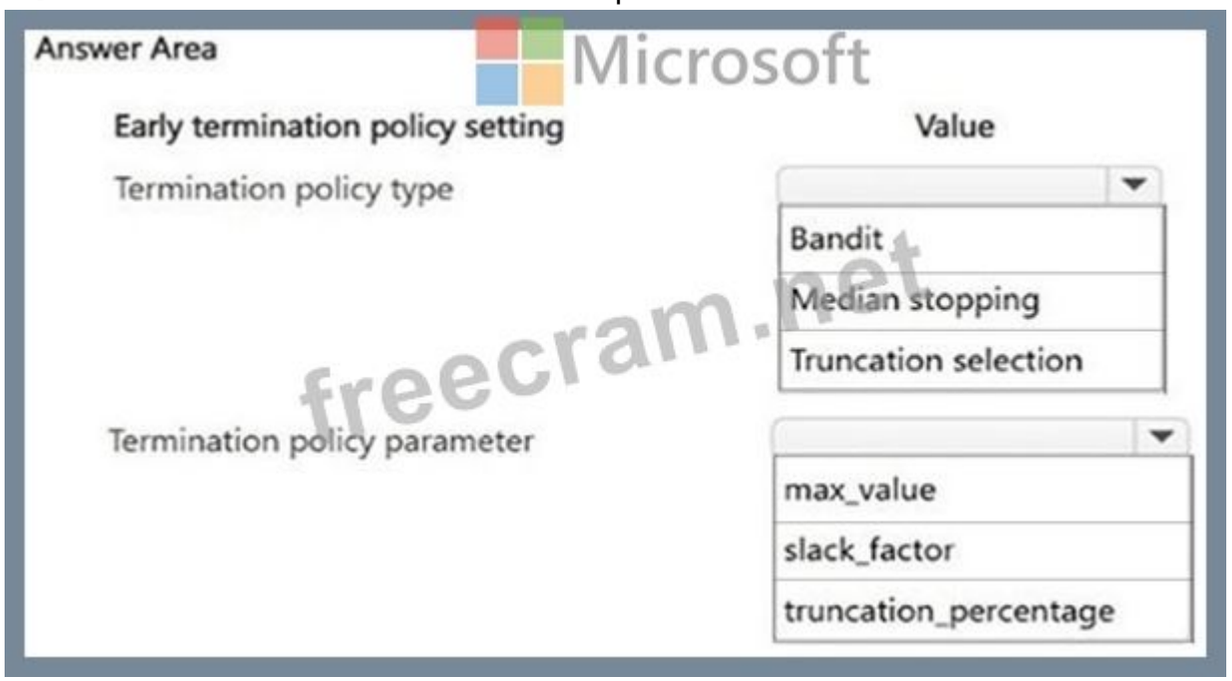
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow-cli-runs

NEW QUESTION: 48
An organization maintains separate Azure Machine Learning workspaces for development and production.
Both environments must use the same validated assets without duplicating them.
Assets must be shared across workspaces while maintaining centralized governance and version control.
You need to enable reuse of assets across workspaces without copying them.
What should you do?
A. Enable workspace-level Git integration and sync assets between repositories.
B. Publish the asset as a pipeline component.
C. Create a shared Azure Machine Learning environment that includes the asset.
D. Publish the asset to an Azure Machine Learning registry.
Answer: (SHOW ANSWER)
To share validated assets between development and production workspaces while maintaining centralized governance and version control, you should use Azure Machine Learning Registries.
Microsoft Learn
Unlike standard workspaces, registries are workspace-agnostic. They act as a central catalog that decouples assets from individual environments, allowing you to "promote" a model or environment from Dev to Prod without manual duplication or data drift.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/concept-machine-learning-registries- mlops

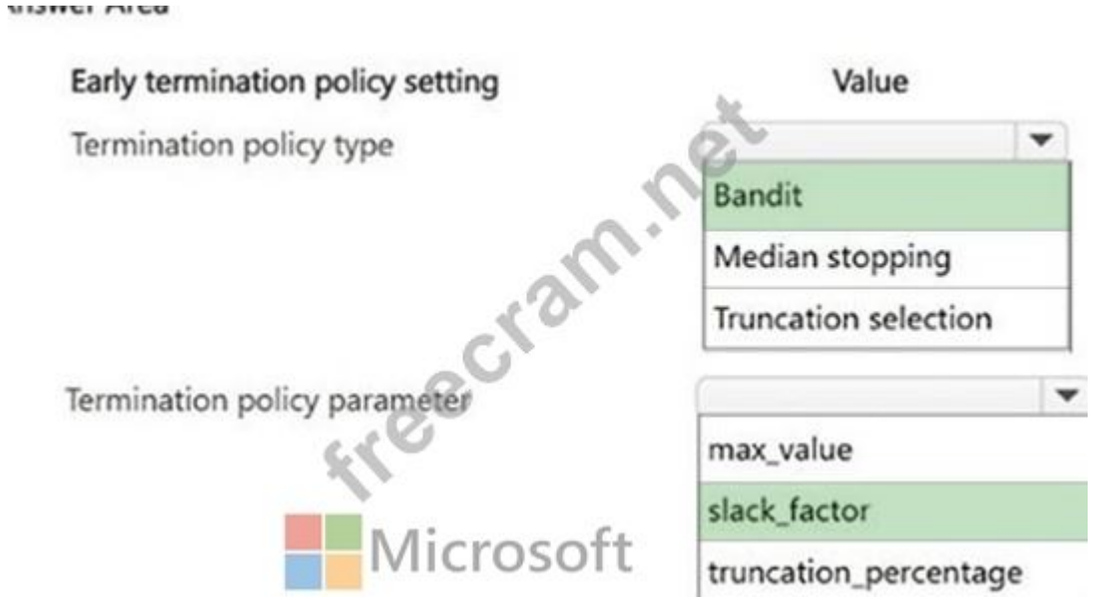
NEW QUESTION: 49
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear on the review screen.
You work in Microsoft Foundry with a prompt flow.
You must manually evaluate prompts and compare results across prompt variants.
You need to capture the inputs, outputs, token usage, and latencies for each flow run for the evaluation.
Solution: Use the prompt flow SDK to enable tracing for the flow before executing runs. Then run the flow to generate traceable results.
Does the solution meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)
Correct:
* In Microsoft Foundry, turn on Tracing for the prompt flow of the project and execute test runs to produce trace data.
Incorrect:
* Create prompt variants and compare their outputs in the Evaluation experience.
* Use the prompt flow SDK to enable tracing for the flow before executing runs. Then run the flow to generate traceable results.
Note:
In Azure AI Foundry, you can capture and compare these metrics by enabling Tracing and using the Bulk Test feature. This allows you to systematically evaluate different prompt variants against a common dataset.

Steps to Evaluate and Compare Prompt Variants
*-> 1. Enable Tracing
Navigate to your Prompt Flow project.
Locate the Tracing toggle at the top of the flow authoring page.
Switch it to On.
This ensures every execution captures latency, token counts, and node-level inputs/outputs.
2. Create Prompt Variants
Within your flow, identify the LLM node you want to test.
Click Variants to create multiple versions of your prompt (e.g., Variant_0, Variant_1).
This allows you to test different instructions or few-shot examples side-by-side.
3. Run a Bulk Test (Evaluation)
4. Analyze the Results
Reference:
https://www.linkedin.com/pulse/streamlining-generative-ai-development-azure-foundry-tracing-taneja-mbwze

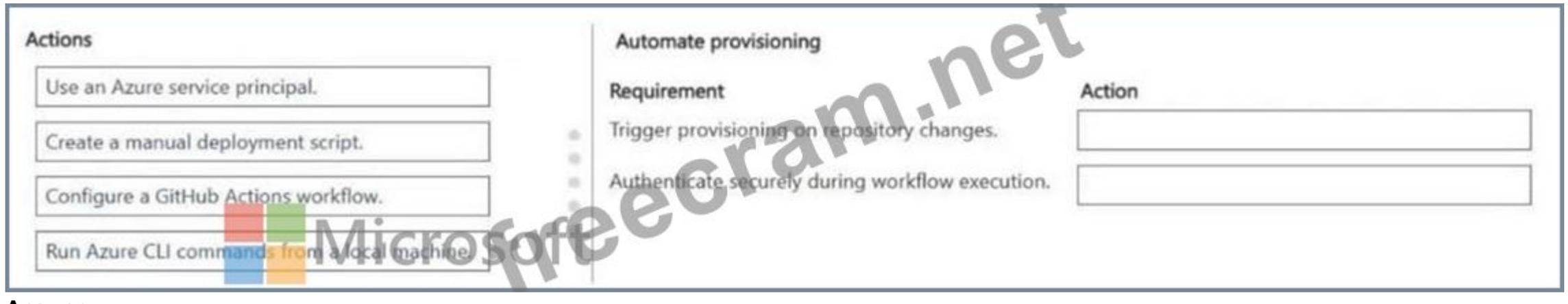
NEW QUESTION: 50
Hotspot Question
You use Azure Machine Learning to implement hyperparameter tuning for an Azure ML Python SDK v2-based model training.
Training runs must terminate when the primary metric is lowered by 25 percent or more compared to the best performing run.
You need to configure an early termination policy to terminate training jobs.
Which values should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 51
Drag and Drop Question
A team maintains Infrastructure as Code (IaC) templates to provision Azure Machine Learning resources.
Provisioning must be triggered by changes in the templates and executed without manual intervention.
You need to automate resource provisioning.
Which action should you take for each requirement? To answer, move the appropriate actions to the correct requirements. You may use each action once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

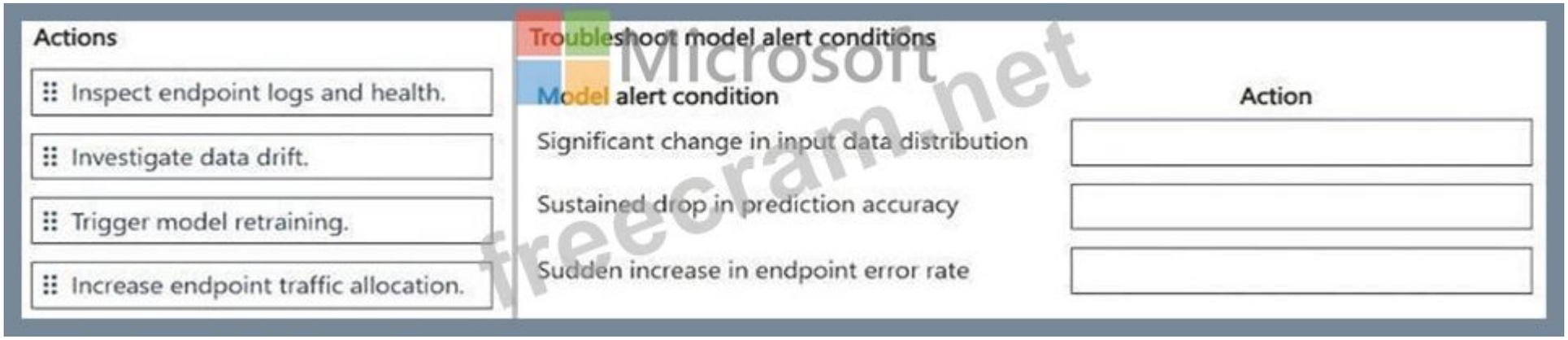


Answer:

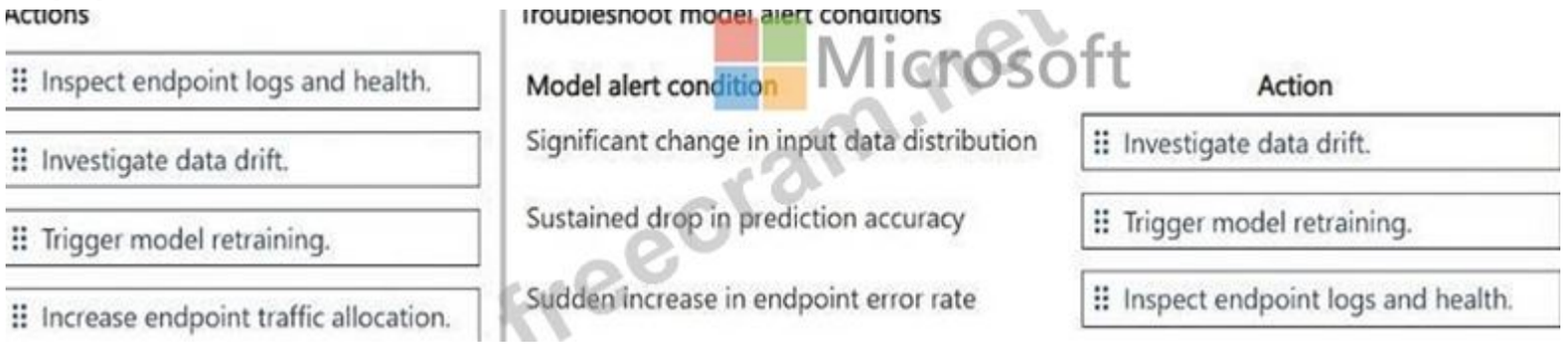


NEW QUESTION: 52
Drag and Drop Question

A team deploys a machine learning model to production and monitors it continuously. Alerts are configured on performance and data quality metrics.
Multiple alerts are triggered during normal operation.
You need to perform the appropriate action for each model alert condition.
Which action should you perform for each alert condition? To answer, move the appropriate actions to the correct model alert conditions. You may use each action once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:



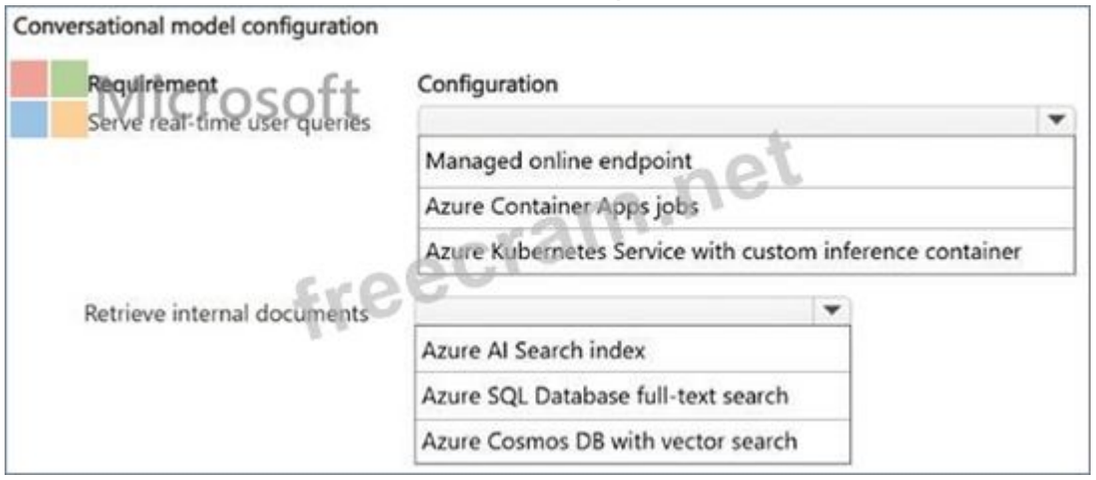
NEW QUESTION: 53
An organization is deploying several generative AI workloads by using Microsoft Foundry. Each workload must meet different requirements related to data governance, task specialization, and operational cost control.
The organization requires models that meet the following requirements:
- Model behavior aligns with the task being performed.
- Data handling aligns with internal governance policies.
- Operational complexity and cost are justified by workload needs.
You need to select the foundation model options that meet the requirements.
Which three models can you select? Each correct answer presents a complete solution. Choose three.
NOTE: Each correct selection is worth one point.
A. A model that is optimized for conversational reasoning when deploying an interactive assistant
B. The largest available model to simplify operational management
C. The smallest available model to minimize the usage cost
D. A model that supports multiple input types when workloads require combined text and image analysis
E. A model that offers enterprise governance controls when workloads process regulated business data
Answer: B,C,E (LEAVE A REPLY)
[B] Operational complexity and cost are justified by workload needs.

In a scenario where a single generative AI workload is deployed via Microsoft Foundry and the associated high operational complexity and costs are already justified, utilizing the largest available model (e.g., GPT-4 class models) is an appropriate strategy to simplify operational management.
This approach aligns with a "production-first" or high-performance mindset where, instead of managing multiple smaller, specialized models, a single powerful model provides the necessary reasoning capabilities, accuracy, and broad applicability for complex, high-stakes tasks, reducing the overhead of engineering, fine-tuning, and maintaining several smaller models.
[C] Model behavior aligns with the task being performed
In Microsoft Azure AI Foundry, the most appropriate model to minimize usage costs while maintaining task alignment is typically a Small Language Model (SLM) from the Phi family or a specialized "mini" partner model.
Top Recommended Small Models for Cost-Efficiency
These models are designed for high efficiency and lower latency, making them ideal for specific, well-defined tasks.
[E] Data handling aligns with internal governance policies
In a Microsoft Foundry (Azure AI Foundry) setup, the Azure OpenAI Service models are the appropriate choice for processing regulated business data due to their built-in enterprise governance and security controls. Unlike standard public models, these "Direct Models" are hosted within your Azure tenant and adhere to strict organizational policies.
Enterprise Governance Controls
For workloads involving regulated data, the following governance features in Microsoft Foundry ensure compliance:
Foundry Control Plane: Centralizes management, observability, and compliance enforcement for all models and agents in one interface.
Azure Policy Integration: Allows administrators to enforce specific model usage, restrict deployments to approved regions, and apply predefined security configurations.
Microsoft Purview Integration: Extends data security posture management (DSPM) to AI workloads, enabling sensitive data discovery, classification, and data loss prevention (DLP) across user prompts and model responses.
Microsoft Entra Agent ID: Automatically assigns a unique identity to every AI agent, enabling granular, role-based access control (RBAC) and auditability of model interactions.
Azure AI Content Safety: Provides configurable filters to block harmful or regulated content and includes "protected material detection" to prevent copyright risks.
Reference:
https://azure.microsoft.com/en-us/products/ai-foundry/models
https://learn.microsoft.com/en-us/azure/foundry/responsible-ai/openai/data-privacy

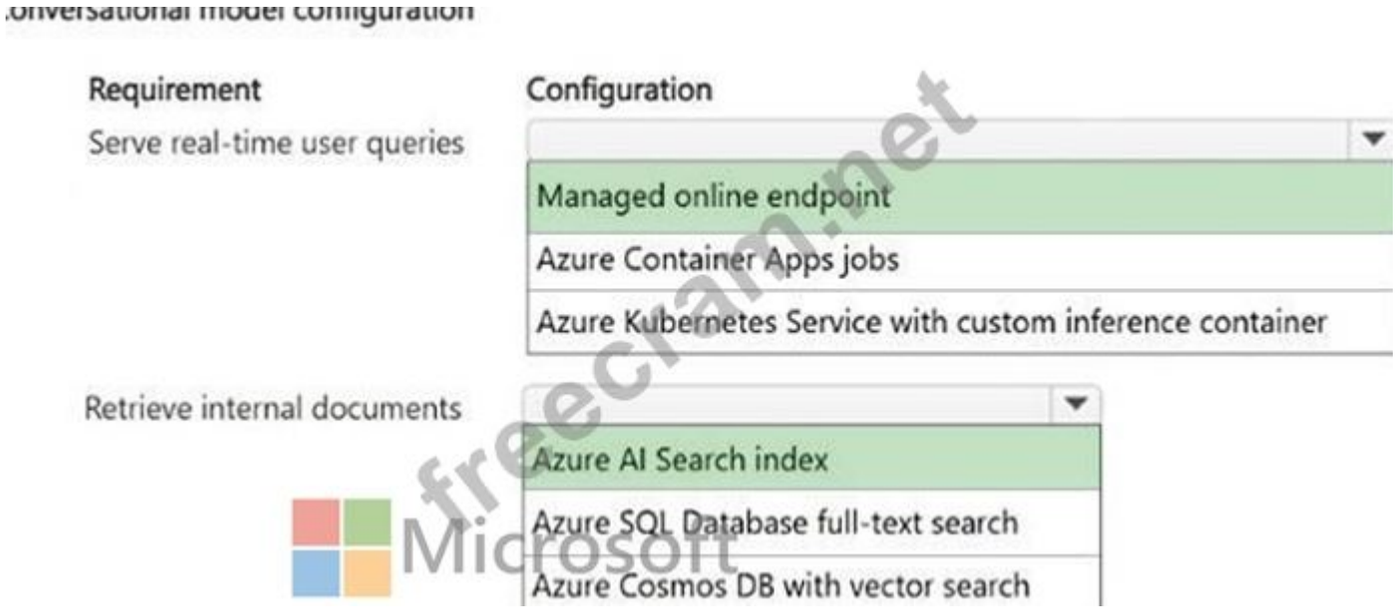
NEW QUESTION: 54
Case Study 1 - Fabrikam Inc.
Background
Fabrikam Inc. is a mid-sized healthcare analytics company that provides population health dashboards and predictive insights to regional hospital systems across the United States.
Fabrikam Inc. customers rely on near real time analytics to monitor patient flow, staffing needs, and readmission risks. They use multiple traditional forecasting machine learning models for predictions.
Fabrikam Inc. has an established Microsoft Azure footprint. The company uses Jupyter Notebooks that run on a local server as the primary development environment. The data science team is experiencing scalability, asset management and code management issues with the current development platform.
Fabrikam Inc. plans to migrate to a cloud-based development environment to mitigate the issues.
Additionally, the company plans to implement a Retrieval-Augmented Generation (RAG)-based chat application for client support. Leadership requires the application to be developed and deployed with a low operational risk.
Current Environment

Fabrikam Inc. operates a single Azure subscription that has the following components:
* Azure Data Lake Storage Gen2 that contains de-identified clinical and operational datasets
* Azure AI Search indexing curated analytical documents and reference materials
* A small set of Python-based training scripts maintained by data scientists
* Azure OpenAI Service with deployed foundational models
* A Microsoft Foundry resource for building a RAG-based solution
Evaluation data has manually defined expected responses.
The current challenges faced by the data science team include the following:
* Model training jobs are run manually from notebooks.
* Experiment tracking is inconsistent
* Model versions are registered without standardized metadata.
* Deployment is performed manually by data scientists, with limited rollback capability.
* The team has no standardized evaluation process for generative AI outputs.
The environment currently allows public network access. Authentication relies on user accounts rather than managed identities. Compute targets are manually created and shared across experiments. This has led to resource contention during peak usage.
Business Requirements
Fabrikam Inc. has the following business requirements for the modernization initiative:
* Provide a conversational interface that answers analytics questions by using internal documents and datasets.
* Ensure that sensitive healthcare-related data is not exposed outside the Fabrikam Inc. Azure tenant.
* Enable repeatable and auditable model training and deployment processes.
* Support experimentation to compare prompt strategies and fine-tuned models.
* Align the model with the ranked preferences and optimize behavior for the long term.
* Minimize disruption to existing analytics workloads during rollout.
Technical Requirements
To support the business goals, Fabrikam Inc. identifies these technical requirements:
* Use Azure Machine Learning workspaces to centrally manage data assets, models, and environments.
* Implement experiment tracking and model versioning for all training jobs.
* Orchestrate training and evaluation by using pipelines rather than manually running notebooks.
* Deploy traditional machine learning models with support for staged rollout and rollback.
* Improve RAG-based solution output quality.
* Use the existing evaluation datasets that are based on real data with input-output pairs.
* Apply advanced fine-tuning techniques only when prompt engineering is insufficient Issues and Constraints Fabrikam Inc. must comply with internal security policies that require the company to restrict network access and avoid long-lived secrets. The data science team has limited Azure DevOps experience, so solutions must favor managed services and automation over custom infrastructure.
Cost predictability is important. Leadership prefers serverless or managed compute options where possible but is willing to approve dedicated compute for stable production workloads.
Problem Statement
Fabrikam Inc. must design and implement an Azure-based AI operations solution that enables reliable training, evaluation, deployment, and iteration of generative AI models. The solution must support experimentation and gradual rollout while ensuring governance, security, and operational stability. The data science and platform teams must collaborate to deliver this solution by using Azure Machine Learning and Microsoft Foundry capabilities.
Hotspot Question

You need to deploy the RAG-based chat application that meets Fabrikam Inc.'s business and technical requirements.
Which configuration should you use for each requirement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 55
Drag and Drop Question
A company is standardizing generative AI development across multiple teams.
Each team requires an isolated workspace. Governance and shared connections must be centrally managed.
You need to implement a Microsoft Foundry environment structure that supports centralized governance and team isolation.
Which type of configuration should you use for each requirement? To answer, move the appropriate configurations to the correct requirements. You may use each configuration once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 56

An Azure Machine Learning workspace contains multiple registered versions of a model that is used in production.

An older model version must no longer be deployable, but it must remain available for compliance review and potential rollback.

You need to change the state of the model version to meet the requirements.

What should you do?

A. Archive the training dataset for the model version.

B. Delete the model version.

C. Archive the model version.

D. Unregister the model version.

Answer: (SHOW ANSWER)

To ensure that older Azure Machine Learning model versions are no longer deployable but remain available for reference or historical tracking, you should archive those specific versions.

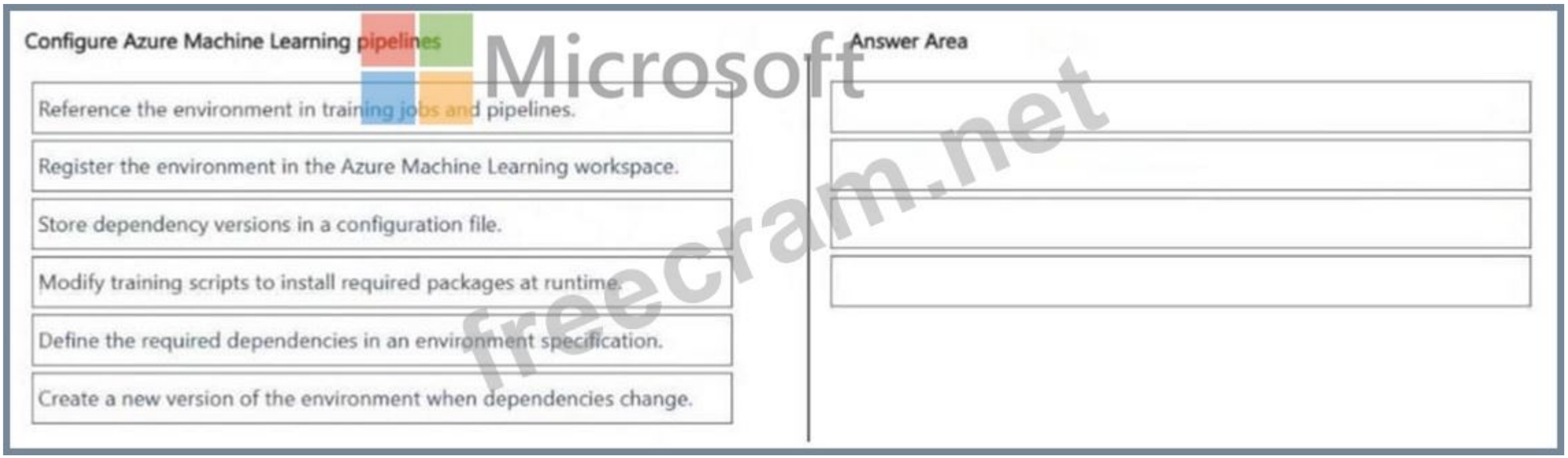
How to Archive Model Versions

Archiving a model version effectively hides it from standard list queries and management interfaces, preventing accidental deployment while maintaining its metadata and files in the workspace.

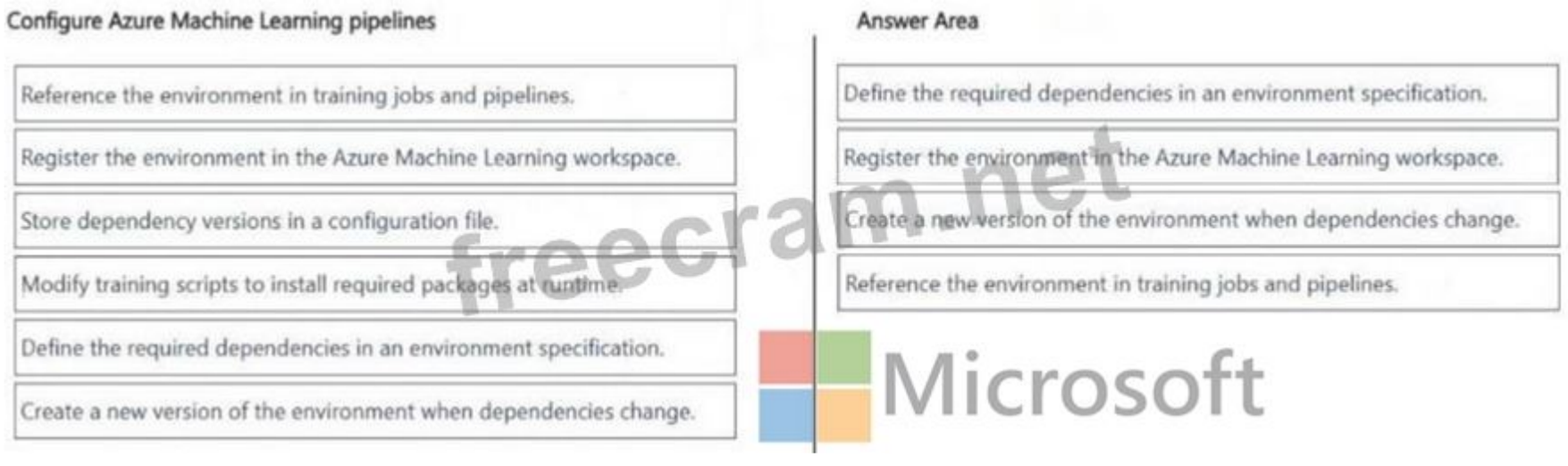
Behavior of Archived Models:

Hidden by Default: They will no longer appear in the Azure Machine Learning Studio model list unless you explicitly filter for archived assets.
Referenceable: You can still reference and use an archived model version in existing workflows if you have its specific version number.
Restoreable: If you need to redeploy the model in the future, you can use the restore command to make it active again.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-manage-models

NEW QUESTION: 57
Drag and Drop Question
A team runs training jobs by using multiple Azure Machine Learning pipelines.
The team must ensure that all runs use the same Python packages and system libraries. The solution must allow dependency updates to be versioned without modifying training code.
You need to configure the workspace so that runtime dependencies are consistent and reusable.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



NEW QUESTION: 58
You manage a Microsoft Foundry project. You build a multi-turn chatbot application.
You plan to filter your traces to identify issues while observing how the application is responding.
The solution must not use an external knowledge base.
You need to select an evaluation metric.
Which built-in evaluator should you use?
A. RelevanceEvaluator
B. SimilarityEvaluator
C. QAEvaluator
D. CoherenceEvaluator
Answer: (SHOW ANSWER)
In a Microsoft Foundry project for a multi-turn chatbot, the best built-in evaluator to use-- especially when you want to avoid an external knowledge base--is the Coherence evaluator or a combination of Agentic Evaluators.
Because your application is multi-turn, you need to observe how the AI maintains a logical flow and resolves user intent over several exchanges. Since you specifically want to avoid an external knowledge base (eliminating "Groundedness" or "Retrieval" metrics), you should focus on quality metrics that only require the conversation history itself.
Reference:
https://www.cekura.ai/blogs/why-single-turn-testing-falls-short-in-evaluating-conversational-ai

NEW QUESTION: 59
Hotspot Question
You are monitoring a fine-tuned large language model deployed in Microsoft Foundry.
You evaluate the model before and after fine-tuning by using the same evaluation dataset.
You review the following evaluation results:

Evaluation metric	Baseline model	Fine-tuned model
Groundedness score	0.82	0.91
Relevance score	0.85	0.88
Hallucination rate	0.06	0.05
Response latency (ms)	420	430

freecram.net Microsoft

You need to determine whether the fine-tuned model shows improved performance without introducing regression. For each of the following statements, select Yes if the statement is true.
Otherwise, select No.
NOTE: Each correct selection is worth one point.

ine-tuned model evaluation metrics

Statement	Yes	No
The fine-tuned model shows improved grounding compared to the baseline model.	○	○
The fine-tuned model introduced a regression in hallucination behavior.	○	○
The fine-tuned model's response latency indicates a functional failure.	○	○
The evaluation results indicate that fine-tuning improved overall response quality.	○	○

Answer:

ine-tuned model evaluation metrics

Statement	Yes	No
The fine-tuned model shows improved grounding compared to the baseline model.	○	○
The fine-tuned model introduced a regression in hallucination behavior.	○	○
The fine-tuned model's response latency indicates a functional failure.	○	○
The evaluation results indicate that fine-tuning improved overall response quality.	○	○

NEW QUESTION: 60

Drag and Drop Question

An organization operates a generative AI application in production by using Microsoft Foundry.

The application serves live user traffic and is updated by a data scientist team regularly as prompts and models evolve.

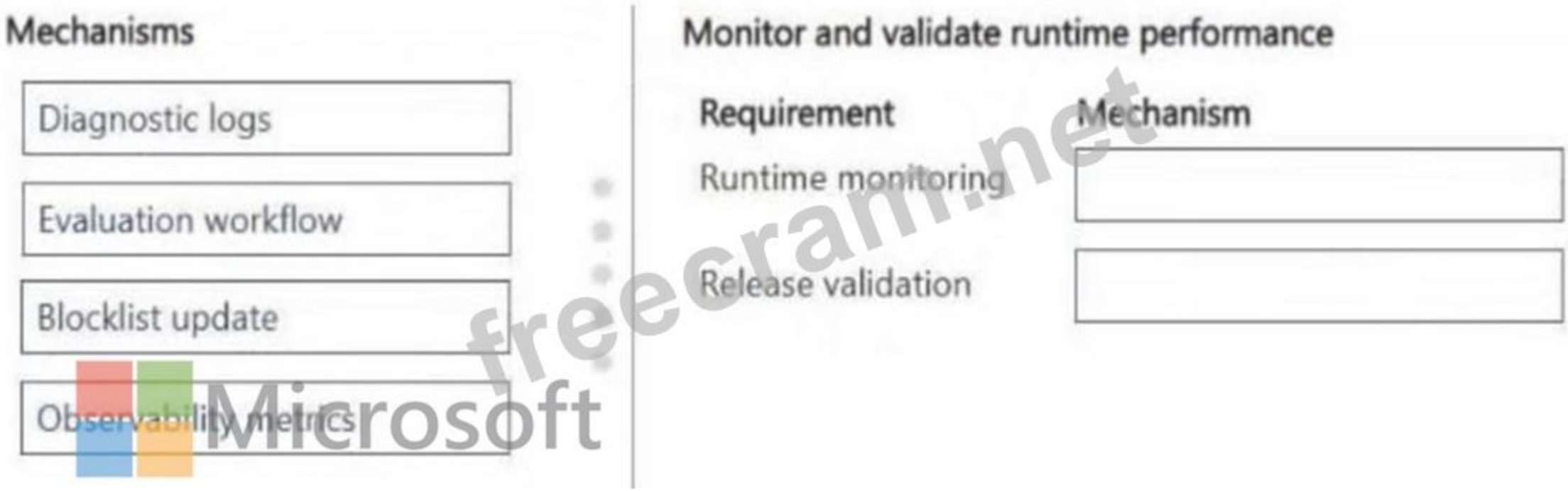
The application intermittently times out during production use, which requires ongoing visibility into runtime behavior.

The team must also validate model quality and safety before releasing new updates to avoid introducing regressions.

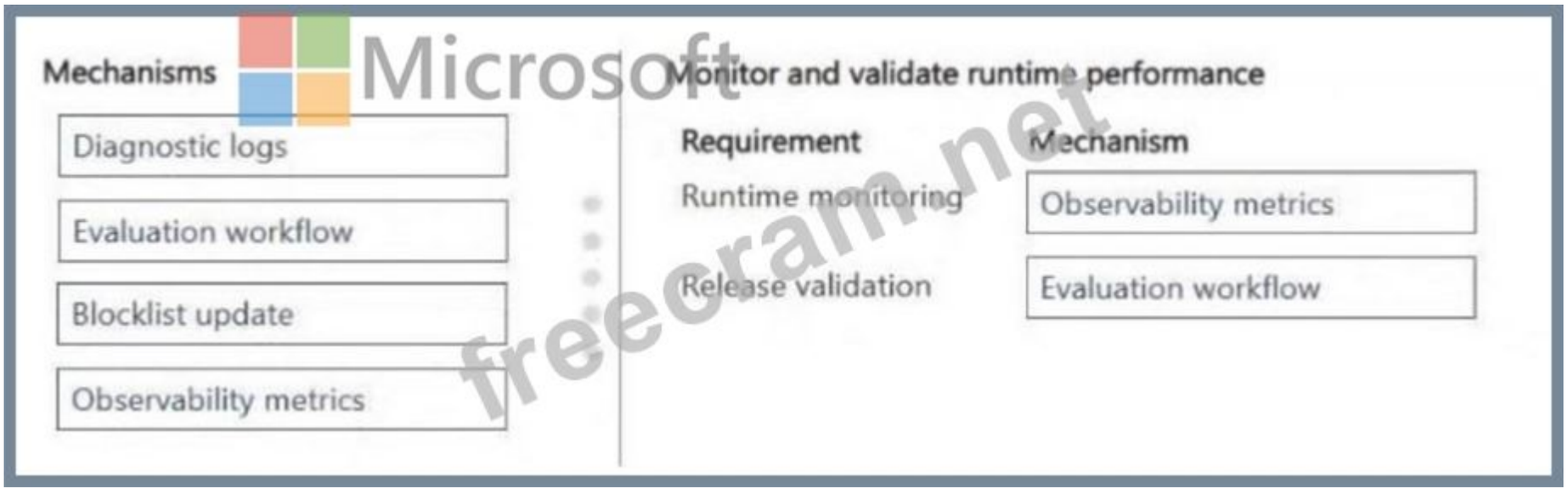
You need to apply the correct mechanisms for continuous runtime monitoring and for release time validation.

Which mechanisms should you use for each requirement? To answer, move the appropriate mechanisms to the correct requirements. You may use each mechanism once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 61

You are designing a new machine learning solution to predict customer churn by using Azure Machine Learning. You have raw data in CSV format stored in Azure Data Lake.

You need to design the solution so that it can efficiently handle large-scale model training and iterative development.

Which two actions should you perform? Each correct answer presents part of the solution.

Choose two.

NOTE: Each correct selection is worth one point.

A. Convert the data into the JSONL format and upload into Blob Storage.

B. Schedule training using an Azure Data Factory pipeline.

C. Configure an Azure Machine Learning compute cluster for model training.

D. Register the data as a tabular dataset in the Azure Machine Learning workspace.

E. Configure an Azure Machine Learning compute instance for model training.

Answer: C,D (LEAVE A REPLY)
Here is the correct and optimized architectural design for the solution.
1. Data Engineering & Registration
Connect Data Lake: Create an Azure ML Datastore pointing to your Azure Data Lake Storage (ADLS) Gen2.
[D] -> Modern Data Asset: Register the CSV data as an MLtable (the modern replacement for Tabular Datasets).
Schema Enforcement: Define types (e.g., CustomerID as string, Churn as boolean) in the MLTable YAML file.
Lazy Loading: MLtable loads data lazily during training to prevent out-of-memory errors on massive datasets.
2. Compute Strategy
[C] -> Compute Cluster (Scale-Out): Use an Azure ML Compute Cluster (AmlCompute) for actual large-scale model training.
Auto-Scaling: Configure the cluster to scale from 0 to N nodes so you only pay for compute during active jobs.Spot VMs: Leverage Azure Spot Virtual Machines on the cluster to reduce training costs by up to 80%.
Incorrect:
[Not E] Compute Instance (Development): Use this only as your workstation for writing code, Jupyter notebooks, and light debugging.
Reference:
https://medium.com/henkel-data-and-analytics/how-to-use-azure-ml-studio-an-eye-opening-model-training-tutorial-for-beginners-from-henkels-data-5035ee10a6d2

Valid AI-300 Dumps shared by ExamDiscuss.com for Helping Passing AI-300 Exam! ExamDiscuss.com now offer the **newest AI-300 exam dumps**, the ExamDiscuss.com AI-300 exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com AI-300 dumps with Test Engine here: https://www.examdiscuss.com/Microsoft/exam/AI-300/premium/ (**232** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

NEW QUESTION: 62
Case Study 1 - Fabrikam Inc.
Background
Fabrikam Inc. is a mid-sized healthcare analytics company that provides population health dashboards and predictive insights to regional hospital systems across the United States.
Fabrikam Inc. customers rely on near real time analytics to monitor patient flow, staffing needs, and readmission risks. They use multiple traditional forecasting machine learning models for predictions.
Fabrikam Inc. has an established Microsoft Azure footprint. The company uses Jupyter Notebooks that run on a local server as the primary development environment. The data science team is experiencing scalability, asset management and code management issues with the current development platform.
Fabrikam Inc. plans to migrate to a cloud-based development environment to mitigate the issues.
Additionally, the company plans to implement a Retrieval-Augmented Generation (RAG)-based chat application for client support. Leadership requires the application to be developed and deployed with a low operational risk.
Current Environment
Fabrikam Inc. operates a single Azure subscription that has the following components:
* Azure Data Lake Storage Gen2 that contains de-identified clinical and operational datasets
* Azure AI Search indexing curated analytical documents and reference materials
* A small set of Python-based training scripts maintained by data scientists
* Azure OpenAI Service with deployed foundational models

* A Microsoft Foundry resource for building a RAG-based solution

Evaluation data has manually defined expected responses.

The current challenges faced by the data science team include the following:

* Model training jobs are run manually from notebooks.
* Experiment tracking is inconsistent
* Model versions are registered without standardized metadata.
* Deployment is performed manually by data scientists, with limited rollback capability.
* The team has no standardized evaluation process for generative AI outputs.

The environment currently allows public network access. Authentication relies on user accounts rather than managed identities. Compute targets are manually created and shared across experiments. This has led to resource contention during peak usage.

Business Requirements

Fabrikam Inc. has the following business requirements for the modernization initiative:

* Provide a conversational interface that answers analytics questions by using internal documents and datasets.
* Ensure that sensitive healthcare-related data is not exposed outside the Fabrikam Inc. Azure tenant.
* Enable repeatable and auditable model training and deployment processes.
* Support experimentation to compare prompt strategies and fine-tuned models.
* Align the model with the ranked preferences and optimize behavior for the long term.
* Minimize disruption to existing analytics workloads during rollout.

Technical Requirements

To support the business goals, Fabrikam Inc. identifies these technical requirements:

* Use Azure Machine Learning workspaces to centrally manage data assets, models, and environments.
* Implement experiment tracking and model versioning for all training jobs.
* Orchestrate training and evaluation by using pipelines rather than manually running notebooks.
* Deploy traditional machine learning models with support for staged rollout and rollback.
* Improve RAG-based solution output quality.
* Use the existing evaluation datasets that are based on real data with input-output pairs.
* Apply advanced fine-tuning techniques only when prompt engineering is insufficient Issues and Constraints Fabrikam Inc. must comply with internal security policies that require the company to restrict network access and avoid long-lived secrets. The data science team has limited Azure DevOps experience, so solutions must favor managed services and automation over custom infrastructure.

Cost predictability is important. Leadership prefers serverless or managed compute options where possible but is willing to approve dedicated compute for stable production workloads.

Problem Statement

Fabrikam Inc. must design and implement an Azure-based AI operations solution that enables reliable training, evaluation, deployment, and iteration of generative AI models. The solution must support experimentation and gradual rollout while ensuring governance, security, and operational stability. The data science and platform teams must collaborate to deliver this solution by using Azure Machine Learning and Microsoft Foundry capabilities.

You need to recommend a solution to address Fabrikam Inc.'s limited rollback capability. Which deployment approach should you recommend?

A. VM-hosted REST APIs

B. Azure Kubernetes Service with blue-green switching

C. Managed online endpoints with traffic splitting

D. Batch endpoints

Answer: (SHOW ANSWER)

Scenario:
Current Environment: Deployment is performed manually by data scientists, with limited rollback capability.
The correct deployment type to address this problem is Managed online endpoints with traffic splitting.
Native traffic control: Managed online endpoints naturally support native blue-green deployments under a single HTTP endpoint.
Granular traffic splitting: You can deploy a new model version (e.g., green) with 0% live traffic, safely test it, and then incrementally shift traffic from the old version (e.g., blue).Instant rollbacks:
If the new model shows bugs or degrades performance, you can immediately change the traffic percentage configuration back to 100% for the old deployment.
This eliminates the risks of limited rollback capabilities.
No infrastructure overhead: Unlike setting up manual routing, Azure handles the underlying infrastructure and routing mechanisms in a turnkey fashion.
Incorrect:
[Not A]
VM-hosted REST APIs: This forces you to build, maintain, and configure your own custom load balancers and deployment scripts to handle routing and rollbacks, which increases operational risk and complexity.
[Not B]
Azure Kubernetes Service with blue-green switching: While it supports blue-green deployments, it requires you to manage complex Kubernetes infrastructure, service meshes, or ingress controllers manually to handle the traffic switching.
[Not D]
Batch endpoints: These are designed for long-running, asynchronous processing of large data batches rather than real-time requests where instant live-traffic rollback is required.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-safely-rollout-online-endpoints

NEW QUESTION: 63
Hotspot Question
You have an Azure Machine Learning workspace.
You plan to set up logging and tracking experiments by using MLflow Tracking.
You need to log the accuracy as a numerical value and the training loss as a plot.
How should you complete the commands? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Logging experiments with MLflow

```
import mlflow
from matplotlib import pyplot as plt

mlflow.set_experiment("mlflow-experiment")

plt.savefig('plot.png')

num1 = 82
img1 = "plot.png"

with mlflow.start_run() as run:
    mlflow.
```

- log_artifact("log", num1)
- log_metric("log", num1)
- log_metrics("log", num1)
- log_params("log", num1)

```
    mlflow.
```

- log_artifact(img1)
- log_artifacts(img1)
- log_figure(fig, img1)
- log_image(img, img1)

Answer:

Microsoft
freecram.net

Logging experiments with MLflow

```
import mlflow
from matplotlib import pyplot as plt

mlflow.set_experiment("mlflow-experiment")

plt.savefig('plot.png')

num1 = 82
img1 = "plot.png"

with mlflow.start_run() as run:
```

mlflow. [log_artifact("log", num1) / log_metric("log", num1) / log_metrics("log", num1) / log_params("log", num1)]

mlflow. [log_artifact(img1) / log_artifacts(img1) / log_figure(fig, img1) / log_image(img, img1)]

Microsoft freecram.net

Explanation:
Box 1: log_metric("log", num1)
Log Numerical Metrics
To track a single numerical value that changes over time (like accuracy or evaluation scores), you use the mlflow.log_metric() function. This enables Azure Machine Learning to plot the value on a performance graph automatically.
Box 2: log_artifact(img1)
Log Plots and FilesTo save external files, images, or plots generated during training, you use the mlflow.log_artifact() function. This uploads the specified file (in this case, your saved Matplotlib image) to the run's artifact storage.
Reference:
https://levelup.gitconnected.com/mlflow-made-easy-your-beginners-guide-bf63f8fed915

NEW QUESTION: 64
Drag and Drop Question
You develop a flow for a Microsoft Foundry project.
You plan to use outputs generated by running the flow to determine the following information:
- the number of tokens used by each large language model (LLM) node of
the flow
- the accuracy of the model used by the flow
You need to examine the output that provides the required information.
Which output type should you examine? To answer, move the appropriate output types to the correct evaluations. You may use each output type once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.



Answer:



Explanation:

Box 1: Traces

The appropriate output flow type to determine the number of tokens used by each Large Language Model (LLM) node is Traces.

While Metrics provide a macro-level, aggregated overview of total token usage across an entire application or resource, Traces capture detailed, node-by-node execution details.

When a prompt flow runs in Microsoft Foundry, Traces track:

The precise execution path of individual items

Inputs and outputs for each specific LLM node

Detailed telemetry records-such as input_tokens, output_tokens, and total_tokens-bound to that exact step Box 2: Metrics The most appropriate output flow type to determine the accuracy of the model is Metrics.

Metrics are quantitative measurements (such as accuracy, F1-score, precision, recall, or mean squared error) specifically calculated by evaluating model predictions against ground truth data.

Reference:

https://huggingface.co/docs/evaluate/a_quick_tour

NEW QUESTION: 65

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure Machine Learning workspace. You connect to a terminal session from the Notebooks page in Azure Machine Learning studio.

You plan to add a new Jupyter kernel that will be accessible from the same terminal session.

You need to perform the task that must be completed before you can add the new kernel.

Solution: Create an environment.

Does the solution meet the goal?

A. Yes
B. No
Answer: (SHOW ANSWER)
Correct:
* Create an environment.
Incorrect:
* Delete the Python 3.6 - AzureML kernel.
* Delete the Python 3.8 - AzureML kernel.
Note:
Before you can add a new Jupyter kernel on an Azure Machine Learning compute instance terminal, you must create a Conda environment.
Required Workflow
To officially provision and expose the new kernel to your Azure Machine Learning studio Notebooks, you need to execute the following full process from your terminal session:
Create the environment: Provision a new isolated environment (e.g., using conda create -n newenv python=3.10).
Activate the environment: Run conda activate newenv.
Install dependencies: Add the required ipykernel package using conda install ipykernel or pip install ipykernel.
Register the kernel: Bind the new environment configuration to the global Jupyter directory by running:
python -m ipykernel install --user --name newenv --display-name "My New Kernel" Reference:
https://docs.azure.cn/en-us/machine-learning/how-to-access-terminal

NEW QUESTION: 66
During training, pipelines occasionally fail due to schema mismatch caused by upstream data changes. You need a robust and automated solution that prevents invalid data from reaching training steps. What is the BEST approach?
A. Ignore schema differences
B. Add a data validation component in pipeline
C. Retrain manually when failure occurs
D. Use larger compute
Answer: B (LEAVE A REPLY)
A data validation component ensures that incoming data matches the expected schema before training begins. This prevents pipeline failures and avoids training on corrupted or incomplete data. Ignoring schema mismatches can introduce silent errors, making debugging difficult and compromising model quality.

NEW QUESTION: 67
Your ML pipeline contains independent feature engineering steps that currently execute sequentially, increasing overall runtime. You want to optimize execution without modifying logic.
What is the BEST solution?
A. Combine all steps
B. Enable parallel step execution
C. Increase compute size
D. Reduce dataset size
Answer: (SHOW ANSWER)

Parallel execution allows independent pipeline steps to run simultaneously, reducing total execution time without altering logic. This is more efficient than increasing compute resources, which may raise costs without guaranteeing proportional performance improvements.

NEW QUESTION: 68
A team is working in Microsoft Foundry to test and compare large language model (LLM) prompt variants in a development environment.
The team requires consistent inputs to evaluate prompt variants without relying on live user traffic.
You need to create a controlled evaluation of input data.
Which action should you perform first?
A. Generate synthetic interaction data.
B. Configure content filters.
C. Apply a blocklist.
D. Enable observability metrics.
Answer: (SHOW ANSWER)
In Microsoft Foundry, you can test and compare LLM prompt variants using Prompt Flow. This environment allows you to create a controlled development workflow with consistent inputs, bypassing the need for live user traffic.
1. Generate Synthetic Interaction Data
Generating synthetic data is the recommended first step to build a robust, diverse test dataset without manual effort.
2. Compare Prompt Variants
Once your synthetic dataset is ready, use Prompt Flow to iterate on your prompts.
3. Controlled Evaluation
To measure performance objectively, run automated Eval Runs.
Reference:
https://learn.microsoft.com/en-us/azure/foundry-classic/concepts/prompt-flow

NEW QUESTION: 69
Hotspot Question
A team trains an MLflow model that scores customer churn risk. The model will be consumed by different downstream systems.
One system requests predictions synchronously during customer interactions.
Another system submits files containing millions of records for scheduled scoring.
You need to deploy the model by using managed inference options that match each usage pattern.
Which option should you use for each usage pattern? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

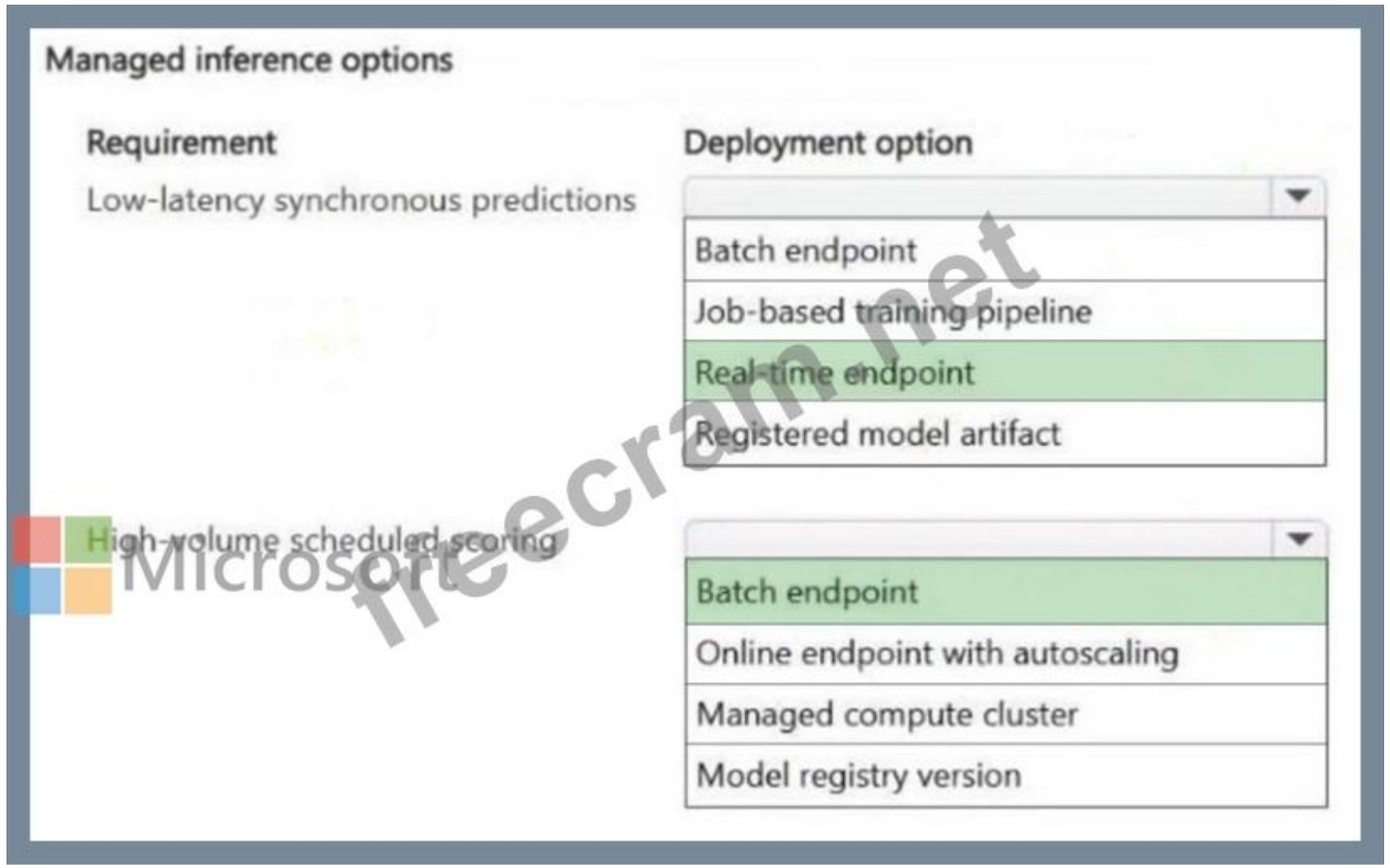
Managed inference options

Requirement	Deployment option
Low-latency synchronous predictions	Batch endpoint Job-based training pipeline Real-time endpoint Registered model artifact
High-volume scheduled scoring	Batch endpoint Online endpoint with autoscaling Managed compute cluster Model registry version

freecram.net

Microsoft

Answer:



NEW QUESTION: 70

An organization validates generative AI applications during CI/CD Microsoft Foundry.
Evaluation must run automatically and block releases when quality thresholds are NOT met.
Manual evaluation is no longer acceptable.
Evaluation must use both predefined quality metrics and custom safety checks.
You need to implement an automated evaluation workflow that supports both built-in and custom metrics.
What should you do?

A. Enable application tracing to collect runtime telemetry.

B. Review evaluation results manually after deployment.

C. Monitor latency metrics during model inference.

D. Implement an evaluation step by using GitHub Actions.

Answer: (SHOW ANSWER)

To implement an automated evaluation step in GitHub Actions for Microsoft Foundry AI, you can use the Microsoft Foundry Evaluation GitHub Action (or the Azure AI Evaluation SDK).

This setup allows you to run both built-in metrics (like groundedness or coherence) and custom safety checks, then fail the build if scores fall below your defined thresholds.

Implementation Steps

1. Define Your Evaluators

You need to configure which metrics to use. Microsoft Foundry supports two main types:

Built-in Metrics: Pre-trained models that score quality (coherence, fluency) and safety (hate, violence, self-harm).
Custom Metrics: Python-based evaluators you define to check domain-specific requirements.
2. Configure the GitHub Actions Workflow
Create a .yml file in your .github/workflows directory. This workflow will:
Trigger on a pull request or commit.
Authenticate with Azure/Foundry.
Run Evaluation using the microsoft/ai-agent-evals action.
Enforce Thresholds to block the release if quality is insufficient.
Key Components for "Block Release" Logic
To ensure the release is blocked, your workflow must include a gating step that interprets the evaluation results:
Reference:
https://learn.microsoft.com/en-us/training/modules/automated-evaluation-genaiops

NEW QUESTION: 71
You have a Microsoft Foundry project.
You plan to use the Microsoft Foundry portal to fine-tune a base Azure OpenAI Service model that can accept both text and images as input.
You need to choose the suitable model.
Which model should you choose?
A. davinci-002
B. gpt-4o
C. gpt-35-turbo
D. gpt-4
Answer: (SHOW ANSWER)
An appropriate base model for fine-tuning with text and image inputs in the Microsoft Foundry portal is GPT-4o (specifically versions like gpt-4o-2024-05-13 or newer). This model supports multi-modal training data, allowing you to improve performance on tasks requiring both visual understanding and textual response within Azure OpenAI Service.
Reference:
https://techcommunity.microsoft.com/blog/azure-ai-foundry-blog/announcing-new-fine-tuning-capabilities-with-images-on-azure-openai-service/4303695

NEW QUESTION: 72
A data science team trains a classification model that predicts loan approval outcomes.
Before registering the model, the team must ensure the following:
- Predictions must not disproportionately impact protected groups.
- Prediction errors can be evaluated across different data segments.
You need to assess whether the model meets Responsible AI expectations.
Which two approaches should you use? Each correct answer presents part of the solution.
Choose two.
NOTE: Each correct selection is worth one point.
A. Analyze error rates across the global cohort.
B. Measure endpoint latency under load.
C. Validate inference schema compatibility.

D. Evaluate feature importance for prediction transparency.
E. Analyze error rates across defined demographic cohorts.
Answer: (SHOW ANSWER)
[D]
To evaluate a trained loan classification model for Responsible AI expectations--ensuring no disproportionate impact on protected groups, evaluating error across segments, and verifying prediction transparency--you can employ SHAP (SHapley Additive exPlanations) values to assess feature importance.
This approach allows you to identify which variables (e.g., credit history, debt levels) drive the model's predictions, fostering trust and fairness.
Feature Importance for Transparency: Use SHAP (model-agnostic) or LIME (local approximations) to explain why the model approved or denied a loan. These techniques identify how each feature contributes to individual predictions.
[E]
To ensure a trained loan approval classification model meets responsible AI expectations-- specifically, that it does not disproportionately impact protected groups and that errors can be evaluated across segments--you should analyze error rates across defined demographic cohorts using Fairness-Aware Machine Learning metrics.
Reference:
https://urfpublishers.com/journal/artificial-intelligence/article/view/explainable-aiml-testing-ensuring-transparency-accountability-and-compliance
https://timvero.com/blog/ethics-in-automated-lending-can-ai-make-fair-credit-decisions

NEW QUESTION: 73
Hotspot Question
You monitor an Azure Machine Learning classification training experiment named train_classification on Azure Notebooks.
You must store a table named table as an artifact in Azure Machine Learning Studio during model training.
You need to collect and list the metrics by using MLflow.
How should you complete the code segment? To answer, select the appropriate option in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
from mlflow.tracking import MlflowClient
import json
mlflow.set_experiment("train_classification")
mlflow_run = mlflow.start_run()
row1 = {"table.col1": 5, "table.col2": 10}
mlflow.[ ▼ ](row1)
```

- log_artifact
- log_image
- log_metrics

```
with open("table.json", 'w') as f:
json.dump(table, f)
mlflow.[ ▼ ]("table.json")
```

- log_artifact
- log_metric
- log_metrics

```
client = MlflowClient()
finished_mlflow_run =
MlflowClient().get_run( [ ▼ ] )
```

- mlflow_run.info.run_id
- mlflow_run.info.run_id, path="accuracy_table.json"
- mlflow_run.info.experiment_id

Microsoft
freecram.net

Answer:

Answer Area

```
from mlflow.tracking import MlflowClient
import json
mlflow.set_experiment("train_classification")
mlflow_run = mlflow.start_run()
row1 = {"table.col1": 5, "table.col2": 10}
mlflow.[ ▼ ](row1)
```

- log_artifact
- log_image
- log_metrics

```
with open("table.json", 'w') as f:
json.dump(table, f)
mlflow.[ ▼ ]("table.json")
```

- log_artifact
- log_metric
- log_metrics

```
client = MlflowClient()
finished_mlflow_run =
MlflowClient().get_run( [ ▼ ] )
```

- mlflow_run.info.run_id
- mlflow_run.info.run_id, path="accuracy_table.json"
- mlflow_run.info.experiment_id

freecram.net
Microsoft

Explanation:

Box 1: log_metrics

Log Metrics dictionary

The dictionary row1 contains key-value pairs representing metrics. The mlflow.log_metrics() function is used to log multiple metrics simultaneously.

Box 2: log_artifact

Save Table Artifact

The code writes a JSON file locally and needs to upload it to the experiment run. The mlflow.log_artifact() function logs a local file or directory as an artifact in Azure Machine Learning Studio.

Box 3: mlflow_run.info.run_id

Retrieve Run ID

To fetch the finalized run data using the MlflowClient, you need to pass the unique run ID string.

This identifier is accessed via the active run object using mlflow_run.info.run_id. Note that you also need to use the instantiated client object variable instead of the class name MlflowClient to call the method correctly.

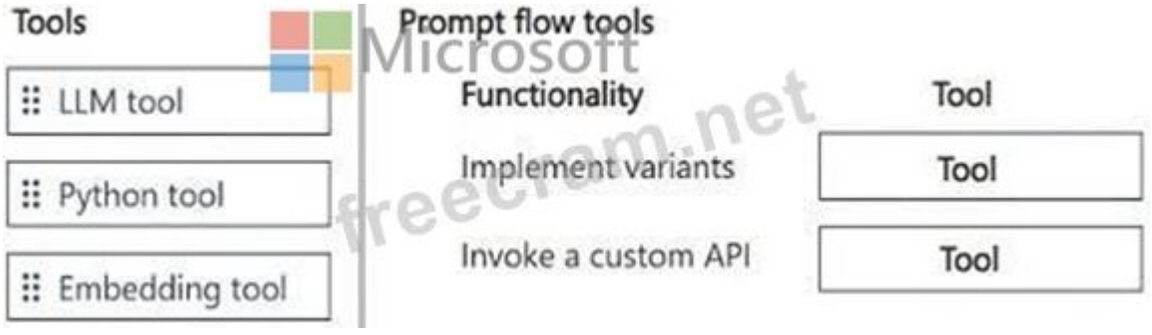
Reference:

https://levelup.gitconnected.com/mlops-mastering-mlflow-unlocking-efficient-model-management-and-experiment-tracking-d9d0e71cc697

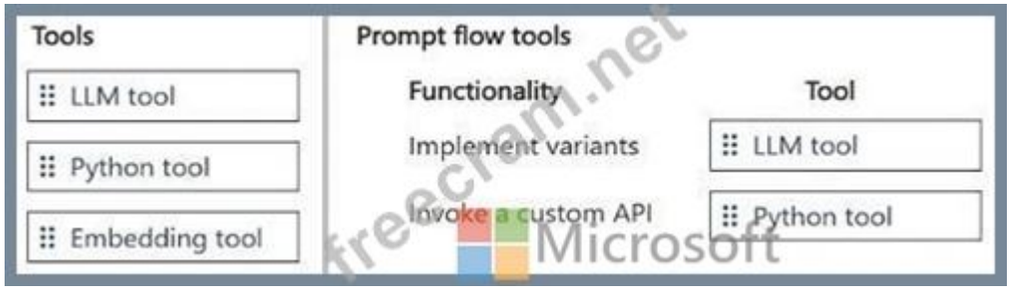
NEW QUESTION: 74

Drag and Drop Question

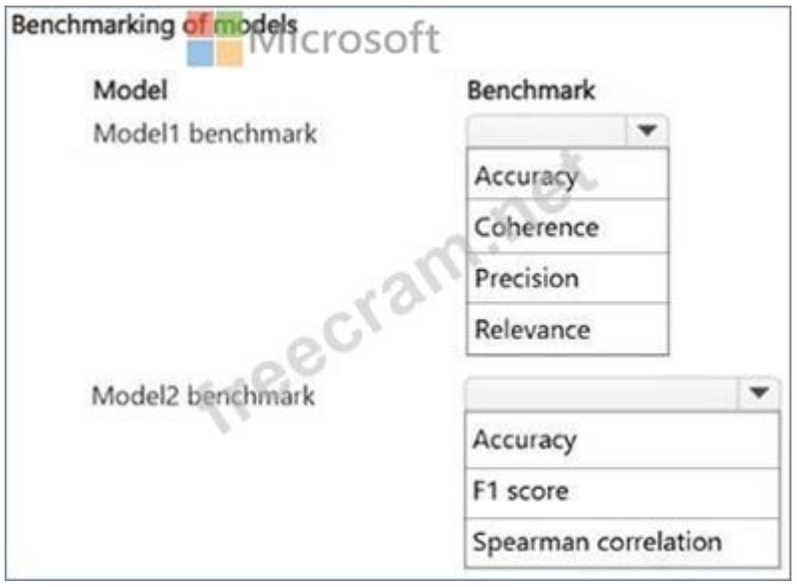
You develop a Prompt flow in Microsoft Foundry project.
You plan to use variants and invoke a custom API in the flow.
You need to add tools to the flow that will implement the planned functionality. Your solution must minimize development efforts.
Which tools should you use? To answer, move the appropriate tools to the correct functionalities.
You may use each tool once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



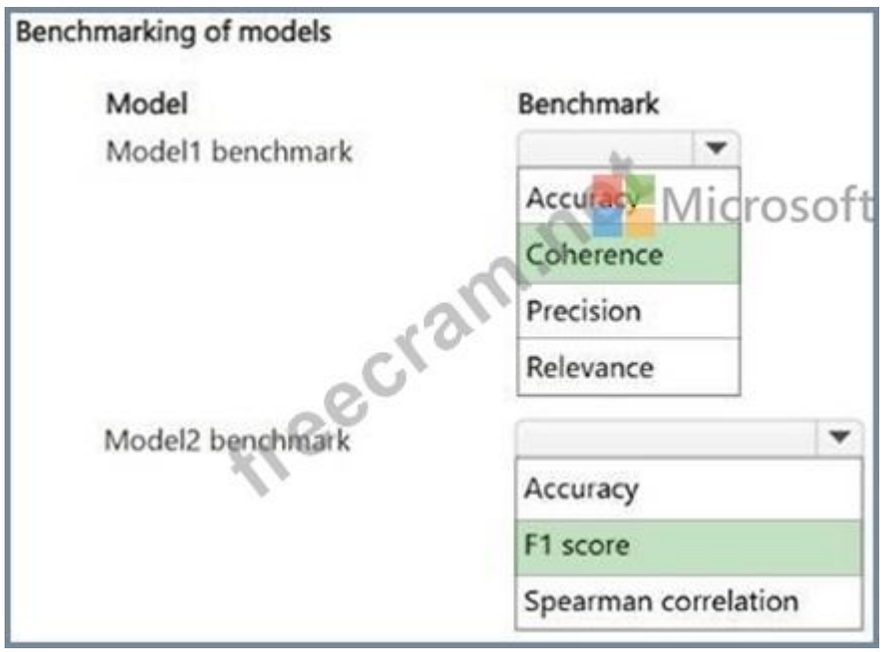
Answer:



NEW QUESTION: 75
Hotspot Question
You manage a Microsoft Foundry project.
You plan to build a RAG solution.
The solution must include two models:
- One for text output, named Model1. This model must resemble human
language and read naturally.
- One for creating embeddings, named Model2. This model must maximize
the retrieval of relevant results (high recall) while minimizing
irrelevant or incorrect matches (high precision).
You need to compare different models by using benchmarking metrics to select the appropriate models for Model1 and Model2.
Which benchmarking metric should you select for each model? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 76
A company requires that only models meeting predefined performance thresholds are registered and deployed. The solution must be fully automated within the ML workflow. What should you implement?
A. Manual approval gate
B. Conditional logic in pipeline
C. Azure Monitor alerts
D. GitHub pull request checks
Answer: (SHOW ANSWER)
Conditional logic in pipelines allows automated decisions based on evaluation metrics, ensuring only models that meet performance thresholds are registered and deployed. This maintains full automation. Manual approval steps interrupt automation and are not scalable in continuous integration and deployment workflows.

Valid AI-300 Dumps shared by ExamDiscuss.com for Helping Passing AI-300 Exam! ExamDiscuss.com now offer the **newest AI-300 exam dumps**, the ExamDiscuss.com AI-300 exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com AI-300 dumps with Test Engine here: https://www.examdiscuss.com/Microsoft/exam/AI-300/premium/ (**232** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)

Valid AI-300 Dumps shared by ExamDiscuss.com for Helping Passing AI-300 Exam! ExamDiscuss.com now offer the **newest AI-300 exam dumps**, the ExamDiscuss.com AI-300 exam **questions have been updated** and **answers have been corrected** get the **newest** ExamDiscuss.com AI-300 dumps with Test Engine here: https://www.examdiscuss.com/Microsoft/exam/AI-300/premium/ (**232** Q&As Dumps, **35%OFF** **Special Discount Code:** **freecram**)