

New Databricks-Generative-AI-Engineer-Associate Practice Questions | Customized Databricks-Generative-AI-Engineer-Associate Lab Simulation



2025 Latest Prep4cram Databricks-Generative-AI-Engineer-Associate PDF Dumps and Databricks-Generative-AI-Engineer-Associate Exam Engine Free Share: <https://drive.google.com/open?id=1q9dsFZhEmsk3AfbfhIIISOB2fuvOF6pK>

Our system is high effective and competent. After the clients pay successfully for the Databricks-Generative-AI-Engineer-Associate certification material the system will send the products to the clients by the mails. The clients click on the links in the mails and then they can use the Databricks-Generative-AI-Engineer-Associate prep guide dump immediately. Our system provides safe purchase procedures to the clients and we guarantee the system won't bring the virus to the clients' computers and the successful payment for our Databricks-Generative-AI-Engineer-Associate learning file. Our system is strictly protect the clients' privacy and sets strict interception procedures to forestall the disclosure of the clients' private important information. Our system will automatically send the updates of the Databricks-Generative-AI-Engineer-Associate learning file to the clients as soon as the updates are available. So our system is wonderful.

Databricks Databricks-Generative-AI-Engineer-Associate Exam Syllabus Topics:

Topic	Details
Topic 1	<ul style="list-style-type: none"> Data Preparation: Generative AI Engineers covers a chunking strategy for a given document structure and model constraints. The topic also focuses on filter extraneous content in source documents. Lastly, Generative AI Engineers also learn about extracting document content from provided source data and format.

Topic 2	<ul style="list-style-type: none"> • Evaluation and Monitoring: This topic is all about selecting an LLM choice and key metrics. Moreover, Generative AI Engineers learn about evaluating model performance. Lastly, the topic includes sub-topics about inference logging and usage of Databricks features.
Topic 3	<ul style="list-style-type: none"> • Assembling and Deploying Applications: In this topic, Generative AI Engineers get knowledge about coding a chain using a pyfunc mode, coding a simple chain using langchain, and coding a simple chain according to requirements. Additionally, the topic focuses on basic elements needed to create a RAG application. Lastly, the topic addresses sub-topics about registering the model to Unity Catalog using MLflow.

>> New Databricks-Generative-AI-Engineer-Associate Practice Questions <<

Customized Databricks Databricks-Generative-AI-Engineer-Associate Lab Simulation | New Databricks-Generative-AI-Engineer-Associate Test Fee

Our company is a multinational company with sales and after-sale service of Databricks-Generative-AI-Engineer-Associate exam torrent compiling departments throughout the world. In addition, our company has become the top-notch one in the fields, therefore, if you are preparing for the exam in order to get the related certification, then the Databricks Certified Generative AI Engineer Associate exam question compiled by our company is your solid choice. All employees worldwide in our company operate under a common mission: to be the best global supplier of electronic Databricks-Generative-AI-Engineer-Associate Exam Torrent for our customers through product innovation and enhancement of customers' satisfaction. Wherever you are in the world we will provide you with the most useful and effectively Databricks-Generative-AI-Engineer-Associate guide torrent in this website, which will help you to pass the exam as well as getting the related certification with a great ease.

Databricks Certified Generative AI Engineer Associate Sample Questions (Q38-Q43):

NEW QUESTION # 38

Generative AI Engineer at an electronics company just deployed a RAG application for customers to ask questions about products that the company carries. However, they received feedback that the RAG response often returns information about an irrelevant product.

What can the engineer do to improve the relevance of the RAG's response?

- A. Use a different semantic similarity search algorithm
- B. Use a different LLM to improve the generated response
- **C. Assess the quality of the retrieved context**
- D. Implement caching for frequently asked questions

Answer: C

Explanation:

In a Retrieval-Augmented Generation (RAG) system, the key to providing relevant responses lies in the quality of the retrieved context. Here's why option A is the most appropriate solution:

* **Context Relevance:** The RAG model generates answers based on retrieved documents or context. If the retrieved information is about an irrelevant product, it suggests that the retrieval step is failing to select the right context. The Generative AI Engineer must first assess the quality of what is being retrieved and ensure it is pertinent to the query.

* **Vector Search and Embedding Similarity:** RAG typically uses vector search for retrieval, where embeddings of the query are matched against embeddings of product descriptions. Assessing the semantic similarity search process ensures that the closest matches are actually relevant to the query.

* **Fine-tuning the Retrieval Process:** By improving the retrieval quality, such as tuning the embeddings or adjusting the retrieval strategy, the system can return more accurate and relevant product information.

* **Why Other Options Are Less Suitable:**

* **B (Caching FAQs):** Caching can speed up responses for frequently asked questions but won't improve the relevance of the retrieved content for less frequent or new queries.

* **C (Use a Different LLM):** Changing the LLM only affects the generation step, not the retrieval process, which is the core issue here.

* **D (Different Semantic Search Algorithm):** This could help, but the first step is to evaluate the current retrieval context before

replacing the search algorithm.

Therefore, improving and assessing the quality of the retrieved context (option A) is the first step to fixing the issue of irrelevant product information.

NEW QUESTION # 39

A Generative AI Engineer is building a production-ready LLM system which replies directly to customers.

The solution makes use of the Foundation Model API via provisioned throughput. They are concerned that the LLM could potentially respond in a toxic or otherwise unsafe way. They also wish to perform this with the least amount of effort.

Which approach will do this?

- A. Ask users to report unsafe responses
- B. Add a regex expression on inputs and outputs to detect unsafe responses.
- **C. Host Llama Guard on Foundation Model API and use it to detect unsafe responses**
- D. Add some LLM calls to their chain to detect unsafe content before returning text

Answer: C

Explanation:

The task is to prevent toxic or unsafe responses in an LLM system using the Foundation Model API with minimal effort. Let's assess the options.

* Option A: Host Llama Guard on Foundation Model API and use it to detect unsafe responses

* Llama Guard is a safety-focused model designed to detect toxic or unsafe content. Hosting it via the Foundation Model API (a Databricks service) integrates seamlessly with the existing system, requiring minimal setup (just deployment and a check step), and leverages provisioned throughput for performance.

* Databricks Reference: "Foundation Model API supports hosting safety models like Llama Guard to filter outputs efficiently" ("Foundation Model API Documentation," 2023).

* Option B: Add some LLM calls to their chain to detect unsafe content before returning text

* Using additional LLM calls (e.g., prompting an LLM to classify toxicity) increases latency, complexity, and effort (crafting prompts, chaining logic), and lacks the specificity of a dedicated safety model.

* Databricks Reference: "Ad-hoc LLM checks are less efficient than purpose-built safety solutions" ("Building LLM Applications with Databricks").

* Option C: Add a regex expression on inputs and outputs to detect unsafe responses

* Regex can catch simple patterns (e.g., profanity) but fails for nuanced toxicity (e.g., sarcasm, context-dependent harm), requiring significant manual effort to maintain and update rules.

* Databricks Reference: "Regex-based filtering is limited for complex safety needs" ("Generative AI Cookbook").

* Option D: Ask users to report unsafe responses

* User reporting is reactive, not preventive, and places burden on users rather than the system. It doesn't limit unsafe outputs proactively and requires additional effort for feedback handling.

* Databricks Reference: "Proactive guardrails are preferred over user-driven monitoring" ("Databricks Generative AI Engineer Guide").

Conclusion: Option A (Llama Guard on Foundation Model API) is the least-effort, most effective approach, leveraging Databricks' infrastructure for seamless safety integration.

NEW QUESTION # 40

Which indicator should be considered to evaluate the safety of the LLM outputs when qualitatively assessing LLM responses for a translation use case?

- A. The latency of the response and the length of text generated
- **B. The accuracy and relevance of the responses**
- C. The similarity to the previous language
- D. The ability to generate responses in code

Answer: B

Explanation:

* Problem Context: When assessing the safety and effectiveness of LLM outputs in a translation use case, it is essential to ensure that the translations accurately and relevantly convey the intended message. The evaluation should focus on how well the LLM understands and processes different languages and contexts.

* Explanation of Options:

- * Option A: The ability to generate responses in code- This is not relevant to translation quality or safety.
- * Option B: The similarity to the previous language- While ensuring that translations preserve the original's intent is important, this doesn't directly address the overall quality or safety of the translation.
- * Option C: The latency of the response and the length of text generated- These operational metrics are less critical in assessing the qualitative aspects of translation safety.
- * Option D: The accuracy and relevance of the responses- This is crucial in translation to ensure that the translated content is true to the original in meaning and appropriateness. Accuracy and relevance directly impact the effectiveness and safety of translations, especially in sensitive or nuanced contexts.

Thus, Option D is the most important indicator when evaluating the safety of LLM outputs in translation, focusing on the core aspects that determine the utility and trustworthiness of translated content.

NEW QUESTION # 41

A Generative AI Engineer is developing a RAG application and would like to experiment with different embedding models to improve the application performance.

Which strategy for picking an embedding model should they choose?

- A. Pick an embedding model with multilingual support to support potential multilingual user questions
- B. Pick the most recent and most performant open LLM released at the time
- C. pick the embedding model ranked highest on the Massive Text Embedding Benchmark (MTEB) leaderboard hosted by HuggingFace
- **D. Pick an embedding model trained on related domain knowledge**

Answer: D

Explanation:

The task involves improving a Retrieval-Augmented Generation (RAG) application's performance by experimenting with embedding models. The choice of embedding model impacts retrieval accuracy, which is critical for RAG systems. Let's evaluate the options based on Databricks Generative AI Engineer best practices.

* Option A: Pick an embedding model trained on related domain knowledge

* Embedding models trained on domain-specific data (e.g., industry-specific corpora) produce vectors that better capture the semantics of the application's context, improving retrieval relevance. For RAG, this is a key strategy to enhance performance.

* Databricks Reference: "For optimal retrieval in RAG systems, select embedding models aligned with the domain of your data" ("Building LLM Applications with Databricks," 2023).

* Option B: Pick the most recent and most performant open LLM released at the time

* LLMs are not embedding models; they generate text, not embeddings for retrieval. While recent LLMs may be performant for generation, this doesn't address the embedding step in RAG. This option misunderstands the component being selected.

* Databricks Reference: Embedding models and LLMs are distinct in RAG workflows:

"Embedding models convert text to vectors, while LLMs generate responses" ("Generative AI Cookbook").

* Option C: Pick the embedding model ranked highest on the Massive Text Embedding Benchmark (MTEB) leaderboard hosted by HuggingFace

* The MTEB leaderboard ranks models across general tasks, but high overall performance doesn't guarantee suitability for a specific domain. A top-ranked model might excel in generic contexts but underperform on the engineer's unique data.

* Databricks Reference: General performance is less critical than domain fit. "Benchmark rankings provide a starting point, but domain-specific evaluation is recommended" ("Databricks Generative AI Engineer Guide").

* Option D: Pick an embedding model with multilingual support to support potential multilingual user questions

* Multilingual support is useful only if the application explicitly requires it. Without evidence of multilingual needs, this adds complexity without guaranteed performance gains for the current use case.

* Databricks Reference: "Choose features like multilingual support based on application requirements" ("Building LLM-Powered Applications").

Conclusion: Option A is the best strategy because it prioritizes domain relevance, directly improving retrieval accuracy in a RAG system-aligning with Databricks' emphasis on tailoring models to specific use cases.

NEW QUESTION # 42

A Generative AI Engineer would like an LLM to generate formatted JSON from emails. This will require parsing and extracting the following information: order ID, date, and sender email. Here's a sample email:


```
Date: April 23, 2024
Time: 4:22 PM
From: anjali.thayer@computex.org
To: cust_service@realtek.com
Subject: Shipment details
```

Hey there,

I have a shipment (order ID is CD34RFT) can you please send me an update?

Thank you,
Anjali

They will need to write a prompt that will extract the relevant information in JSON format with the highest level of output accuracy. Which prompt will do that?

- A. You will receive customer emails and need to extract date, sender email, and order ID. Return the extracted information in a human-readable format.
- B. You will receive customer emails and need to extract date, sender email, and order ID. Return the extracted information in JSON format.
Here's an example: {"date": "April 16, 2024", "sender_email": "sarah.lee925@gmail.com", "order_id": "RE987D"}
- C. You will receive customer emails and need to extract date, sender email, and order ID. You should return the date, sender email, and order ID information in JSON format.
- D. You will receive customer emails and need to extract date, sender email, and order ID. Return the extracted information in JSON format.

Answer: B

Explanation:

Problem Context: The goal is to parse emails to extract certain pieces of information and output this in a structured JSON format. Clarity and specificity in the prompt design will ensure higher accuracy in the LLM's responses.

Explanation of Options:

- * Option A: Provides a general guideline but lacks an example, which helps an LLM understand the exact format expected.
- * Option B: Includes a clear instruction and a specific example of the output format. Providing an example is crucial as it helps set the pattern and format in which the information should be structured, leading to more accurate results.
- * Option C: Does not specify that the output should be in JSON format, thus not meeting the requirement.
- * Option D: While it correctly asks for JSON format, it lacks an example that would guide the LLM on how to structure the JSON correctly.

Therefore, Option B is optimal as it not only specifies the required format but also illustrates it with an example, enhancing the likelihood of accurate extraction and formatting by the LLM.

NEW QUESTION # 43

.....

The Prep4cram is committed to making the Databricks Databricks-Generative-AI-Engineer-Associate certification exam preparation simple, smart, and successful. To achieve this objective Prep4cram is offering top-notch and real Databricks-Generative-AI-Engineer-Associate exam questions in three different formats. The names of these Databricks Certified Generative AI Engineer Associate (Databricks-Generative-AI-Engineer-Associate) exam questions formats are PDF files, desktop practice test software, and web-based Databricks-Generative-AI-Engineer-Associate practice test software.

Customized Databricks-Generative-AI-Engineer-Associate Lab Simulation: https://www.prep4cram.com/Databricks-Generative-AI-Engineer-Associate_exam-questions.html

- 2026 Professional Databricks-Generative-AI-Engineer-Associate – 100% Free New Practice Questions | Customized Databricks Certified Generative AI Engineer Associate Lab Simulation ☐ Open ☒ www.validtorrent.com ☒ enter “Databricks-Generative-AI-Engineer-Associate” and obtain a free download ☐ Reliable Databricks-Generative-AI-Engineer-Associate Dumps Files
- Reliable Databricks-Generative-AI-Engineer-Associate Dumps Files ☐ Test Databricks-Generative-AI-Engineer-Associate Voucher ☐ Databricks-Generative-AI-Engineer-Associate Exam Brain Dumps ☐ Search for ➡ Databricks-

[illegible]

DOWNLOAD the newest Prep4cram Databricks-Generative-AI-Engineer-Associate PDF dumps from Cloud Storage for free:
<https://drive.google.com/open?id=1q9dsFZhEmsk3AfbhfIIISOB2fuvOF6pK>