

# Reliable Databricks-Generative-AI-Engineer-Associate Source Pass Certify| Efficient Exam Databricks-Generative-AI-Engineer-Associate Tips: Databricks Certified Generative AI Engineer Associate



Different with other similar education platforms on the internet, the Databricks-Generative-AI-Engineer-Associate guide torrent has a high hit rate, in the past, according to data from the students' learning to use the Databricks-Generative-AI-Engineer-Associate test torrent, 99% of these students can pass the qualification test and acquire the qualification of their yearning, this powerfully shows that the information provided by the Databricks-Generative-AI-Engineer-Associate Study Tool suit every key points perfectly, targeted training students a series of patterns and problem solving related routines, and let students answer up to similar topic.

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

Topic	Details
Topic 1	<ul style="list-style-type: none"><li>• 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.</li></ul>
Topic 2	<ul style="list-style-type: none"><li>• 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.</li></ul>
Topic 3	<ul style="list-style-type: none"><li>• Design Applications: The topic focuses on designing a prompt that elicits a specifically formatted response. It also focuses on selecting model tasks to accomplish a given business requirement. Lastly, the topic covers chain components for a desired model input and output.</li></ul>

>> Reliable Databricks-Generative-AI-Engineer-Associate Source <<

## High Pass-Rate Reliable Databricks-Generative-AI-Engineer-Associate Source Provide Prefect Assistance in Databricks-Generative-AI-Engineer-Associate Preparation

Opportunities are very important in this society. With the opportunity you can go further. However, it is difficult to seize the opportunity. Is your strength worthy of the opportunity before you? In any case, you really need to make yourself better by using our

Databricks-Generative-AI-Engineer-Associate training engine. With our Databricks-Generative-AI-Engineer-Associate Exam Questions, you can equip yourself with the most specialized knowledge of the subject. What is more, our Databricks-Generative-AI-Engineer-Associate study materials can help you get the certification. Imagine you're coming good future maybe you will make a better choice!

## Databricks Certified Generative AI Engineer Associate Sample Questions (Q40-Q45):

### NEW QUESTION # 40

A Generative AI Engineer has been asked to build an LLM-based question-answering application. The application should take into account new documents that are frequently published. The engineer wants to build this application with the least cost and least development effort and have it operate at the lowest cost possible.

Which combination of chaining components and configuration meets these requirements?

- A. The LLM needs to be frequently with the new documents in order to provide most up-to-date answers.
- B. For the question-answering application, prompt engineering and an LLM are required to generate answers.
- C. For the application a prompt, a retriever, and an LLM are required. The retriever output is inserted into the prompt which is given to the LLM to generate answers.
- D. For the application a prompt, an agent and a fine-tuned LLM are required. The agent is used by the LLM to retrieve relevant content that is inserted into the prompt which is given to the LLM to generate answers.

**Answer: C**

Explanation:

Problem Context: The task is to build an LLM-based question-answering application that integrates new documents frequently with minimal costs and development efforts.

Explanation of Options:

\* Option A: Utilizes a prompt and a retriever, with the retriever output being fed into the LLM. This setup is efficient because it dynamically updates the data pool via the retriever, allowing the LLM to provide up-to-date answers based on the latest documents without needing to frequently retrain the model. This method offers a balance of cost-effectiveness and functionality.

\* Option B: Requires frequent retraining of the LLM, which is costly and labor-intensive.

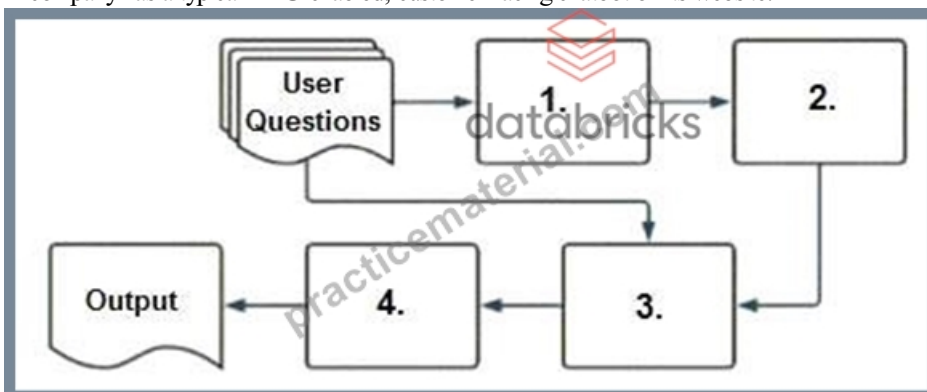
\* Option C: Only involves prompt engineering and an LLM, which may not adequately handle the requirement for incorporating new documents unless it's part of an ongoing retraining or updating mechanism, which would increase costs.

\* Option D: Involves an agent and a fine-tuned LLM, which could be overkill and lead to higher development and operational costs.

Option A is the most suitable as it provides a cost-effective, minimal development approach while ensuring the application remains up-to-date with new information.

### NEW QUESTION # 41

A company has a typical RAG-enabled, customer-facing chatbot on its website.



Select the correct sequence of components a user's questions will go through before the final output is returned. Use the diagram above for reference.

- A. 1.response-generating LLM, 2.context-augmented prompt, 3.vector search, 4.embedding model
- B. 1.embedding model, 2.vector search, 3.context-augmented prompt, 4.response-generating LLM
- C. 1.context-augmented prompt, 2.vector search, 3.embedding model, 4.response-generating LLM
- D. 1.response-generating LLM, 2.vector search, 3.context-augmented prompt, 4.embedding model

**Answer: B**

Explanation:

To understand how a typical RAG-enabled customer-facing chatbot processes a user's question, let's go through the correct sequence as depicted in the diagram and explained in option A:

- \* **Embedding Model (1):** The first step involves the user's question being processed through an embedding model. This model converts the text into a vector format that numerically represents the text. This step is essential for allowing the subsequent vector search to operate effectively.
- \* **Vector Search (2):** The vectors generated by the embedding model are then used in a vector search mechanism. This search identifies the most relevant documents or previously answered questions that are stored in a vector format in a database.
- \* **Context-Augmented Prompt (3):** The information retrieved from the vector search is used to create a context-augmented prompt. This step involves enhancing the basic user query with additional relevant information gathered to ensure the generated response is as accurate and informative as possible.
- \* **Response-Generating LLM (4):** Finally, the context-augmented prompt is fed into a response-generating large language model (LLM). This LLM uses the prompt to generate a coherent and contextually appropriate answer, which is then delivered as the final output to the user.

Why Other Options Are Less Suitable:

\* B, C, D: These options suggest incorrect sequences that do not align with how a RAG system typically processes queries. They misplace the role of embedding models, vector search, and response generation in an order that would not facilitate effective information retrieval and response generation.

Thus, the correct sequence is embedding model, vector search, context-augmented prompt, response-generating LLM, which is option A.

#### NEW QUESTION # 42

A Generative AI Engineer is using an LLM to classify species of edible mushrooms based on text descriptions of certain features. The model is returning accurate responses in testing and the Generative AI Engineer is confident they have the correct list of possible labels, but the output frequently contains additional reasoning in the answer when the Generative AI Engineer only wants to return the label with no additional text.

Which action should they take to elicit the desired behavior from this LLM?

- A. Use few shot prompting to instruct the model on expected output format
- B. Use zero shot prompting to instruct the model on expected output format
- C. Use zero shot chain-of-thought prompting to prevent a verbose output format
- **D. Use a system prompt to instruct the model to be succinct in its answer**

**Answer: D**

Explanation:

The LLM classifies mushroom species accurately but includes unwanted reasoning text, and the engineer wants only the label. Let's assess how to control output format effectively.

- \* **Option A:** Use few shot prompting to instruct the model on expected output format
- \* Few-shot prompting provides examples (e.g., input: description, output: label). It can work but requires crafting multiple examples, which is effort-intensive and less direct than a clear instruction.
- \* **Databricks Reference:** "Few-shot prompting guides LLMs via examples, effective for format control but requires careful design" ("Generative AI Cookbook").
- \* **Option B:** Use zero shot prompting to instruct the model on expected output format
- \* Zero-shot prompting relies on a single instruction (e.g., "Return only the label") without examples. It's simpler than few-shot but may not consistently enforce succinctness if the LLM's default behavior is verbose.
- \* **Databricks Reference:** "Zero-shot prompting can specify output but may lack precision without examples" ("Building LLM Applications with Databricks").
- \* **Option C:** Use zero shot chain-of-thought prompting to prevent a verbose output format
- \* Chain-of-Thought (CoT) encourages step-by-step reasoning, which increases verbosity-opposite to the desired outcome. This contradicts the goal of label-only output.
- \* **Databricks Reference:** "CoT prompting enhances reasoning but often results in detailed responses" ("Databricks Generative AI Engineer Guide").
- \* **Option D:** Use a system prompt to instruct the model to be succinct in its answer
- \* A system prompt (e.g., "Respond with only the species label, no additional text") sets a global instruction for the LLM's behavior. It's direct, reusable, and effective for controlling output style across queries.
- \* **Databricks Reference:** "System prompts define LLM behavior consistently, ideal for enforcing concise outputs" ("Generative AI Cookbook," 2023).

Conclusion: Option D is the most effective and straightforward action, using a system prompt to enforce succinct, label-only responses, aligning with Databricks' best practices for output control.

### NEW QUESTION # 43

A Generative AI Engineer is developing an LLM application that users can use to generate personalized birthday poems based on their names.

Which technique would be most effective in safeguarding the application, given the potential for malicious user inputs?

- **A. Implement a safety filter that detects any harmful inputs and ask the LLM to respond that it is unable to assist**
- B. Reduce the time that the users can interact with the LLM
- C. Ask the LLM to remind the user that the input is malicious but continue the conversation with the user
- D. Increase the amount of compute that powers the LLM to process input faster

**Answer: A**

Explanation:

In this case, the Generative AI Engineer is developing an application to generate personalized birthday poems, but there's a need to safeguard against malicious user inputs. The best solution is to implement a safety filter (option A) to detect harmful or inappropriate inputs.

\* Safety Filter Implementation: Safety filters are essential for screening user input and preventing inappropriate content from being processed by the LLM. These filters can scan inputs for harmful language, offensive terms, or malicious content and intervene before the prompt is passed to the LLM.

\* Graceful Handling of Harmful Inputs: Once the safety filter detects harmful content, the system can provide a message to the user, such as "I'm unable to assist with this request," instead of processing or responding to malicious input. This protects the system from generating harmful content and ensures a controlled interaction environment.

\* Why Other Options Are Less Suitable:

\* B (Reduce Interaction Time): Reducing the interaction time won't prevent malicious inputs from being entered.

\* C (Continue the Conversation): While it's possible to acknowledge malicious input, it is not safe to continue the conversation with harmful content. This could lead to legal or reputational risks.

\* D (Increase Compute Power): Adding more compute doesn't address the issue of harmful content and would only speed up processing without resolving safety concerns.

Therefore, implementing a safety filter that blocks harmful inputs is the most effective technique for safeguarding the application.

### NEW QUESTION # 44

What is the most suitable library for building a multi-step LLM-based workflow?

- A. TensorFlow
- **B. LangChain**
- C. Pandas
- D. PySpark

**Answer: B**

Explanation:

\* Problem Context: The Generative AI Engineer needs a tool to build a multi-step LLM-based workflow. This type of workflow often involves chaining multiple steps together, such as query generation, retrieval of information, response generation, and post-processing, with LLMs integrated at several points.

\* Explanation of Options:

\* Option A: Pandas: Pandas is a powerful data manipulation library for structured data analysis, but it is not designed for managing or orchestrating multi-step workflows, especially those involving LLMs.

\* Option B: TensorFlow: TensorFlow is primarily used for training and deploying machine learning models, especially deep learning models. It is not designed for orchestrating multi-step tasks in LLM-based workflows.

\* Option C: PySpark: PySpark is a distributed computing framework used for large-scale data processing. While useful for handling big data, it is not specialized for chaining LLM-based operations.

\* Option D: LangChain: LangChain is a purpose-built framework designed specifically for orchestrating multi-step workflows with large language models (LLMs). It enables developers to easily chain different tasks, such as retrieving documents, summarizing information, and generating responses, all in a structured flow. This makes it the best tool for building complex LLM-based workflows.

Thus, LangChain is the most suitable library for creating multi-step LLM-based workflows.

• • • • •

**Exam Databricks-Generative-AI-Engineer-Associate Tips:** <https://www.practicematerial.com/Databricks-Generative-AI-Engineer-Associate-exam-materials.html>

- [illegible]

Disposable vapes