

# Databricks-Generative-AI-Engineer-Associate テスト問題集 & Databricks-Generative-AI-Engineer-Associate 問題数



無料でクラウドストレージから最新のJpexam Databricks-Generative-AI-Engineer-Associate PDFダンプをダウンロードする: [https://drive.google.com/open?id=12oOIY8\\_UJWEdr\\_ZyJmKAIFxf7QAGz6pU](https://drive.google.com/open?id=12oOIY8_UJWEdr_ZyJmKAIFxf7QAGz6pU)

私たちのウェブサイトから見ると、Databricks-Generative-AI-Engineer-Associate 学習教材は3つのバージョンがあります。PDF、ソフトウェアとオンライン版です。Databricks-Generative-AI-Engineer-Associate PDF版は印刷できます。ソフトウェアとオンライン版はコンピュータで使用できます。コンピュータで学ぶことが難しい場合は、Databricks-Generative-AI-Engineer-Associate 学習教材の印刷資料で勉強できます。また、Databricks-Generative-AI-Engineer-Associate 学習教材の価格は合理的に設定されています。

## Databricks Databricks-Generative-AI-Engineer-Associate 認定試験の出題範囲:

トピック	出題範囲
トピック 1	<ul style="list-style-type: none"><li>評価と監視: このトピックでは、LLM の選択と主要なメトリックについて説明します。さらに、Generative AI エンジニアはモデルのパフォーマンスの評価について学習します。最後に、このトピックには推論ログと Databricks 機能の使用に関するサブトピックが含まれています。</li></ul>

トピック 2	<ul style="list-style-type: none"> <li>ガバナンス:試験を受けるジェネレーティブ AI エンジニアは、このトピックのマスキング手法、ガードレール手法、および法的</li> <li>ライセンス要件に関する知識を習得します。</li> </ul>
トピック 3	<ul style="list-style-type: none"> <li>データ準備:Generative AI エンジニアは、特定のドキュメント構造とモデル制約のチャンキング戦略について説明します。このトピックでは、ソースドキュメント内の不要なコンテンツのフィルター処理にも重点を置いています。最後に、Generative AI エンジニアは、提供されたソースデータと形式からドキュメントコンテンツを抽出する方法についても学習します。</li> </ul>
トピック 4	<ul style="list-style-type: none"> <li>アプリケーションの組み立てとデプロイ:このトピックでは、Generative AI エンジニアは、pyfunc モードを使用してチェーンをコーディングする方法、langchain を使用してシンプルなチェーンをコーディングする方法、要件に従ってシンプルなチェーンをコーディングする方法を学びます。さらに、このトピックでは、RAG アプリケーションを作成するために必要な基本要素に焦点を当てています。最後に、このトピックでは、MLflow を使用してモデルを Unity Catalog に登録する方法に関するサブトピックを取り上げます。</li> </ul>
トピック 5	<ul style="list-style-type: none"> <li>アプリケーション開発:このトピックでは、Generative AI エンジニアは、データの抽出に必要なツール、Langchain</li> <li>類似ツール、一般的な問題を特定するための応答の評価について学習します。さらに、このトピックには、LLM の応答の調整、LLM ガードレール、およびアプリケーションの属性に基づいた最適な LLM に関する質問が含まれています。</li> </ul>

>> Databricks-Generative-AI-Engineer-Associate テスト問題集 <<

## Databricks-Generative-AI-Engineer-Associate 問題数 & Databricks-Generative-AI-Engineer-Associate 的中率

Databricks-Generative-AI-Engineer-Associate トレーニング資料の PDF バージョン: Databricks Certified Generative AI Engineer Associate は読みやすく、覚えやすく、印刷要求をサポートしているため、紙で印刷して練習することができます。練習資料のソフトウェアバージョンは、シミュレーションテストシステムをサポートし、セットアップの時間を与えることには制限がありません。このバージョンは Windows システムユーザーのみをサポートすることに注意してください。Databricks-Generative-AI-Engineer-Associate 試験問題のオンライン版は、Databricks あらゆる種類の機器やデジタルデバイスに適しています。モバイルデータなしで練習することを条件に、オフラインでの運動をサポートします。豊富な練習資料はお客様のさまざまなニーズに対応でき、これらの Databricks-Generative-AI-Engineer-Associate 模擬練習にはすべて、Databricks テストに合格するために知っておく必要がある新しい情報が含まれています。あなたの個人的な好みに応じてそれらを選択することができます。

## Databricks Certified Generative AI Engineer Associate 認定 Databricks-Generative-AI-Engineer-Associate 試験問題 (Q70-Q75):

### 質問 # 70

A Generative AI Engineer is helping a cinema extend its website's chat bot to be able to respond to questions about specific showtimes for movies currently playing at their local theater. They already have the location of the user provided by location services to their agent, and a Delta table which is continually updated with the latest showtime information by location. They want to implement this new capability In their RAG application.

Which option will do this with the least effort and in the most performant way?

- A. Set up a task in Databricks Workflows to write the information in the Delta table periodically to an external database such as MySQL and query the information from there as part of the agent logic / tool implementation.
- B. implementation. Write the Delta table contents to a text column, then embed those texts using an embedding model and store these in the vector index. Look up the information based on the embedding as part of the agent logic / tool implementation.
- C. Query the Delta table directly via a SQL query constructed from the user's input using a text-to-SQL LLM in the agent logic / tool
- D. Create a Feature Serving Endpoint from a FeatureSpec that references an online store synced from the Delta table. Query

## the Feature Serving Endpoint as part of the agent logic / tool implementation.

正解: D

解説:

The task is to extend a cinema chatbot to provide movie showtime information using a RAG application, leveraging user location and a continuously updated Delta table, with minimal effort and high performance.

Let's evaluate the options.

\* Option A: Create a Feature Serving Endpoint from a FeatureSpec that references an online store synced from the Delta table.

Query the Feature Serving Endpoint as part of the agent logic / tool implementation

\* Databricks Feature Serving provides low-latency access to real-time data from Delta tables via an online store. Syncing the Delta table to a Feature Serving Endpoint allows the chatbot to query showtimes efficiently, integrating seamlessly into the RAG agent's tool logic. This leverages Databricks' native infrastructure, minimizing effort and ensuring performance.

\* Databricks Reference: "Feature Serving Endpoints provide real-time access to Delta table data with low latency, ideal for production systems" ("Databricks Feature Engineering Guide," 2023).

\* Option B: Query the Delta table directly via a SQL query constructed from the user's input using a text-to-SQL LLM in the agent logic / tool

\* Using a text-to-SQL LLM to generate queries adds complexity (e.g., ensuring accurate SQL generation) and latency (LLM inference + SQL execution). While feasible, it's less performant and requires more effort than a pre-built serving solution.

\* Databricks Reference: "Direct SQL queries are flexible but may introduce overhead in real-time applications" ("Building LLM Applications with Databricks").

\* Option C: Write the Delta table contents to a text column, then embed those texts using an embedding model and store these in the vector index. Look up the information based on the embedding as part of the agent logic / tool implementation

\* Converting structured Delta table data (e.g., showtimes) into text, embedding it, and using vector search is inefficient for structured lookups. It's effort-intensive (preprocessing, embedding) and less precise than direct queries, undermining performance.

\* Databricks Reference: "Vector search excels for unstructured data, not structured tabular lookups" ("Databricks Vector Search Documentation").

\* Option D: Set up a task in Databricks Workflows to write the information in the Delta table periodically to an external database such as MySQL and query the information from there as part of the agent logic / tool implementation

\* Exporting to an external database (e.g., MySQL) adds setup effort (workflow, external DB management) and latency (periodic updates vs. real-time). It's less performant and more complex than using Databricks' native tools.

\* Databricks Reference: "Avoid external systems when Delta tables provide real-time data natively" ("Databricks Workflows Guide").

Conclusion: Option A minimizes effort by using Databricks Feature Serving for real-time, low-latency access to the Delta table, ensuring high performance in a production-ready RAG chatbot.

## 質問 #71

A Generative AI Engineer is building an LLM to generate article summaries in the form of a type of poem, such as a haiku, given the article content. However, the initial output from the LLM does not match the desired tone or style.

Which approach will NOT improve the LLM's response to achieve the desired response?

- A. Provide the LLM with a prompt that explicitly instructs it to generate text in the desired tone and style
- B. Fine-tune the LLM on a dataset of desired tone and style
- C. Include few-shot examples in the prompt to the LLM
- D. Use a neutralizer to normalize the tone and style of the underlying documents

正解: D

解説:

The task at hand is to improve the LLM's ability to generate poem-like article summaries with the desired tone and style. Using a neutralizer to normalize the tone and style of the underlying documents (option B) will not help improve the LLM's ability to generate the desired poetic style. Here's why:

Neutralizing Underlying Documents:

A neutralizer aims to reduce or standardize the tone of input data. However, this contradicts the goal, which is to generate text with a specific tone and style (like haikus). Neutralizing the source documents will strip away the richness of the content, making it harder for the LLM to generate creative, stylistic outputs like poems.

Why Other Options Improve Results:

A (Explicit Instructions in the Prompt): Directly instructing the LLM to generate text in a specific tone and style helps align the output with the desired format (e.g., haikus). This is a common and effective technique in prompt engineering.

C (Few-shot Examples): Providing examples of the desired output format helps the LLM understand the expected tone and structure, making it easier to generate similar outputs.

D (Fine-tuning the LLM): Fine-tuning the model on a dataset that contains examples of the desired tone and style is a powerful way to improve the model's ability to generate outputs that match the target format.

Therefore, using a neutralizer (option B) is not an effective method for achieving the goal of generating stylized poetic summaries.

### 質問 # 72

A Generative AI Engineer is building a system that will answer questions on currently unfolding news topics.

As such, it pulls information from a variety of sources including articles and social media posts. They are concerned about toxic posts on social media causing toxic outputs from their system.

Which guardrail will limit toxic outputs?

- A. Log all LLM system responses and perform a batch toxicity analysis monthly.
- **B. Use only approved social media and news accounts to prevent unexpected toxic data from getting to the LLM.**
- C. Reduce the amount of context items the system will include in consideration for its response.
- D. Implement rate limiting

正解: B

解説:

The system answers questions on unfolding news topics using articles and social media, with a concern about toxic outputs from toxic inputs. A guardrail must limit toxicity in the LLM's responses. Let's evaluate the options.

\* Option A: Use only approved social media and news accounts to prevent unexpected toxic data from getting to the LLM

\* Curating input sources (e.g., verified accounts) reduces exposure to toxic content at the data ingestion stage, directly limiting toxic outputs. This is a proactive guardrail aligned with data quality control.

\* Databricks Reference: "Control input data quality to mitigate unwanted LLM behavior, such as toxicity" ("Building LLM Applications with Databricks," 2023).

\* Option B: Implement rate limiting

\* Rate limiting controls request frequency, not content quality. It prevents overload but doesn't address toxicity in social media inputs or outputs.

\* Databricks Reference: Rate limiting is for performance, not safety: "Use rate limits to manage compute load" ("Generative AI Cookbook").

\* Option C: Reduce the amount of context items the system will include in consideration for its response

\* Reducing context might limit exposure to some toxic items but risks losing relevant information, and it doesn't specifically target toxicity. It's an indirect, imprecise fix.

\* Databricks Reference: Context reduction is for efficiency, not safety: "Adjust context size based on performance needs" ("Databricks Generative AI Engineer Guide").

\* Option D: Log all LLM system responses and perform a batch toxicity analysis monthly

\* Logging and analyzing responses is reactive, identifying toxicity after it occurs rather than preventing it. Monthly analysis doesn't limit real-time toxic outputs.

\* Databricks Reference: Monitoring is for auditing, not prevention: "Log outputs for post-hoc analysis, but use input filters for safety" ("Building LLM-Powered Applications").

Conclusion: Option A is the most effective guardrail, proactively filtering toxic inputs from unverified sources, which aligns with Databricks' emphasis on data quality as a primary safety mechanism for LLM systems.

### 質問 # 73

A Generative AI Engineer is developing a chatbot designed to assist users with insurance-related queries. The chatbot is built on a large language model (LLM) and is conversational. However, to maintain the chatbot's focus and to comply with company policy, it must not provide responses to questions about politics. Instead, when presented with political inquiries, the chatbot should respond with a standard message:

"Sorry, I cannot answer that. I am a chatbot that can only answer questions around insurance." Which framework type should be implemented to solve this?

- **A. Safety Guardrail**
- B. Contextual Guardrail
- C. Compliance Guardrail
- D. Security Guardrail

正解: A

解説:

In this scenario, the chatbot must avoid answering political questions and instead provide a standard message for such inquiries. Implementing a Safety Guardrail is the appropriate solution for this:

What is a Safety Guardrail?

Safety guardrails are mechanisms implemented in Generative AI systems to ensure the model behaves within specific bounds. In this case, it ensures the chatbot does not answer politically sensitive or irrelevant questions, which aligns with the business rules.

Preventing Responses to Political Questions:

The Safety Guardrail is programmed to detect specific types of inquiries (like political questions) and prevent the model from generating responses outside its intended domain. When such queries are detected, the guardrail intervenes and provides a pre-defined response: "Sorry, I cannot answer that. I am a chatbot that can only answer questions around insurance." How It Works in Practice:

The LLM system can include a classification layer or trigger rules based on specific keywords related to politics. When such terms are detected, the Safety Guardrail blocks the normal generation flow and responds with the fixed message.

Why Other Options Are Less Suitable:

B (Security Guardrail): This is more focused on protecting the system from security vulnerabilities or data breaches, not controlling the conversational focus.

C (Contextual Guardrail): While context guardrails can limit responses based on context, safety guardrails are specifically about ensuring the chatbot stays within a safe conversational scope.

D (Compliance Guardrail): Compliance guardrails are often related to legal and regulatory adherence, which is not directly relevant here.

Therefore, a Safety Guardrail is the right framework to ensure the chatbot only answers insurance-related queries and avoids political discussions.

#### 質問 # 74

A team wants to serve a code generation model as an assistant for their software developers. It should support multiple programming languages. Quality is the primary objective.

Which of the Databricks Foundation Model APIs, or models available in the Marketplace, would be the best fit?

- A. BGE-large
- B. Llama2-70b
- C. MPT-7b
- **D. CodeLlama-34B**

正解: D

解説:

For a code generation model that supports multiple programming languages and where quality is the primary objective, CodeLlama-34B is the most suitable choice. Here's the reasoning:

Specialization in Code Generation:

CodeLlama-34B is specifically designed for code generation tasks. This model has been trained with a focus on understanding and generating code, which makes it particularly adept at handling various programming languages and coding contexts.

Capacity and Performance:

The "34B" indicates a model size of 34 billion parameters, suggesting a high capacity for handling complex tasks and generating high-quality outputs. The large model size typically correlates with better understanding and generation capabilities in diverse scenarios.

Suitability for Development Teams:

Given that the model is optimized for code, it will be able to assist software developers more effectively than general-purpose models. It understands coding syntax, semantics, and the nuances of different programming languages.

Why Other Options Are Less Suitable:

A (Llama2-70b): While also a large model, it's more general-purpose and may not be as fine-tuned for code generation as CodeLlama.

B (BGE-large): This model may not specifically focus on code generation.

C (MPT-7b): Smaller than CodeLlama-34B and likely less capable in handling complex code generation tasks at high quality.

Therefore, for a high-quality, multi-language code generation application, CodeLlama-34B (option D) is the best fit.

#### 質問 # 75

.....

ユーザーのオフライン読書を促進するために、Databricks-Generative-AI-Engineer-Associateスタディブレインダンプは、特にユーザー向けのPDFモードを開発するために、破片の時間をより有効に活用して学習できます。この

