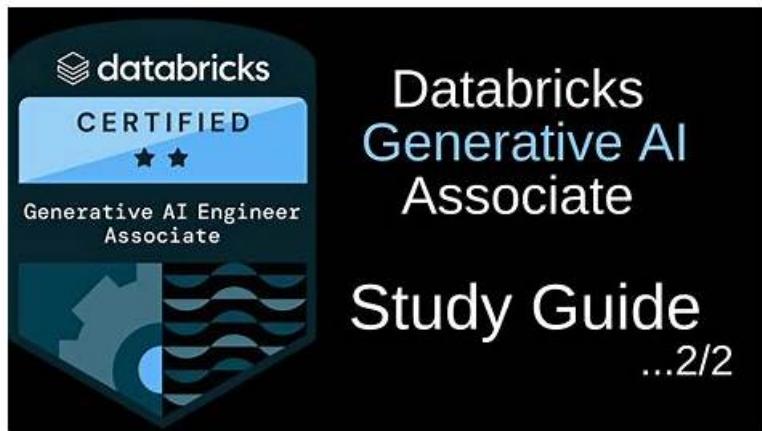


ユニークなDatabricks-Generative-AI-Engineer-Associate最新問題試験-試験の準備方法-信頼的なDatabricks-Generative-AI-Engineer-Associate日本語版参考書



BONUS! ! ! CertShiken Databricks-Generative-AI-Engineer-Associateダンプの一部を無料でダウンロード：https://drive.google.com/open?id=1XCQLf149t4FKgrGFPihL202_Q3MiwF2V

あなたに相応しいCertShiken問題集を探していますか。Databricks-Generative-AI-Engineer-Associate試験備考資料の整理を悩んでいますか。専業化のIT認定試験資料提供者CertShikenとして、かねてより全面的な資料を準備します。あなたの資料を探す時間を節約し、Databricks Databricks-Generative-AI-Engineer-Associate試験の復習をやっています。

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">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.
トピック 2	<ul style="list-style-type: none">Application Development: In this topic, Generative AI Engineers learn about tools needed to extract data, Langchainsimilar tools, and assessing responses to identify common issues. Moreover, the topic includes questions about adjusting an LLM's response, LLM guardrails, and the best LLM based on the attributes of the application.
トピック 3	<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.

>> Databricks-Generative-AI-Engineer-Associate最新問題 <<

最も有効なDatabricks-Generative-AI-Engineer-Associate最新問題だけが、Databricks Certified Generative AI Engineer Associateに合格の見込みを示すことができます

何の努力と時間もなくDatabricksのDatabricks-Generative-AI-Engineer-Associate試験に合格するのは不可能です。しかし、我々CertShikenチームは力を尽くしてあなたのDatabricksのDatabricks-Generative-AI-Engineer-Associate試験

を準備する圧力を減少して規範的な模擬問題と理解しやすい解答分析はあなたにDatabricksのDatabricks-Generative-AI-Engineer-Associate試験に合格するコツを把握させます。試験に失敗したら、あなたのDatabricksのDatabricks-Generative-AI-Engineer-Associate試験の成績書を提供して確認してから我々はすべての費用をあなたに払い戻します。CertShikenはあなたの信頼を得る足ります。

Databricks Certified Generative AI Engineer Associate 認定 Databricks-Generative-AI-Engineer-Associate 試験問題 (Q21-Q26):

質問 # 21

A Generative AI Engineer received the following business requirements for an external chatbot.

The chatbot needs to know what types of questions the user asks and routes to appropriate models to answer the questions. For example, the user might ask about upcoming event details. Another user might ask about purchasing tickets for a particular event. What is an ideal workflow for such a chatbot?

- A. The chatbot should be implemented as a multi-step LLM workflow. First, identify the type of question asked, then route the question to the appropriate model. If it's an upcoming event question, send the query to a text-to-SQL model. If it's about ticket purchasing, the customer should be redirected to a payment platform
- B. There should be two different chatbots handling different types of user queries.
- C. The chatbot should only look at previous event information
- D. The chatbot should only process payments

正解: A

解説:

* Problem Context: The chatbot must handle various types of queries and intelligently route them to the appropriate responses or systems.

* Explanation of Options:

* Option A: Limiting the chatbot to only previous event information restricts its utility and does not meet the broader business requirements.

* Option B: Having two separate chatbots could unnecessarily complicate user interaction and increase maintenance overhead.

* Option C: Implementing a multi-step workflow where the chatbot first identifies the type of question and then routes it accordingly is the most efficient and scalable solution. This approach allows the chatbot to handle a variety of queries dynamically, improving user experience and operational efficiency.

* Option D: Focusing solely on payments would not satisfy all the specified user interaction needs, such as inquiring about event details.

Option C offers a comprehensive workflow that maximizes the chatbot's utility and responsiveness to different user needs, aligning perfectly with the business requirements.

質問 # 22

A Generative AI Engineer is ready to deploy an LLM application written using Foundation Model APIs. They want to follow security best practices for production scenarios. Which authentication method should they choose?

- A. Use OAuth machine-to-machine authentication
- B. Use an access token belonging to any workspace user
- C. Use an access token belonging to service principals
- D. Use a frequently rotated access token belonging to either a workspace user or a service principal

正解: C

解説:

The task is to deploy an LLM application using Foundation Model APIs in a production environment while adhering to security best practices. Authentication is critical for securing access to Databricks resources, such as the Foundation Model API. Let's evaluate the options based on Databricks' security guidelines for production scenarios.

* Option A: Use an access token belonging to service principals

* Service principals are non-human identities designed for automated workflows and applications in Databricks. Using an access token tied to a service principal ensures that the authentication is scoped to the application, follows least-privilege principles (via role-based access control), and avoids reliance on individual user credentials. This is a security best practice for production deployments.

* Databricks Reference: "For production applications, use service principals with access tokens to authenticate securely, avoiding user-specific credentials" ("Databricks Security Best Practices,"

2023). Additionally, the "Foundation Model API Documentation" states: "Service principal tokens are recommended for programmatic access to Foundation Model APIs."

* Option B: Use a frequently rotated access token belonging to either a workspace user or a service principal

* Frequent rotation enhances security by limiting token exposure, but tying the token to a workspace user introduces risks (e.g., user account changes, broader permissions). Including both user and service principal options dilutes the focus on application-specific security, making this less ideal than a service-principal-only approach. It also adds operational overhead without clear benefits over Option A.

* Databricks Reference: "While token rotation is a good practice, service principals are preferred over user accounts for application authentication" ("Managing Tokens in Databricks," 2023).

* Option C: Use OAuth machine-to-machine authentication

* OAuth M2M (e.g., client credentials flow) is a secure method for application-to-service communication, often using service principals under the hood. However, Databricks' Foundation Model API primarily supports personal access tokens (PATs) or service principal tokens over full OAuth flows for simplicity in production setups. OAuth M2M adds complexity (e.g., managing refresh tokens) without a clear advantage in this context.

* Databricks Reference: "OAuth is supported in Databricks, but service principal tokens are simpler and sufficient for most API-based workloads" ("Databricks Authentication Guide," 2023).

* Option D: Use an access token belonging to any workspace user

* Using a user's access token ties the application to an individual's identity, violating security best practices. It risks exposure if the user leaves, changes roles, or has overly broad permissions, and it's not scalable or auditable for production.

* Databricks Reference: "Avoid using personal user tokens for production applications due to security and governance concerns" ("Databricks Security Best Practices," 2023).

Conclusion: Option A is the best choice, as it uses a service principal's access token, aligning with Databricks' security best practices for production LLM applications. It ensures secure, application-specific authentication with minimal complexity, as explicitly recommended for Foundation Model API deployments.

質問 # 23

A Generative AI Engineer has created a RAG application which can help employees retrieve answers from an internal knowledge base, such as Confluence pages or Google Drive. The prototype application is now working with some positive feedback from internal company testers. Now the Generative AI Engineer wants to formally evaluate the system's performance and understand where to focus their efforts to further improve the system.

How should the Generative AI Engineer evaluate the system?

- A. Curate a dataset that can test the retrieval and generation components of the system separately. Use MLflow's built in evaluation metrics to perform the evaluation on the retrieval and generation components.
- B. Use an LLM-as-a-judge to evaluate the quality of the final answers generated.
- C. Use cosine similarity score to comprehensively evaluate the quality of the final generated answers.
- D. Benchmark multiple LLMs with the same data and pick the best LLM for the job.

正解: A

解説:

* Problem Context: After receiving positive feedback for the RAG application prototype, the next step is to formally evaluate the system to pinpoint areas for improvement.

* Explanation of Options:

* Option A: While cosine similarity scores are useful, they primarily measure similarity rather than the overall performance of an RAG system.

* Option B: This option provides a systematic approach to evaluation by testing both retrieval and generation components separately. This allows for targeted improvements and a clear understanding of each component's performance, using MLflow's metrics for a structured and standardized assessment.

* Option C: Benchmarking multiple LLMs does not focus on evaluating the existing system's components but rather on comparing different models.

* Option D: Using an LLM as a judge is subjective and less reliable for systematic performance evaluation.

Option B is the most comprehensive and structured approach, facilitating precise evaluations and improvements on specific components of the RAG system.

質問 # 24

A Generative AI Engineer is developing a patient-facing healthcare-focused chatbot. If the patient's question is not a medical emergency, the chatbot should solicit more information from the patient to pass to the doctor's office and suggest a few relevant

pre-approved medical articles for reading. If the patient's question is urgent, direct the patient to calling their local emergency services.

Given the following user input:

"I have been experiencing severe headaches and dizziness for the past two days." Which response is most appropriate for the chatbot to generate?

- A. Headaches can be tough. Hope you feel better soon!
- B. Please provide your age, recent activities, and any other symptoms you have noticed along with your headaches and dizziness.
- C. Here are a few relevant articles for your browsing. Let me know if you have questions after reading them.
- D. Please call your local emergency services.

正解: D

解説:

* Problem Context: The task is to design responses for a healthcare-focused chatbot that appropriately addresses the urgency of a patient's symptoms.

* Explanation of Options:

* Option A: Suggesting articles might be suitable for less urgent inquiries but is inappropriate for symptoms that could indicate a serious condition.

* Option B: Given the description of severe symptoms like headaches and dizziness, directing the patient to emergency services is prudent. This aligns with medical guidelines that recommend immediate professional attention for such severe symptoms.

* Option C: Offering well-wishes does not address the potential seriousness of the symptoms and lacks appropriate action.

* Option D: While gathering more information is part of a detailed assessment, the immediate need here suggests a more urgent response.

Given the potential severity of the described symptoms, Option D is the most appropriate, ensuring the chatbot directs patients to seek urgent care when needed, potentially saving lives.

質問 # 25

A Generative AI Engineer is tasked with improving the RAG quality by addressing its inflammatory outputs.

Which action would be most effective in mitigating the problem of offensive text outputs?

- A. Restrict access to the data sources to a limited number of users
- B. Curate upstream data properly that includes manual review before it is fed into the RAG system
- C. Increase the frequency of upstream data updates
- D. Inform the user of the expected RAG behavior

正解: B

解説:

Addressing offensive or inflammatory outputs in a Retrieval-Augmented Generation (RAG) system is critical for improving user experience and ensuring ethical AI deployment. Here's why D is the most effective approach:

* Manual data curation: The root cause of offensive outputs often comes from the underlying data used to train the model or populate the retrieval system. By manually curating the upstream data and conducting thorough reviews before the data is fed into the RAG system, the engineer can filter out harmful, offensive, or inappropriate content.

* Improving data quality: Curating data ensures the system retrieves and generates responses from a high-quality, well-vetted dataset. This directly impacts the relevance and appropriateness of the outputs from the RAG system, preventing inflammatory content from being included in responses.

* Effectiveness: This strategy directly tackles the problem at its source (the data) rather than just mitigating the consequences (such as informing users or restricting access). It ensures that the system consistently provides non-offensive, relevant information.

Other options, such as increasing the frequency of data updates or informing users about behavior expectations, may not directly mitigate the generation of inflammatory outputs.

質問 # 26

.....

時間は何もありません。タイミングが全てだ。しないでください。Databricks-Generative-AI-Engineer-Associate VCEダンプは、試験をクリアする時間を節約するのに役立ちます。有効な試験ファイルを選択した場合、試験は一発で合格します。Databricks VCEダンプで最短時間で認定資格を取得できます。今すぐ上級職に就くと、

他の人よりも絶対に有利になります。 これで、時間を無駄にせずに、Databricks-Generative-AI-Engineer-Associate VCEダンプから始めてください。 優れた有効なVCEダンプは、あなたの夢を実現し、他の仲間よりも先に人生のピークを迎えます。

Databricks-Generative-AI-Engineer-Associate日本語版参考書: <https://www.certshiken.com/Databricks-Generative-AI-Engineer-Associate-shiken.html>

BONUS！！！ CertShiken Databricks-Generative-AI-Engineer-Associateダンプの一部を無料でダウンロード：https://drive.google.com/open?id=1XCQLfI49t4FKgrGFPihL202_Q3MiwF2V