

Databricks

Exam Questions Databricks-Generative-AI-Engineer-Associate

Databricks Certified Generative AI Engineer Associate



NEW QUESTION 1

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. Use a neutralizer to normalize the tone and style of the underlying documents
- C. Include few-shot examples in the prompt to the LLM
- D. Fine-tune the LLM on a dataset of desired tone and style

Answer: B

NEW QUESTION 2

A Generative AI Engineer is creating an LLM-powered application that will need access to up-to-date news articles and stock prices.

The design requires the use of stock prices which are stored in Delta tables and finding the latest relevant news articles by searching the internet.

How should the Generative AI Engineer architect their LLM system?

- A. Use an LLM to summarize the latest news articles and lookup stock tickers from the summaries to find stock prices.
- B. Query the Delta table for volatile stock prices and use an LLM to generate a search query to investigate potential causes of the stock volatility.
- C. Download and store news articles and stock price information in a vector store.
- D. Use a RAG architecture to retrieve and generate at runtime.
- E. Create an agent with tools for SQL querying of Delta tables and web searching, provide retrieved values to an LLM for generation of response.

Answer: D

NEW QUESTION 3

A Generative AI Engineer is tasked with developing an application that is based on an open source large language model (LLM). They need a foundation LLM with a large context window.

Which model fits this need?

- A. DistilBERT
- B. MPT-30B
- C. Llama2-70B
- D. DBRX

Answer: C

NEW QUESTION 4

A Generative AI Engineer needs to design an LLM pipeline to conduct multi-stage reasoning that leverages external tools. To be effective at this, the LLM will need to plan and adapt actions while performing complex reasoning tasks.

Which approach will do this?

- A. Train the LLM to generate a single, comprehensive response without interacting with any external tools, relying solely on its pre-trained knowledge.
- B. Implement a framework like ReAct which allows the LLM to generate reasoning traces and perform task-specific actions that leverage external tools if necessary.
- C. Encourage the LLM to make multiple API calls in sequence without planning or structuring the calls, allowing the LLM to decide when and how to use external tools spontaneously.
- D. Use a Chain-of-Thought (CoT) prompting technique to guide the LLM through a series of reasoning steps, then manually input the results from external tools for the final answer.

Answer: B

NEW QUESTION 5

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

NEW QUESTION 6

A Generative AI Engineer interfaces with an LLM with prompt/response behavior that has been trained on customer calls inquiring about product availability. The LLM is designed to output "In Stock" if the product is available or only the term "Out of Stock" if not.

Which prompt will work to allow the engineer to respond to call classification labels correctly?

- A. Respond with "In Stock" if the customer asks for a product.
- B. You will be given a customer call transcript where the customer asks about product availability.
- C. The outputs are either "In Stock" or "Out of Stock". Format the output in JSON, for example: {"call_id": "123", "label": "In Stock"}.
- D. Respond with "Out of Stock" if the customer asks for a product.
- E. You will be given a customer call transcript where the customer inquires about product availability.
- F. Respond with "In Stock" if the product is available or "Out of Stock" if not.

Answer: B

NEW QUESTION 7

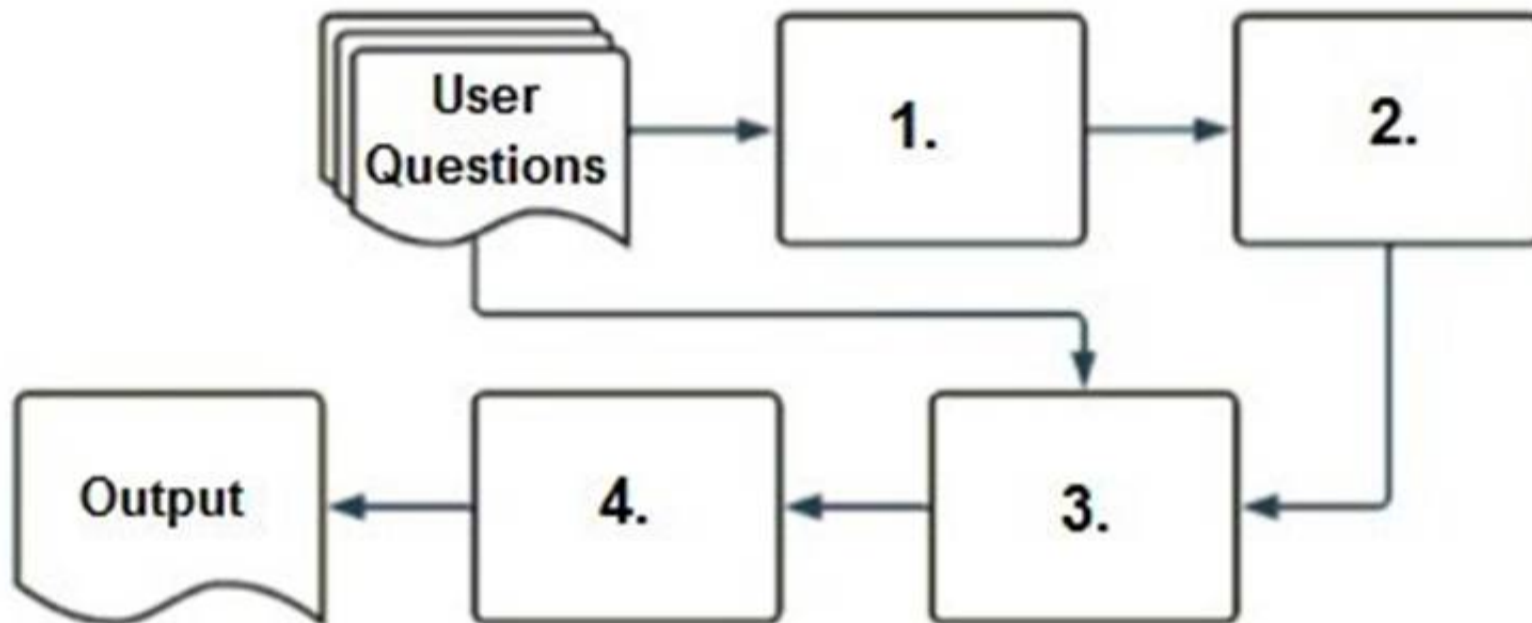
A small and cost-conscious startup in the cancer research field wants to build a RAG application using Foundation Model APIs. Which strategy would allow the startup to build a good-quality RAG application while being cost-conscious and able to cater to customer needs?

- A. Limit the number of relevant documents available for the RAG application to retrieve from
- B. Pick a smaller LLM that is domain-specific
- C. Limit the number of queries a customer can send per day
- D. Use the largest LLM possible because that gives the best performance for any general queries

Answer: B

NEW QUESTION 8

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

Answer: A

NEW QUESTION 9

A Generative AI Engineer is working with a retail company that wants to enhance its customer experience by automatically handling common customer inquiries. They are working on an LLM-powered AI solution that should improve response times while maintaining a personalized interaction. They want to define the appropriate input and LLM task to do this. Which input/output pair will do this?

- A. Input: Customer reviews; Output Group the reviews by users and aggregate per-user average rating, then respond
- B. Input: Customer service chat logs; Output Group the chat logs by users, followed by summarizing each user's interactions, then respond
- C. Input: Customer service chat logs; Output: Find the answers to similar questions and respond with a summary
- D. Input: Customer reviews; Output Classify review sentiment

Answer: C

NEW QUESTION 10

A Generative AI Engineer has already trained an LLM on Databricks and it is now ready to be deployed. Which of the following steps correctly outlines the easiest process for deploying a model on Databricks?

- A. Log the model as a pickle object, upload the object to Unity Catalog Volume, register it to Unity Catalog using MLflow, and start a serving endpoint
- B. Log the model using MLflow during training, directly register the model to Unity Catalog using the MLflow API, and start a serving endpoint
- C. Save the model along with its dependencies in a local directory, build the Docker image, and run the Docker container
- D. Wrap the LLM's prediction function into a Flask application and serve using Gunicorn

Answer: B

NEW QUESTION 10

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. Here are a few relevant articles for your browsin
- B. Let me know if you have questions after reading them.
- C. Please call your local emergency services.

- D. Headaches can be tough
- E. Hope you feel better soon!
- F. Please provide your age, recent activities, and any other symptoms you have noticed along with your headaches and dizziness.

Answer: B

NEW QUESTION 11

When developing an LLM application, it's crucial to ensure that the data used for training the model complies with licensing requirements to avoid legal risks. Which action is NOT appropriate to avoid legal risks?

- A. Reach out to the data curators directly before you have started using the trained model to let them know.
- B. Use any available data you personally created which is completely original and you can decide what license to use.
- C. Only use data explicitly labeled with an open license and ensure the license terms are followed.
- D. Reach out to the data curators directly after you have started using the trained model to let them know.

Answer: D

NEW QUESTION 16

A Generative AI Engineer is designing a chatbot for a gaming company that aims to engage users on its platform while its users play online video games. Which metric would help them increase user engagement and retention for their platform?

- A. Randomness
- B. Diversity of responses
- C. Lack of relevance
- D. Repetition of responses

Answer: B

NEW QUESTION 17

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

Answer: B

NEW QUESTION 19

A Generative AI Engineer is creating an LLM system that will retrieve news articles from the year 1918 and related to a user's query and summarize them. The engineer has noticed that the summaries are generated well but often also include an explanation of how the summary was generated, which is undesirable. Which change could the Generative AI Engineer perform to mitigate this issue?

- A. Split the LLM output by newline characters to truncate away the summarization explanation.
- B. Tune the chunk size of news articles or experiment with different embedding models.
- C. Revisit their document ingestion logic, ensuring that the news articles are being ingested properly.
- D. Provide few shot examples of desired output format to the system and/or user prompt.

Answer: D

NEW QUESTION 20

A Generative AI Engineer just deployed an LLM application at a digital marketing company that assists with answering customer service inquiries. Which metric should they monitor for their customer service LLM application in production?

- A. Number of customer inquiries processed per unit of time
- B. Energy usage per query
- C. Final perplexity scores for the training of the model
- D. HuggingFace Leaderboard values for the base LLM

Answer: A

NEW QUESTION 22

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

Databricks-Generative-AI-Engineer-Associate Practice Exam Features:

- * Databricks-Generative-AI-Engineer-Associate Questions and Answers Updated Frequently
- * Databricks-Generative-AI-Engineer-Associate Practice Questions Verified by Expert Senior Certified Staff
- * Databricks-Generative-AI-Engineer-Associate Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * Databricks-Generative-AI-Engineer-Associate Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The Databricks-Generative-AI-Engineer-Associate Practice Test Here](#)