Databricks

Exam Questions Databricks-Generative-AI-Engineer-Associate

Databricks Certified Generative AI Engineer Associate

NEW QUESTION 1
A Generative Al 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. Increase the frequency of upstream data updates
B. Inform the user of the expected RAG behavior
C. Restrict access to the data sources to a limited number of users
D. Curate upstream data properly that includes manual review before it is fed into the RAG system

Answer: D

NEW QUESTION 2
A Generative AI Engineer is creating an agent-based LLM system for their favorite monster truck team. The system can answer text based questions about the monster truck team, lookup event dates via an API call, or query tables on the team??s latest standings.
How could the Generative AI Engineer best design these capabilities into their system?

A. Ingest PDF documents about the monster truck team into a vector store and query it in a RAG architecture.
B. Write a system prompt for the agent listing available tools and bundle it into an agent system that runs a number of calls to solve a query.
C. Instruct the LLM to respond with ??RAG??, ??API??, or ??TABLE?? depending on the query, then use text parsing and conditional statements to resolve the query.
D. Build a system prompt with all possible event dates and table information in the system promp
E. Use a RAG architecture to lookup generic text questions and otherwise leverage the information in the system prompt.

Answer: B

NEW QUESTION 3
A Generative Al Engineer is developing a RAG system for their company to perform internal document Q&A for structured HR policies, but the answers returned are frequently incomplete and unstructured It seems that the retriever is not returning all relevant context The Generative Al Engineer has experimented with different embedding and response generating LLMs but that did not improve results.
Which TWO options could be used to improve the response quality? Choose 2 answers

A. Add the section header as a prefix to chunks
B. Increase the document chunk size
C. Split the document by sentence
D. Use a larger embedding model
E. Fine tune the response generation model

Answer: AB

NEW QUESTION 4
A Generative Al Engineer is building an LLM-based application that has an important transcription (speech-to-text) task. Speed is essential for the success of the application
Which open Generative Al models should be used?

A. L!ama-2-70b-chat-hf
B. MPT-30B-Instruct
C. DBRX
D. whisper-large-v3 (1.6B)

Answer: D

NEW QUESTION 5
A Generative AI Engineer has been asked to design an LLM-based application that accomplishes the following business objective: answer employee HR questions using HR PDF documentation.
Which set of high level tasks should the Generative AI Engineer's system perform?

A. Calculate averaged embeddings for each HR document, compare embeddings to user query to find the best documen
B. Pass the best document with the user query into an LLM with a large context window to generate a response to the employee.
C. Use an LLM to summarize HR documentatio
D. Provide summaries of documentation and user query into an LLM with a large context window to generate a response to the user.
E. Create an interaction matrix of historical employee questions and HR documentatio
F. Use ALS to factorize the matrix and create embedding
G. Calculate the embeddings of new queries and use them to find the best HR documentatio
H. Use an LLM to generate a response to the employee question based upon the documentation retrieved.
I. Split HR documentation into chunks and embed into a vector stor
J. Use the employee question to retrieve best matched chunks of documentation, and use the LLM to generate a response to the employee based upon the documentation retrieved.

Answer: D

NEW QUESTION 6
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. Llama2-70b

B. BGE-large
C. MPT-7b
D. CodeLlama-34B

Answer: D

NEW QUESTION 7
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 availabilit
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 availabilit
F. Respond with ??In Stock?? if the product is available or ??Out of Stock?? if not.

Answer: B

NEW QUESTION 8
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 9
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 only look at previous event information
B. There should be two different chatbots handling different types of user queries.
C. The chatbot should be implemented as a multi-step LLM workflo
D. First, identify the type of question asked, then route the question to the appropriate mode
E. If it??s an upcoming event question, send the query to a text-to-SQL mode
F. If it??s about ticket purchasing, the customer should be redirected to a payment platform.
G. The chatbot should only process payments

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 tasked with deploying an application that takes advantage of a custom MLflow Pyfunc model to return some interim results.
How should they configure the endpoint to pass the secrets and credentials?

A. Use spark.conf.set ()
B. Pass variables using the Databricks Feature Store API
C. Add credentials using environment variables
D. Pass the secrets in plain text

Answer: C

NEW QUESTION 13
What is an effective method to preprocess prompts using custom code before sending them to an LLM?

A. Directly modify the LLM??s internal architecture to include preprocessing steps
B. It is better not to introduce custom code to preprocess prompts as the LLM has not been trained with examples of the preprocessed prompts
C. Rather than preprocessing prompts, it??s more effective to postprocess the LLM outputs to align the outputs to desired outcomes
D. Write a MLflow PyFunc model that has a separate function to process the prompts

Answer: D

NEW QUESTION 14
A Generative Al Engineer is deciding between using LSH (Locality Sensitive Hashing) and HNSW (Hierarchical Navigable Small World) for indexing their vector database Their top priority is semantic accuracy
Which approach should the Generative Al Engineer use to evaluate these two techniques?

A. Compare the cosine similarities of the embeddings of returned results against those of a representative sample of test inputs
B. Compare the Bilingual Evaluation Understudy (BLEU) scores of returned results for a representative sample of test inputs
C. Compare the Recall-Onented-Understudy for Gistmg Evaluation (ROUGE) scores of returned results for a representative sample of test inputs
D. Compare the Levenshtein distances of returned results against a representative sample of test inputs

Answer: A

NEW QUESTION 15
A Generative Al 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 trained on related domain knowledge
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 with multilingual support to support potential multilingual user questions

Answer: A

NEW QUESTION 19
.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons 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