Exam NCA-GENL questions and answers

BTW, DOWNLOAD part of DumpsFree NCA-GENL dumps from Cloud Storage: https://drive.google.com/open?id=1N2XkcPyLy5GKPikI5vhcYMXcjDR3-vSK

DumpsFree offers accurate and reliable study materials to help you prepare for the NVIDIA NCA-GENL Exam. They have prepared the best NVIDIA NCA-GENL Exam Questions that provide authentic and reliable material. With DumpsFree, many candidates have succeeded in passing the NVIDIA NCA-GENL Exam.

One of the most important functions of our APP online vesion which is contained in our NCA-GENL preparation questions are that can support almost all electronic equipment, including the computer, mobile phone and so on. If you want to prepare for your exam by the computer, you can buy our NCA-GENL training quiz, because our products can work well by the computer. Of course, if you prefer to study by your mobile phone, our NCA-GENL study materials also can meet your demand.

>> NCA-GENL Exam Dumps Pdf <<

Reliable NCA-GENL Exam Testking, Dumps NCA-GENL Torrent

With passing rate up to 98 to 100 percent, the quality and accuracy of our NCA-GENL training materials are unquestionable. You may wonder their price must be equally steep. While it is not truth. On the contrary everyone can afford them easily. By researching on the frequent-tested points in the real exam, our experts have made both clear outlines and comprehensive questions into our NCA-GENL Exam Prep. So our NCA-GENL practice engine is easy for you to understand.

NVIDIA NCA-GENL Exam Syllabus Topics:

Topic	Details

Topic 1	• Fundamentals of Machine Learning and Neural Networks: This section of the exam measures the skills of AI Researchers and covers the foundational principles behind machine learning and neural networks, focusing on how these concepts underpin the development of large language models (LLMs). It ensures the learner understands the basic structure and learning mechanisms involved in training generative AI systems.
Topic 2	• This section of the exam measures skills of AI Product Developers and covers how to strategically plan experiments that validate hypotheses, compare model variations, or test model responses. It focuses on structure, controls, and variables in experimentation.
Topic 3	• LLM Integration and Deployment: This section of the exam measures skills of AI Platform Engineers and covers connecting LLMs with applications or services through APIs, and deploying them securely and efficiently at scale. It also includes considerations for latency, cost, monitoring, and updates in production environments.
Topic 4	• Prompt Engineering: This section of the exam measures the skills of Prompt Designers and covers how to craft effective prompts that guide LLMs to produce desired outputs. It focuses on prompt strategies, formatting, and iterative refinement techniques used in both development and real-world applications of LLMs.
Topic 5	• Python Libraries for LLMs: This section of the exam measures skills of LLM Developers and covers using Python tools and frameworks like Hugging Face Transformers, LangChain, and PyTorch to build, fine-tune, and deploy large language models. It focuses on practical implementation and ecosystem familiarity.
Topic 6	• Experiment Design

NVIDIA Generative AI LLMs Sample Questions (Q29-Q34):

NEW QUESTION # 29
What metrics would you use to evaluate the performance of a RAG workflow in terms of the accuracy of responses generated in relation to the input query? (Choose two.)

- A. Retriever latency
- B. Response relevancy
- C. Generator latency
- D. Tokens generated per second
- E. Context precision

Answer: B,E

Explanation:
In a Retrieval-Augmented Generation (RAG) workflow, evaluating the accuracy of responses relative to the input query focuses on the quality of the retrieved context and the generated output. As covered in NVIDIA's Generative AI and LLMs course, two key metrics are response relevancy and context precision. Response relevancy measures how well the generated response aligns with the input query, often assessed through human evaluation or automated metrics like ROUGE or BLEU, ensuring the output is pertinent and accurate.
Context precision evaluates the retriever's ability to fetch relevant documents or passages from the knowledge base, typically measured by metrics like precision@k, which assesses the proportion of retrieved items that are relevant to the query. Options A (generator latency), B (retriever latency), and C (tokens generated per second) are incorrect, as they measure performance efficiency (speed) rather than accuracy. The course notes:
"In RAG workflows, response relevancy ensures the generated output matches the query intent, while context precision evaluates the accuracy of retrieved documents, critical for high-quality responses." References: NVIDIA Building Transformer-Based Natural Language Processing Applications course; NVIDIA Introduction to Transformer-Based Natural Language Processing.

NEW QUESTION # 30
When comparing and contrasting the ReLU and sigmoid activation functions, which statement is true?

- A. ReLU is more computationally efficient, but sigmoid is better for predicting probabilities.

- B. ReLU is a linear function while sigmoid is non-linear.
- C. ReLU is less computationally efficient than sigmoid, but it is more accurate than sigmoid.
- D. ReLU and sigmoid both have a range of 0 to 1.

Answer: A

Explanation:
ReLU (Rectified Linear Unit) and sigmoid are activation functions used in neural networks. According to NVIDIA's deep learning documentation (e.g., cuDNN and TensorRT), ReLU, defined as f(x) = max(0, x), is computationally efficient because it involves simple thresholding, avoiding expensive exponential calculations required by sigmoid, f(x) = 1/(1 + e