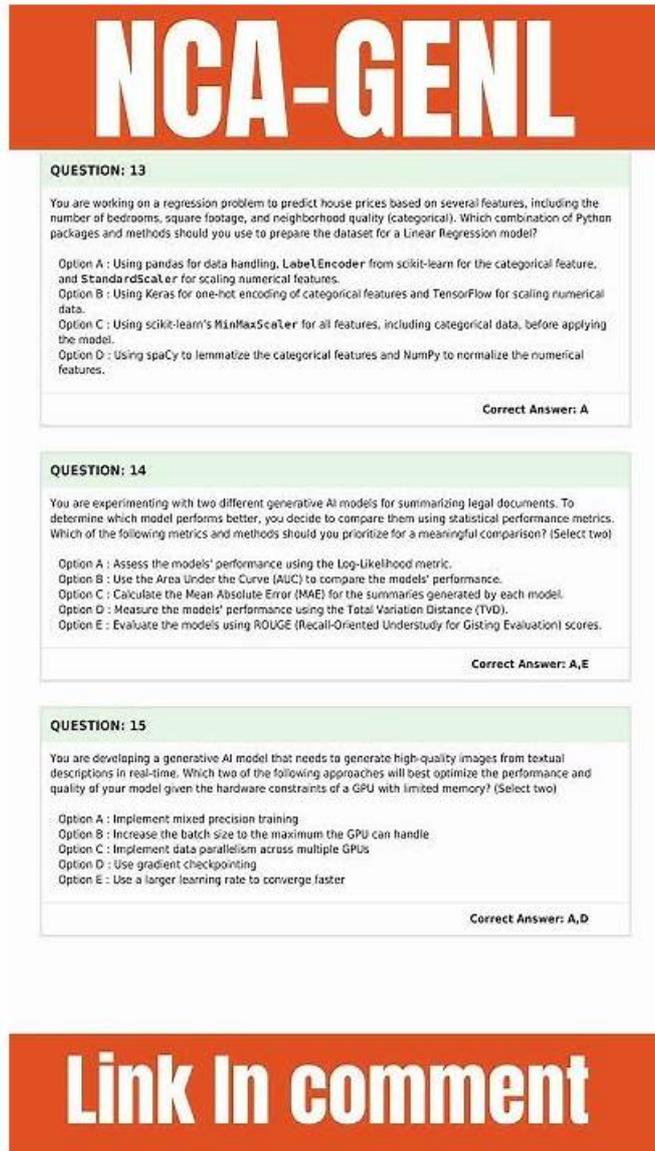


# Test NCA-GENL King - NCA-GENL Exam Dumps Demo



**NCA-GENL**

**QUESTION: 13**

You are working on a regression problem to predict house prices based on several features, including the number of bedrooms, square footage, and neighborhood quality (categorical). Which combination of Python packages and methods should you use to prepare the dataset for a Linear Regression model?

Option A : Using pandas for data handling, LabelEncoder from scikit-learn for the categorical feature, and StandardScaler for scaling numerical features.  
Option B : Using Keras for one-hot encoding of categorical features and TensorFlow for scaling numerical data.  
Option C : Using scikit-learn's MinMaxScaler for all features, including categorical data, before applying the model.  
Option D : Using spaCy to lemmatize the categorical features and NumPy to normalize the numerical features.

**Correct Answer: A**

**QUESTION: 14**

You are experimenting with two different generative AI models for summarizing legal documents. To determine which model performs better, you decide to compare them using statistical performance metrics. Which of the following metrics and methods should you prioritize for a meaningful comparison? (Select two)

Option A : Assess the models' performance using the Log-Likelihood metric.  
Option B : Use the Area Under the Curve (AUC) to compare the models' performance.  
Option C : Calculate the Mean Absolute Error (MAE) for the summaries generated by each model.  
Option D : Measure the models' performance using the Total Variation Distance (TVD).  
Option E : Evaluate the models using ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores.

**Correct Answer: A,E**

**QUESTION: 15**

You are developing a generative AI model that needs to generate high-quality images from textual descriptions in real-time. Which two of the following approaches will best optimize the performance and quality of your model given the hardware constraints of a GPU with limited memory? (Select two)

Option A : Implement mixed precision training  
Option B : Increase the batch size to the maximum the GPU can handle  
Option C : Implement data parallelism across multiple GPUs  
Option D : Use gradient checkpointing  
Option E : Use a larger learning rate to converge faster

**Correct Answer: A,D**

**Link In comment**

2025 Latest DumpTorrent NCA-GENL PDF Dumps and NCA-GENL Exam Engine Free Share: <https://drive.google.com/open?id=123IkvszH9sCPyIPmLJGcB6tpoyIRf44D>

Do you want to pass your NVIDIA Generative AI LLMs exam? If so, DumpTorrent is the ideal place to begin. DumpTorrent provides comprehensive NCA-GENL exam questions preparation in two simple formats: a pdf file format and an NVIDIA NCA-GENL online practice test engine. If you fail your NVIDIA Generative AI LLMs (NCA-GENL) Exam, you can obtain a full refund and a 20% discount! Continue reading to discover more about the essential aspects of these excellent NCA-GENL exam questions.

## NVIDIA NCA-GENL Exam Syllabus Topics:

Topic	Details

Topic 1	<ul style="list-style-type: none"> <li>• <b>LLM Integration and Deployment:</b> 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.</li> </ul>
Topic 2	<ul style="list-style-type: none"> <li>• <b>Experimentation:</b> This section of the exam measures the skills of ML Engineers and covers how to conduct structured experiments with LLMs. It involves setting up test cases, tracking performance metrics, and making informed decisions based on experimental outcomes.:</li> </ul>
Topic 3	<ul style="list-style-type: none"> <li>• <b>Software Development:</b> This section of the exam measures the skills of Machine Learning Developers and covers writing efficient, modular, and scalable code for AI applications. It includes software engineering principles, version control, testing, and documentation practices relevant to LLM-based development.</li> </ul>
Topic 4	<ul style="list-style-type: none"> <li>• <b>Fundamentals of Machine Learning and Neural Networks:</b> 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.</li> </ul>
Topic 5	<ul style="list-style-type: none"> <li>• <b>Prompt Engineering:</b> 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.</li> </ul>
Topic 6	<ul style="list-style-type: none"> <li>• 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.</li> </ul>
Topic 7	<ul style="list-style-type: none"> <li>• <b>Experiment Design</b></li> </ul>
Topic 8	<ul style="list-style-type: none"> <li>• <b>Data Preprocessing and Feature Engineering:</b> This section of the exam measures the skills of Data Engineers and covers preparing raw data into usable formats for model training or fine-tuning. It includes cleaning, normalizing, tokenizing, and feature extraction methods essential to building robust LLM pipelines.</li> </ul>
Topic 9	<ul style="list-style-type: none"> <li>• <b>Python Libraries for LLMs:</b> 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.</li> </ul>
Topic 10	<ul style="list-style-type: none"> <li>• <b>Alignment:</b> This section of the exam measures the skills of AI Policy Engineers and covers techniques to align LLM outputs with human intentions and values. It includes safety mechanisms, ethical safeguards, and tuning strategies to reduce harmful, biased, or inaccurate results from models.</li> </ul>

>> Test NCA-GENL King <<

## NCA-GENL Exam Dumps Demo | NCA-GENL Pass Exam

We can provide you with efficient online services during the whole day, no matter what kind of problems or consultants about our NCA-GENL quiz torrent; we will spare no effort to help you overcome them sooner or later. First of all, we have professional staff with dedication to check and update our NCA-GENL Exam Torrent materials on a daily basis, so that you can get the latest information from our NCA-GENL exam torrent at any time. Besides our after-sales service engineers will be always online to give remote guidance and assistance for you on NCA-GENL study questions if necessary.

## NVIDIA Generative AI LLMs Sample Questions (Q46-Q51):

### NEW QUESTION # 46

What are the main advantages of instructed large language models over traditional, small language models (< 300M parameters)? (Pick the 2 correct responses)

- A. Single generic model can do more than one task.
- B. Smaller latency, higher throughput.
- C. It is easier to explain the predictions.
- D. Cheaper computational costs during inference.
- E. Trained without the need for labeled data.

**Answer: A,D**

Explanation:

Instructed large language models (LLMs), such as those supported by NVIDIA's NeMo framework, have significant advantages over smaller, traditional models:

\* Option D: LLMs often have cheaper computational costs during inference for certain tasks because they can generalize across multiple tasks without requiring task-specific retraining, unlike smaller models that may need separate models per task.

References:

NVIDIA NeMo Documentation: <https://docs.nvidia.com/deeplearning/nemo/user-guide/docs/en/stable/nlp/intro.html> Brown, T., et al. (2020). "Language Models are Few-Shot Learners."

### NEW QUESTION # 47

You are working on developing an application to classify images of animals and need to train a neural model.

However, you have a limited amount of labeled data. Which technique can you use to leverage the knowledge from a model pre-trained on a different task to improve the performance of your new model?

- A. Early stopping
- B. Dropout
- C. Transfer learning
- D. Random initialization

**Answer: C**

Explanation:

Transfer learning is a technique where a model pre-trained on a large, general dataset (e.g., ImageNet for computer vision) is fine-tuned for a specific task with limited data. NVIDIA's Deep Learning AI documentation, particularly for frameworks like NeMo and TensorRT, emphasizes transfer learning as a powerful approach to improve model performance when labeled data is scarce. For example, a pre-trained convolutional neural network (CNN) can be fine-tuned for animal image classification by reusing its learned features (e.g., edge detection) and adapting the final layers to the new task. Option A (dropout) is a regularization technique, not a knowledge transfer method. Option B (random initialization) discards pre-trained knowledge. Option D (early stopping) prevents overfitting but does not leverage pre-trained models.

References:

NVIDIA NeMo Documentation: [https://docs.nvidia.com/deeplearning/nemo/user-guide/docs/en/stable/nlp/model\\_finetuning.html](https://docs.nvidia.com/deeplearning/nemo/user-guide/docs/en/stable/nlp/model_finetuning.html)

NVIDIA Deep Learning AI: <https://www.nvidia.com/en-us/deep-learning-ai/>

### NEW QUESTION # 48

What is the main difference between forward diffusion and reverse diffusion in diffusion models of Generative AI?

- A. Forward diffusion focuses on generating a sample from a given noise vector, while reverse diffusion reverses the process by estimating the latent space representation of a given sample.
- B. Forward diffusion uses bottom-up processing, while reverse diffusion uses top-down processing to generate samples from noise vectors.
- C. Forward diffusion focuses on progressively injecting noise into data, while reverse diffusion focuses on generating new samples from the given noise vectors.
- D. Forward diffusion uses feed-forward networks, while reverse diffusion uses recurrent networks.

**Answer: C**

Explanation:

Diffusion models, a class of generative AI models, operate in two phases: forward diffusion and reverse diffusion. According to NVIDIA's documentation on generative AI (e.g., in the context of NVIDIA's work on generative models), forward diffusion progressively injects noise into a data sample (e.g., an image or text embedding) over multiple steps, transforming it into a noise

distribution. Reverse diffusion, conversely, starts with a noise vector and iteratively denoises it to generate a new sample that resembles the training data distribution. This process is central to models like DDPM (Denoising Diffusion Probabilistic Models). Option A is incorrect, as forward diffusion adds noise, not generates samples. Option B is false, as diffusion models typically use convolutional or transformer-based architectures, not recurrent networks. Option C is misleading, as diffusion does not align with bottom-up/top-down processing paradigms.

References:

NVIDIA Generative AI Documentation: <https://www.nvidia.com/en-us/ai-data-science/generative-ai/> Ho, J., et al. (2020). "Denoising Diffusion Probabilistic Models."

#### NEW QUESTION # 49

In the transformer architecture, what is the purpose of positional encoding?

- A. To encode the importance of each token in the input sequence.
- B. To remove redundant information from the input sequence.
- C. To encode the semantic meaning of each token in the input sequence.
- **D. To add information about the order of each token in the input sequence.**

**Answer: D**

Explanation:

Positional encoding is a vital component of the Transformer architecture, as emphasized in NVIDIA's Generative AI and LLMs course. Transformers lack the inherent sequential processing of recurrent neural networks, so they rely on positional encoding to incorporate information about the order of tokens in the input sequence. This is typically achieved by adding fixed or learned vectors (e.g., sine and cosine functions) to the token embeddings, where each position in the sequence has a unique encoding. This allows the model to distinguish the relative or absolute positions of tokens, enabling it to understand word order in tasks like translation or text generation. For example, in the sentence "The cat sleeps," positional encoding ensures the model knows "cat" is the second token and "sleeps" is the third. Option A is incorrect, as positional encoding does not remove information but adds positional context. Option B is wrong because semantic meaning is captured by token embeddings, not positional encoding. Option D is also inaccurate, as the importance of tokens is determined by the attention mechanism, not positional encoding. The course notes: "Positional encodings are used in Transformers to provide information about the order of tokens in the input sequence, enabling the model to process sequences effectively." References: NVIDIA Building Transformer-Based Natural Language Processing Applications course; NVIDIA Introduction to Transformer-Based Natural Language Processing.

#### NEW QUESTION # 50

In the context of a natural language processing (NLP) application, which approach is most effective for implementing zero-shot learning to classify text data into categories that were not seen during training?

- A. Use rule-based systems to manually define the characteristics of each category.
- B. Use a large, labeled dataset for each possible category.
- **C. Use a pre-trained language model with semantic embeddings.**
- D. Train the new model from scratch for each new category encountered.

**Answer: C**

Explanation:

Zero-shot learning allows models to perform tasks or classify data into categories without prior training on those specific categories. In NLP, pre-trained language models (e.g., BERT, GPT) with semantic embeddings are highly effective for zero-shot learning because they encode general linguistic knowledge and can generalize to new tasks by leveraging semantic similarity. NVIDIA's NeMo documentation on NLP tasks explains that pre-trained LLMs can perform zero-shot classification by using prompts or embeddings to map input text to unseen categories, often via techniques like natural language inference or cosine similarity in embedding space. Option A (rule-based systems) lacks scalability and flexibility. Option B contradicts zero-shot learning, as it requires labeled data. Option C (training from scratch) is impractical and defeats the purpose of zero-shot learning.

References:

NVIDIA NeMo Documentation: <https://docs.nvidia.com/deeplearning/nemo/user-guide/docs/en/stable/nlp/intro.html>

Brown, T., et al. (2020). "Language Models are Few-Shot Learners."

