Q1)

Domain generalization in Large Language Models (LLMs) depends on model architecture, training data, and fine-tuning techniques. Transformer-based architectures, with self-attention mechanisms, enable LLMs to capture contextual patterns across diverse domains. Training on large, diverse datasets improves generalization by exposing models to varied linguistic patterns, allowing them to adapt better to unseen domains. Fine-tuning techniques, such as few-shot learning and domain-specific adjustments, enhance performance in new tasks while preserving general capabilities. Continual learning can help maintain broad knowledge across multiple domains. Together, these factors ensure LLMs can perform well even in domains not encountered during pretraining.

Q2)

Mode collapse in Generative Adversarial Networks (GANs) occurs when the generator repeatedly produces similar outputs, ignoring parts of the data distribution. Instead of generating diverse results, it focuses on a few patterns or "modes."

Symptoms:

- The generator outputs highly similar or identical results, regardless of varying inputs.
- The model is unable to capture the complete diversity present in the training data.

Impact:

- **Quality**: While individual outputs may still look realistic, the lack of variety diminishes the overall effectiveness.
- **Diversity**: Mode collapse prevents GANs from generating a broad range of outputs, limiting their utility in tasks where diversity and representation of various data patterns are essential.

Q3)

Variational Autoencoders (VAEs) differ from standard autoencoders in how they handle latent space and generative abilities.

Latent Space:

- **Standard Autoencoders**: Map inputs to fixed, deterministic points in latent space, limiting their generative potential.
- **VAEs**: Encode inputs into a probabilistic distribution, allowing sampling from the latent space, enabling the generation of new data.

Generative Abilities:

- **Standard Autoencoders**: Focus on reconstructing inputs and lack robust generative capabilities.
- **VAEs**: Designed as generative models, VAEs can produce new data by sampling from the learned latent distribution, offering more flexibility for data generation.

VAEs provide a more powerful framework for generating diverse and realistic samples due to their probabilistic approach.

Q4)

The runtime complexity of transformer-based text encoder models scales quadratically with input sequence length, specifically **$O(n^2)$**, where **n** is the sequence length. This complexity arises from the self-attention mechanism, which computes attention scores between every pair of tokens in the sequence, resulting in **$n \times n$** interactions.

Implications for Long Sequences:

As sequence length increases, computational and memory requirements grow significantly, making transformers inefficient for processing long sequences. This leads to slower training and inference times, limiting their scalability. Techniques like sparse attention or efficient transformer variants (e.g., Longformer) are often used to mitigate this issue by reducing the number of token interactions and improving performance on long sequences.