U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: VII**
Branch: Artificial Intelligence & Machine Learning **Duration: 3 hrs.**
Course Code: 24AM7PCGAL **Max Marks: 100**
Course: Generative AI with Large Language Models

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

		UNIT - I	*CO*	*PO*	**Marks**
1	a)	Compare traditional autoencoders and Variational Autoencoder's (VAEs) in terms of latent space representation and data generation ability.	*CO1*	*PO2*	**06**
	b)	Explain the benefits of using Generative AI in Applications.	*CO1*	*PO1*	**06**
	c)	Given: Input (X): [0.8, 0.7, 0.6, 0.9] Encoder Weights (W_e): [[0.2, 0.1], [0.4, 0.3], [0.1, 0.2], [0.3, 0.4]] Encoder Biases (b_e): [0.05, 0.1] Decoder Weights (W_d): [[0.3, 0.2, 0.1, 0.4], [0.4, 0.3, 0.2, 0.1]] Decoder Biases (b_d): [0.1, 0.05, 0.2, 0.15] Illustrate the encoder-decoder process of an Autoencoder by i. Constructing a latent space representation Z of the given input vector X. ii. Reconstructing the input vector 'X' using results of (i).	*CO2*	*PO3*	**08**
		OR			
2	a)	Given mean(x)=5, mean(y)=6.1, stddev(x)=1.2, stddev(y)=0.9, random sample for x: ϵ_x=0.3, random sample for y: ϵ_y=0.5. Calculate latent space variable (z) and the reconstructed data points ($\hat{x}$, $\hat{y}$) of a VAE model.	*C 2*	*PO3*	**08**
	b)	For a VAE with latent variable $z \sim N(\mu,\sigma^2)$, derive the gradient of the reparameterization trick $z = \mu + \sigma \cdot \epsilon$, where $\epsilon \sim N(0,1)$. Explain its significance in training.	*CO1*	*PO3*	**06**
	c)	Describe the key components of a Variational Autoencoder architecture and explain the role of the encoder, decoder, and latent space in data reconstruction.	*CO1*	*PO1*	**06**

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

			UNIT - II			
	3	a)	Compare the computational trade-offs of using dense versus sparse models in LLM training.	*CO3*	*PO2*	**06**
		b)	The self-attention mechanism involves calculating the attention matrix A from a sequence of tokens. Given the following example of query and key vectors: $Q=[0.2 \quad 0.4 \quad 0.6]$, $K=[0.5 \quad 0.2 \quad 0.3]$ Calculate the similarity score between Q and K and then compute the scaled dot-product attention score using the formula: Attention Score= $Q \cdot K^T$ / sqrt{d_k}	*CO3*	*PO3*	**08**
		c)	Explain the importance of scaling laws in large language models and it's in the design of compute-optimal models.	*CO2*	*PO2*	**06**
			OR			
	4	a)	Differentiate between one-shot prompting and few-shot prompting.	*CO1*	*PO2*	**06**
		b)	BloombergGPT is pre-trained on a financial corpus with a vocabulary size of 50,000. If each token is represented by a 512-dimensional vector, calculate the memory required to store the embeddings for the entire vocabulary.	*CO2*	*PO3*	**08**
		c)	Apply tokenization, embedding (d_{model} = 4, row major embedding from 0.1 to 2.0) and positional encoding processes of a Transformer Model on the sentence "AI is amazing.(fullstop)" and Compute the Final Embedded Tokens matrix.	*CO2*	*PO3*	**06**
			UNIT - III			
	5	a)	In Reinforcement Learning with Human Feedback (RLHF), given a sequence of 3 actions with rewards [8, 12, 20] and a discount factor of 0.85, calculate the discounted cumulative reward for the sequence.	*CO3*	*PO3*	**08**
		b)	Compare and contrast In-Context fine tuning and Model fine tuning.	*CO2*	*PO2*	**06**
		c)	Is it possible to optimize the objective function in soft prompts? Justify using required mathematical representations of the procedure.	*CO2*	*PO1*	**06**
			OR			
	6	a)	Calculate the total reward received over 5 iterations in RLHF, given a 20% reward decay per iteration and an initial reward of 100.	*CO3*	*PO3*	**08**
		b)	Derive the mathematical formulation of QLoRA.	*CO2*	*PO3*	**06**
		c)	Discuss model evaluation in fine-tuning large language models and the common benchmarks used to assess their performance.	*CO2*	*PO1*	**06**
			UNIT - IV			
	7	a)	In Proximal Policy Optimizer (PPO), given that the reward at time step t is 3, the advantage estimate A^t is 1.2, and the probability ratio $r_t(\theta)$=1.5, calculate the value of the objective function for	*CO3*	*PO3*	**08**

			ϵ=0.1 before and after applying the clip function. Also, compare the change in the objective function if ϵ were increased to 0.5.			
		b)	Identify challenges in implementing ReAct strategies in existing systems and propose viable solutions.	*CO3*	*PO2*	**06**
		c)	Describe quantization in large language models, its impact on efficiency, and the advantages of QLORA for fine-tuning.	*CO2*	*PO1*	**06**
			OR			
	8	a)	Elaborate on the reward model construction process and associate math behind it.	*CO3*	*PO1*	**08**
		b)	Given the loss at each reasoning step $L_{step}=1 / k$ for $k = 5$, calculate the total loss. Also, if the loss decreases exponentially with a decay factor of 0.8, find the total loss after 5 steps.	*CO2*	*PO3*	**06**
		c)	Explain chain-of-thought reasoning and its impact on improving model performance on complex tasks.	*CO3*	*PO1*	**06**
			UNIT - V			
	9	a)	Elaborate on the generator and the discriminator objective function of a Style Transfer Generative Adversarial Network (GAN).	*CO2*	*PO3*	**06**
		b)	Explain the concept of the minimax game in GAN training. How does this concept relate to adversarial loss?	*CO2*	*PO1*	**06**
		c)	Given: Conditional GAN (CGAN): • Discriminator Loss (L_D): 0.3 for the real images. • Generator Loss (L_G): 0.8 for the generated images. Wasserstein GAN (WGAN): • Discriminator Loss (L_D): 0.2 for the real images. • Generator Loss (L_G): 0.7 for the generated images. i. Interpret the Discriminator Loss and Generator Loss scores of both the models in terms of training behavior for both models. Given that both models have 100 neurons in the first layer, and the generator in both cases has a latent vector of size 100 and a class label vector of size 3, total number of biases as 100, calculate the number of parameters in the conditional input layer for both CGAN and WGAN.	*CO3*	*PO3*	**08**
			OR			
	10	a)	In WGAN, the Wasserstein loss is given by $L_D=E[D(x)]-E[D(G(z))]$ Explain how the gradient penalty term is added to enforce the Lipschitz constraint.	*CO3*	*PO3*	**08**
		b)	List out the key differences between DCGAN, WGAN, and CGAN regarding their architecture and training methodologies.	*CO3*	*PO1*	**06**
		c)	Illustrate GAN architecture in detail.	*CO2*	*PO3*	**06**

******