U.S.N.										

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

Autonomous Institute Affiliated to VTU

April 2025 Semester End Make-Up 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)	Given a latent variable $z \sim N(\mu,\sigma^2)$, compute the KL divergence between the approximate posterior q(z|x) and the prior $p(z) \sim N(0,1)$.	*CO1*	*PO2*	**06**
	b)	Are Generative models deterministic or probabilistic in nature? Explain.	*CO2*	*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).	*CO3*	*PO3*	**08**
		OR			
2	a)	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**
	b)	Differentiate between Generative and Discriminative models.	*CO1*	*PO2*	**06**
	c)	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.	*CO3*	*PO3*	**08**
		UNIT - II			
3	a)	Using any 2 specific applications in finance, explain how does BloombergGPT model exhibit better performance in financial contexts compared to general-purpose models.	*CO2*	*PO1*	**05**

<table>
<tr><td></td><td></td><td>b)</td><td>Illustrate the transformer architecture in detail.</td><td>CO1</td><td>PO2</td><td>05</td></tr>
<tr><td></td><td></td><td>c)</td><td>Using the Layer Normalization parameters: Scaling Factor(γ)=1, shifting Factor(β)=0, apply Layer Normalization to the input vector [2,4,6] and determine the transformed output vector.</td><td>CO3</td><td>PO3</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>4</td><td>a)</td><td>Differentiate between one-shot learning and zero-shot learning.</td><td>CO1</td><td>PO2</td><td>06</td></tr>
<tr><td></td><td></td><td>b)</td><td>Given a transformer model with d_{model}=512, calculate the total number of parameters in a single multi-head attention layer with 8 attention heads.</td><td>CO2</td><td>PO3</td><td>06</td></tr>
<tr><td></td><td></td><td>c)</td><td>Apply tokenization, embedding (d_{model} = 6, row major embedding from 0.1 to 3.6) and positional encoding processes of a Transformer Model on the sentence “Deep learning works.(fullstop)” and Compute the Final Embedded Tokens matrix.</td><td>CO3</td><td>PO3</td><td>08</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - III</td><td></td><td></td><td></td></tr>
<tr><td></td><td>5</td><td>a)</td><td>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.</td><td>CO1</td><td>PO3</td><td>08</td></tr>
<tr><td></td><td></td><td>b)</td><td>Derive the mathematical framework of scaling instruct models.</td><td>CO3</td><td>PO3</td><td>06</td></tr>
<tr><td></td><td></td><td>c)</td><td>Explain Bilingual Evaluation Understudy (BLEU) metric in detail.</td><td>CO2</td><td>PO1</td><td>06</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>6</td><td>a)</td><td>Determine the computational cost in FLOPs for multi-task fine-tuning with 100 million parameters, 500,000 examples, 104 operations per forward pass, and 10 epochs.</td><td>CO1</td><td>PO2</td><td>07</td></tr>
<tr><td></td><td></td><td>b)</td><td>Explain Rouge metric in detail.</td><td>CO2</td><td>PO1</td><td>06</td></tr>
<tr><td></td><td></td><td>c)</td><td>Is it possible to optimize the objective function in soft prompts? Justify using required mathematical representations of the procedure.</td><td>CO2</td><td>PO2</td><td>07</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - IV</td><td></td><td></td><td></td></tr>
<tr><td></td><td>7</td><td>a)</td><td>Elaborate on the reward model construction process and associate math behind it.</td><td>CO1</td><td>PO2</td><td>10</td></tr>
<tr><td></td><td></td><td>b)</td><td>A language model is predicting the next word in a sequence. The true distribution P and the model's predicted distribution Q for the next word are:
<table>
<tr><th>Word</th><th>P(w)</th><th>Q(w)</th></tr>
<tr><td>"apple"</td><td>0.4</td><td>0.5</td></tr>
<tr><td>"banana"</td><td>0.3</td><td>0.2</td></tr>
<tr><td>"cherry"</td><td>0.2</td><td>0.2</td></tr>
<tr><td>"date"</td><td>0.1</td><td>0.1</td></tr>
</table>
Calculate the KL Divergence $D_{KL}(P||Q)$ between the true distribution P and the model's distribution Q.</td><td>CO3</td><td>PO3</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>8</td><td>a)</td><td>Elaborate the process of obtaining feedback from humans and integrating into Machine Learning.</td><td>CO1</td><td>PO2</td><td>10</td></tr>
</table>

		b)	A large language model is employed to solve a real-world problem using Chain-of-Thought (CoT) reasoning. The scenario is: A customer purchases 3 pens and 2 notebooks. • Each pen costs $2. • Each notebook costs $5. • The customer pays with a $20 bill. The language model's reasoning process to calculate the change is: • Calculate the total cost of pens. • Calculate the total cost of notebooks. • Add the two totals to get the total cost. • Subtract the total cost from $20 to find the change. i. Analyse the accuracy of the model's reasoning process. ii. Apply the given steps to solve the problem and determine the customer's change. iii. How does Chain-of-Thought (CoT) reasoning help to improve logical consistency and accuracy in solving the stated scenario of multi-step problems?	*CO3*	*PO3*	**10**
			UNIT - V			
	9	a)	A Deep Convolutional Generative AI Network (DCGAN) is trained to generate 100×100 grayscale images. The Generator (G) takes a 200-dimensional noise vector (z) as input, and the Discriminator (D) classifies the generated images as fake or real. i. If the Discriminator (D) outputs a probability of 0.6 for the generated image being real, calculate the Generator Loss. ii. If the Discriminator classifies real images correctly with a probability of 0.95, and the fake images are classified correctly with a probability of 0.7, calculate the Discriminator Loss.	*CO3*	*PO3*	**08**
		b)	Elaborate on any 2 variations available for realization of image synthesis application of a Generative Adversarial Networks (GANs).	*CO1*	*PO2*	**06**
		c)	With the help of mathematical representations, explain the maximization step of a GAN training process.	*CO2*	*PO4*	**06**
			OR			
	10	a)	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. ii. 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**

		b)	With the help of mathematical representations, explain the minimization step of a GAN training process.	*CO2*	*PO4*	**06**
		c)	Illustrate GAN architecture in detail.	*CO1*	*PO2*	**06**

******