Lab with Capstone Project - GenAI

Code: CS32241CS	Course Type: Prct.	Semester: II
Evolution Scheme (TA-MSE-ESE): 25-25-50	Total Marks: 100	L-T-P (Credits): 0-0-4 (2)

Pre-Requisites:

- Advanced Deep Learning and Computer Vision (mandatory)
- Probability Theory (ELBO, KL divergence, Gaussian sampling)
- Python (PyTorch at an intermediate level: custom training loops, hooks)
- Linux command line; GPU environment (CUDA, conda); basic experiment tracking (W&B/MLflow)
- Basic NLP concepts (tokens, embeddings)

Course Outcomes: At the end of the course, the student will be able to

CO-1	Implement VAEs, β-VAEs, and VQ-VAEs from scratch, and analyze latent space geometry through interpolation and disentanglement metrics.
CO-2	Train DCGAN, WGAN-GP, and Conditional GAN models, diagnose training instabilities, and interpret generative evaluation metrics (FID, IS).
CO-3	Implement Diffusion Models (DDPM/DDIM) and manipulate classifier-free guidance, comparing speed–quality tradeoffs across samplers.
CO-4	Fine-tune pre-trained LLMs using Parameter-Efficient Fine-Tuning (QLoRA) and conduct alignment experiments utilizing Direct Preference Optimization (DPO).

Course Articulation Matrix:

PO	PEO-1	PEO-2	PEO-3	PEO-4	PEO-5
CO-1	3	3	2	-	2
CO-2	3	3	2	2	2
CO-3	3	3	3	-	1
CO-4	3	3	3	2	3

1 - Slightly; 2 - Moderately; 3 – Substantially

Syllabus:

Latent variable models including Variational Autoencoders (VAEs), reparameterization techniques, latent space interpolation, and disentanglement methods such as β-VAE; adversarial networks covering GAN architectures (DCGAN, WGAN-GP), conditional generation, mode collapse analysis, and evaluation metrics like FID and IS; diffusion and score-based models including DDPM formulations, DDIM-based accelerated sampling, classifier-free guidance (CFG), and latent diffusion models (LDMs); large language model fine-tuning and alignment through parameter-efficient approaches (PEFT), LoRA/QLoRA techniques, and alignment strategies such as RLHF and DPO; and multimodal systems with orchestration frameworks encompassing retrieval-augmented generation (RAG) pipelines, semantic chunking, and agentic workflows with tool integration.

List of Experiments

Exp-1. VAE Implementation from Scratch: Implement an encoder ($q\varphi$) and decoder ($p\theta$) using the reparameterization trick. Compute reconstruction loss and KL divergence, verifying posterior collapse indicators.

Exp-2. Latent Space Geometry & Disentanglement: Train a β-VAE on a factored dataset. Perform linear/spherical latent space interpolations and compute Mutual Information Gap (MIG).\

Exp-3. Discrete Latent Spaces (VQ-VAE): Implement a Vector Quantized VAE with an embedding codebook. Compute commitment loss, straight-through estimators, and analyze codebook utilization.

Exp-4. Deep Convolutional GANs (DCGAN): Build and train a DCGAN. Monitor discriminator vs. generator loss curves to visually analyze convergence dynamics and early-stage artifacts.

Exp-5. WGAN-GP & Mode Collapse Recovery: Implement the Wasserstein distance metric and gradient penalty loss ($\lambda = 10$). Compare training stability against standard DCGAN setups.

Exp-6. Conditional Generation & Evaluation Suite: Train a Conditional GAN (Pix2Pix) using a U-Net generator. Implement and compute Fréchet Inception Distance (FID) and Inception Score (IS).

Exp-7. DDPM Forward & Reverse Processes: Formulate the forward Gaussian diffusion process (T = 1000 steps) and train a U-Net denoiser (-prediction) from scratch on a foundational dataset.

Exp-8. DDIM Sampling & Classifier-Free Guidance: Compare DDPM ancestral sampling with deterministic DDIM. Implement CFG and plot the tradeoff between generation diversity and fidelity.

Exp-9. Stable Diffusion Fine-Tuning & ControlNet: Utilize HuggingFace Diffusers to finetune a Latent Diffusion Model using LoRA, and deploy ControlNet for edge/depth conditioned text-to-image generation.

Exp-10. LLM Architecture & Inference Optimization: Load an open-weights LLM (e.g., LLaMA3). Inspect tokenizer behavior (BPE), analyze attention mechanisms, and test greedy vs. nucleus (top-p) sampling.

Exp-11. Domain Adaptation via QLoRA: Prepare a domain-specific instruction dataset. Finetune a quantized LLM utilizing Parameter-Efficient Fine-Tuning (QLoRA) and analyze weight norms.

Exp-12. LLM Alignment via Direct Preference Optimization (DPO): Prepare a preference dataset(chosen vs. rejected). Implement DPO using the TRL library and evaluate response quality using MT-Bench style metrics.

Exp-13. End-to-End RAG System Construction: Build a Retrieval-Augmented Generation pipeline combining recursive semantic chunking, a vector database (e.g., FAISS/ChromaDB), and an LLM generator.

Exp-14. Multimodal GenAI (Visual Question Answering): Load a Multimodal LLM (e.g., LLaVA or InternVL). Perform VQA tasks and evaluate visual hallucination using standard benchmarks like POPE.

Exp-15. Agentic AI with Function Calling & Tool Use: Design a multi-step reasoning agent (e.g., ReAct paradigm) integrated with external tools (web search, Python REPL calculator) and log execution traces.

Mini Project / Capstone (Mandatory):

The faculty-in-charge shall assign capstone projects, which students are required to complete and submit

Learning Resources:

Text Books:

- Understanding Deep Learning (Ch 17–21) PrinceS.J.D. MIT Press 2024
- Generative Deep Learning Foster D. O'Reilly Media 2023 (2nd Ed.)
- Natural Language Processing with Transformers Tunstall L., et al. O'Reilly Media 2022

Reference Books:

- LLM Engineer's Handbook Iusztin P., Labonne M. Packt Publishing 2024
- Probabilistic Machine Learning: Advanced Topics Murphy K.P. MIT Press 2023