CS F437 Generative AI
Course Assignment

Objective

This assignment assesses your proficiency in core generative AI implementation and your ability to advance current methodologies through research and innovation. The assignment includes two essential components:

1. **Applied Problem-Solving:** Implement challenging generative AI tasks with specified constraints.

2. **Research & Innovation:** Integrate cutting-edge advancements from recent literature into your implementations and demonstrate measurable improvements.

Guidelines

- Implementations must be in Python using either PyTorch or TensorFlow.

- Use of pre-built models or libraries (e.g., Hugging Face, torchvision, timm, PyTorch Lightning) is strictly prohibited. Core components must be self-developed.

- Each submission must include a technical report detailing your approach, challenges faced, methodologies applied, and results.

- Students must form groups of exactly **2 students**.

Assignment Tasks

Choose **one** of the following tasks for implementation:

Task 1: Structured Latent Space in Variational Autoencoders (VAEs)

Problem Statement:

Design and implement a custom Variational Autoencoder (VAE) explicitly engineered to create a structured and semantically meaningful latent space. The default dataset provided for this task is Omniglot. However, if you choose a different dataset, you must provide a clear, detailed justification explaining how your selected dataset will enable you to effectively demonstrate the required capabilities listed below:

- **Continuous Interpolation:** Demonstrate smooth and meaningful transitions between character classes or semantic categories without abrupt distortions.

- **Style-Preserving Content Transfer:** Clearly separate the style from the fundamental shape/content to enable seamless transfer of writing style across different characters.

- **Semantic Latent Manipulation:** Structure your latent dimensions to encode interpretable semantic variations such as stroke thickness, slant, or complexity.

Constraints:

- Implement a novel custom latent regularization strategy explicitly designed to support structured latent representations (standard KL divergence approaches are not permitted).
- The encoder-decoder architecture must be original and not copied from existing standard VAE implementations.

Deliverables:

- Fully functional, original implementation of the VAE model.
- Comprehensive latent space visualizations clearly demonstrating interpolation, style transfer, and semantic manipulation.
- A detailed technical report explaining architectural choices, latent regularization methods, semantic dimension justifications, and experimental validation.

Task 2: Adversarial Attacks & Defense in Generative Models

Problem Statement:

Critically analyze and evaluate the adversarial robustness of a pre-trained GAN model (examples include StyleGAN or DCGAN). Specifically:

- **Develop a Black-box Adversarial Attack:** Design and implement an adversarial strategy using latent space perturbations to cause subtle yet significant distortions or biases in the GAN-generated outputs.
- **Adaptive Defense Strategy:** Create a novel, adaptive defense mechanism that integrates seamlessly into existing pre-trained GANs without retraining from scratch, effectively countering the adversarial attack.
- **Quantitative and Qualitative Evaluation:** Employ robust perceptual metrics and statistical methods to quantitatively and qualitatively evaluate the effectiveness of your attack and defense strategy.

Constraints:

- The adversarial attack must strictly follow a black-box approach without gradient access.
- The defense strategy must be implemented as a plug-in or post-hoc mechanism compatible with existing GAN frameworks.

Deliverables:

- Source code for both adversarial attack and defense mechanisms.
- Clearly presented visual comparisons of original, attacked, and defended images.
- Comprehensive evaluation report detailing methodology, metrics utilized, effectiveness assessments, and conclusions drawn from the evaluation.

Part 2: Research & Innovation

Integrate recent research into your chosen task:

- Perform a literature review and select one unique research paper relevant to your task.
- Clearly articulate and justify your proposed enhancement inspired by the chosen research paper.
- Implement and empirically demonstrate your enhancement's measurable improvement over the baseline implementation, utilizing the evaluation metrics below.

Evaluation Metrics:

Task	Metrics
VAE Latent Space	KL Divergence, Reconstruction Loss, Clustering Accuracy
GAN Robustness	Attack Success Rate, FID, LPIPS

Timeline & Important Deadlines

- **Group Formation Deadline:** March 31, 2025
- **Part 1 Submission & Research Paper Registration:** April 10, 2025
 (On this date, Part 1 implementation will be reviewed, and each group must officially register their unique research paper for Part 2.)
- **Final Assignment Submission:** April 20, 2025
- **Viva/Presentation Date:** June 21–23, 2025

Group and Paper Registration

- Group registration must be completed by the specified deadline. Contact course TA for registration

- Research paper selection is unique per group and operates on a first-come-first-served basis. Registration is done with Course TA.

Final Deliverables

- **Annotated Literature Survey:** Summarize key insights from the selected research paper.
- **Enhanced Implementation Code:** Clearly demonstrate the implemented enhancement and provide comparative results.
- **Performance Analysis Report:** Provide an in-depth analysis illustrating how the enhancement impacted overall model performance.

Grading Criteria

Component	Weight (%)
Part 1 Implementation + evaluation + analysis	50%
Research & Paper Selection	10%
Proposed Enhancement	10%
Enhancement Implementation	30%
Evaluation & Analysis	10%

Academic Integrity

All submissions will be rigorously reviewed to ensure originality and compliance with academic integrity standards. Unauthorized use of pre-built libraries, models, or plagiarism is strictly prohibited and will result in severe penalties.