# Internship Report

Name: Ojas Sharma

Company: NullClass

Internship Duration: 12 May 2025 – 12 June 2025

Project: Text-to-Image GAN Pipeline with NLP Integration

## 1. Introduction

This report details my internship experience at NullClass, where I worked on a text-to-image generation pipeline integrating natural language processing and deep learning models. The objective was to simulate a real-world application by combining BERT-based tokenization, CLIP embeddings, and GAN-based image generation. The internship offered a hands-on opportunity to implement advanced AI technologies within an end-to-end application.

## 2. Background

Text-to-image generation is a key challenge in AI and vision-language research. It requires converting textual descriptions into meaningful visual representations. GANs (Generative Adversarial Networks) have become essential in this domain. Combined with models like BERT and CLIP, they allow for semantically rich and accurate image generation based on text input. This internship aimed to combine these approaches into a unified system.

## 3. Learning Objectives

- Understand and apply tokenization using BERT and CLIP.

- Build a preprocessing pipeline for natural language input.

- Develop a GAN-based model for image generation.

- Integrate a complete end-to-end system.

- Evaluate model performance with metrics like accuracy, precision, recall.

- Deploy the project using a GUI (Streamlit).

## 4. Activities and Tasks

Task 1: BERT Tokenizer & Encoder

- Used Hugging Face Transformers.

- Tokenized input text and converted into embedding tensors.

Task 2: CLIP Text Embedding

- Used CLIPTokenizer and CLIPTextModel to obtain semantic embeddings.

Task 3: Text-to-Image GAN Pipeline

- Integrated a GAN model with the preprocessed and embedded text.

- Trained a ResNet-based classifier for evaluation.

GUI Integration

- Developed a Streamlit app for live input/output.

- Incorporated all tasks in a single unified interface.

## 5. Skills and Competencies Acquired

- NLP using Transformers (BERT, CLIP)

- PyTorch and Hugging Face Libraries

- GAN and Image Generation Techniques

- Model Evaluation (Confusion Matrix, Precision, Recall)

- Streamlit App Deployment

- GitHub and Google Drive integration for model sharing

## 6. Feedback and Evidence

The outcomes were tested using benchmark datasets and validated through visual inspection and classification metrics. The model achieved over 85% accuracy. Outputs were manually reviewed for semantic alignment with text. Screenshots and sample outputs are available in the GitHub repository.

## 7. Challenges and Solutions

- Challenge: Embedding alignment between CLIP and GAN

Solution: Standardized embedding dimensions and normalized input.

- Challenge: Memory management in model training

Solution: Batch processing and reduced latent vector size.

- Challenge: Streamlit integration

Solution: Modularized codebase for clean integration.

## 8. Outcomes and Impact

- Successfully built an end-to-end system for text-to-image generation.

- Learned practical integration of NLP and CV.

- Gained valuable experience in deploying real-world AI applications.

- Produced a GitHub-hosted project ready for professional portfolios.

## 9. Conclusion

This internship provided me with an incredible opportunity to combine deep learning models and deploy a functional application. It improved my understanding of both language processing and image synthesis. I am grateful to NullClass for this enriching experience and look forward to applying these skills in future AI roles.