Final AI Project Report

Project Title:

Prompt-Based Image Generation using Fine-Tuned Stable Diffusion

1. Project Overview

The goal of this project is to develop a **prompt-to-image generation system** by **fine-tuning a Stable Diffusion model** on a custom dataset. Users can input natural language prompts, and the system generates high-quality images that match the description. This project demonstrates the potential of **generative AI in creative content generation** such as digital art, product design, and marketing materials.

2. Objectives

- Fine-tune the pre-trained **Stable Diffusion** model on a custom image-text dataset.
- Improve image output relevance and quality for specific domains (e.g., healthcare, fashion, architecture).
- Build a user interface where users can enter prompts and receive generated images.
- Evaluate model performance using visual and textual similarity metrics.

3. Technologies and Tools Used

- **Python** – Core programming language.
- **Hugging Face Diffusers** – For loading and fine-tuning Stable Diffusion.
- **PyTorch** – Deep learning backend.
- **CUDA / GPU (NVIDIA)** – For fast training and inference.
- **Transformers** – For tokenizing and understanding text prompts.
- **Gradio / Streamlit** – (optional) for creating a web-based user interface.
- **Datasets** – Custom or publicly available datasets like LAION, Unsplash, etc.

4. Methodology

Step 1: Pre-trained Model Setup

- Loaded the pre-trained **Stable Diffusion model** using Hugging Face Diffusers.

Step 2: Dataset Preparation

- Prepared a dataset of **image-prompt pairs**.
- Preprocessed text using tokenizer (CLIPTokenizer).
- Resized and normalized images.

Step 3: Fine-Tuning Process

- Trained the model using the **DreamBooth or LoRA (Low-Rank Adaptation)** technique for efficient and memory-friendly fine-tuning.
- Used a **learning rate scheduler**, **gradient accumulation**, and **checkpointing** for optimized training.

Step 4: Image Generation

- Accepted a prompt from the user.
- Tokenized the prompt and fed it to the fine-tuned model.
- Generated images using the **denoising diffusion process**.

5. Key Features

- **Prompt-to-Image Synthesis**: Converts user input into unique images.
- **Custom Domain Specialization**: Fine-tuned on a specific dataset (e.g., medical imaging, fashion).
- **Real-Time Inference**: Generates output images in seconds with GPU support.
- **User Interface**: Optionally built with Gradio for demo purposes.

6. Results

- Generated high-quality images that align with custom prompts.
- Improved accuracy and relevance due to fine-tuning on a targeted dataset.

- Demonstrated model's understanding of abstract, descriptive, and domain-specific prompts.

8. Evaluation Metrics

- **FID (Fréchet Inception Distance)** – to evaluate image quality.
- **CLIP Similarity Score** – to measure how well the image matches the prompt.
- **User Study / Visual Assessment** – to get subjective feedback on image realism and relevance.

9. Challenges Faced

- **High VRAM Requirements** – Managed using LoRA and smaller batch sizes.
- **Training Time** – Reduced using pretrained weights and 8-bit optimizers.
- **Prompt Alignment** – Required prompt engineering to improve output quality.

10. Future Scope

- **Model Deployment on Web/App**.
- **Support for Multiple Languages**.
- **Interactive Prompt Editing**.
- **Training on Larger Custom Datasets**.
- **Use in Real-World Applications** such as e-commerce mockups or education.

11. Conclusion

This project successfully demonstrates the power of generative AI using **fine-tuned diffusion models**. By customizing Stable Diffusion with a domain-specific dataset, we enhanced its ability to generate images that are both visually appealing and prompt-accurate. The system showcases practical use cases in design, marketing, and AI-driven creativity.