TEXT TO IMAGE GENERATOR USING STABLE DIFFUSION

BY LAVKESH MADAVI
(2311MC08)

PROJECT OVERVIEW

- Technology Used: Stable Diffusion (stabilityai/stable-diffusion-2-1)

- Frameworks & Libraries:
  - Diffusers: For image generation
  - Torch: Deep learning framework for computations.
  - Gradio: User interface for web-based interaction
  - PIL: Image processing

- Objective:
  - Generate high-quality images from text prompts
  - Allow users to customize parameters for better results

MODEL & PIPELINE SETUP

- Pretrained model: "stabilityai/stable-diffusion-2-1"
- Pipeline Initialization:
  - Uses StableDiffusionPipeline from Hugging Face
  - Runs on CPU (since the free Hugging Face Space does not support GPU)
- Schedulers for Image Generation:
  - Euler Discrete
  - LMS Discrete
  - PNDM
  - DPM Solver Multistep
  - DDIM

IMAGE GENERATION PROCESS

- Inputs:
  - Prompt: Description of the desired image
  - Negative Prompt: What to exclude
  - Inference Steps: Number of denoising steps
  - Guidance Scale: Balances prompt adherence
  - Image Dimensions: Height & width selection
  - Batch Size: Number of images generated
  - Scheduler: Algorithm selection

- Processing:
  - Selects the scheduler based on user input
  - Sets a random seed for reproducibility (if provided)
  - Generates an image output list

GRADIO WEB APP & DEPLOYMENT

- Frontend Features:
  - Textboxes: User input for prompt & negative prompt
  - Sliders: Fine-tune image generation parameters
  - Dropdown: Select different scheduler algorithms
  - Gallery Output: Displays generated images

- Web app deployment:
  - Hugging Face Spaces:
  - Web demo hosted on free tier (CPU-based) → Slower execution.
  - Link: https://huggingface.co/spaces/madavilavkesh/text2image_stable_diffusion

- For Faster Execution:
  - Run the IPYNB file on Google Colab from github.
  - Link: https://github.com/madavilavkesh/Text-to-Image-Stable-Diffusion
  - Enable GPU: Runtime → Change runtime type → GPU
  - Faster processing compared to CPU.