

Problem Statement:

In the rapidly advancing field of artificial intelligence, the ability to generate high-quality images from natural language descriptions represents a significant milestone. Despite recent progress in text-to-image generation, existing models often struggle with capturing intricate details, maintaining coherence, and producing high-fidelity images across diverse domains. The challenge lies in creating a robust text-to-image generative model that can accurately translate complex textual descriptions into realistic images, applicable to a wide range of scenarios including everyday objects, complex scenes, and varied landscapes.

Objective: To develop and deploy a robust pipeline leveraging the Stable Diffusion model and GPT-3 for generating high-quality images from textual descriptions. This project aims to merge advancements in natural language processing and computer vision to enable innovative applications, thereby pushing the boundaries of what is achievable in AI-driven image generation.

Approach-

The Jupyter notebook outlines a process for generating images from text using a Stable Diffusion model. It begins by installing and importing necessary libraries, including ``accelerate``, ``keras``, ``torch``, ``diffusers``, and ``transformers``. Configuration parameters are defined in two classes, ``work`` and ``DATA``, which include settings such as seeds for reproducibility, image size, and model identifiers. The Stable Diffusion model is then loaded and set to run on a CUDA device, with a manual seed set for consistency. A function ``generate_image`` is defined to generate and resize images based on input text prompts using the model. Finally, the notebook demonstrates generating an image from the text prompt "full porsche" to test the implementation.

Output-

