Text-to-Image with Stable Diffusion: Solution & Approach

Mission: Democratize creativity through AI by enabling text-based image generation.

Three Pillars: Parameters, Pipeline, and Output form the foundation.

This open-source model uses diffusion techniques to craft detailed images.

by Gurnoor Singh

Parameters: The Artist's Palette

Prompt

Text guiding the image (e.g., ~~"A futuristic city at sunset"~~)

Guidance Scale

Controls prompt strength; typically set at 7.5 for balance

Steps

Number of refinement iterations; more steps mean finer detail.

Seed

Seed = 42 ensures reproducible results across runs.

sample code: seed = 42

Pipeline: From Text to Image

Tokenizer

Encodes text prompts into numerical tokens for processing.

UNet

Iteratively denoises latent space using text embeddings.

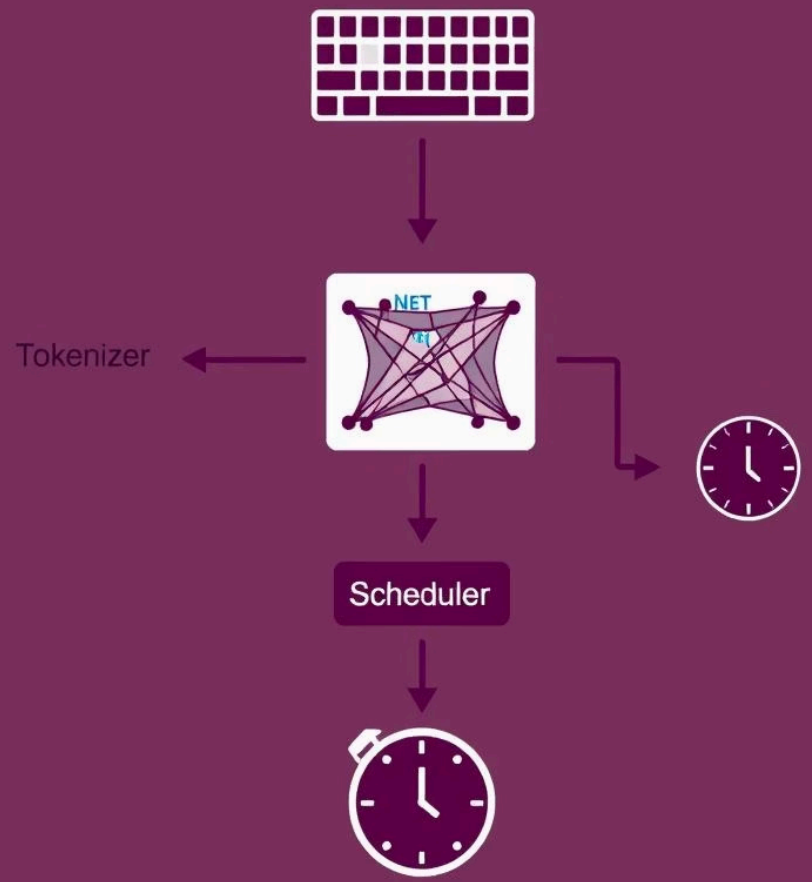
VAE (Variational Autoencoder)

Decodes latent data into the final image pixels.

Scheduler

Controls diffusion steps; example schedulers include DDIM and Euler.

AI Text-to image
AI text to pripline



Output: The Final Creation

Inference Timing

Approximately 20-30 seconds per image on consumer GPUs.

Single vs Batch

Supports single image or batch generation to boost throughput.

Saving

Save images as standard formats like PNG or JPG.

Reproducibility

Same seed and parameters reliably recreate images, aiding iteration.