

Generative Milestone 1 (Group- 7)

GitHub link: https://github.com/Sanika0701/nn-dl_gen_ai_milestone1.git

1. Dataset Used

Dataset: COCO 2017 Validation Subset

Size: 5 000 image-caption pairs

Preprocessing: Each image resized to 512×512 , normalized to $[-1, 1]$.

Annotations: Loaded from captions_val2017.json; first caption per image used.

Split: Single subset used for baseline inference (no training yet).

2. Model Architecture

- **Text Encoder:** openai/clip-vit-base-patch32 from Hugging Face Transformers.
- **Image Generator:** stabilityai/stable-diffusion-2-1 from Diffusers.
- **Frameworks:** PyTorch 2.2 + Diffusers 0.31 + Transformers 4.44.

3. Roles

Sailee Choudhari – Data Preprocessing & Pipeline Lead

Sanika Chaudhari – Model Integration Engineer

Parth Saraykar – Experimentation & Tuning Specialist

Novia Dsilva – Visualization & Reporting Analyst

4. Sample Prompts and Generated Images

"a cat sitting on a couch", "a beautiful sunset over the ocean", "a person riding a bicycle in the park", "a bowl of fresh fruit on a table", "a dog playing with a ball in the garden"

