

Vision From Words Using Artificial Intelligence and Machine Learning

Noothan S N and Thouseef Ulla Khan

Department of MCA

Vidya Vikas Institute of Engineering and Technology, Mysuru, India

snnoothan@gmail.com, thouseef.khan@vidyavikas.edu.in

Abstract: In recent years, the convergence of natural language processing (NLP) and computer vision has enabled remarkable progress in artificial intelligence, particularly in the area of text-to-image generation. This work, titled “Vision from Words Artificial Intelligence and Machine Learning”, presents a system capable of synthesizing novel, realistic images directly from natural language prompts. The project explores both Generative Adversarial Networks (GANs) and diffusion-based models, demonstrating evolution from baseline implementations to state-of-the-art fine-tuning strategies. An initial GAN-based pipeline, comprising a GRU text encoder, convolutional generator, and discriminator, validated the feasibility of text-to-image synthesis but was constrained to low-resolution outputs. To address these limitations, the system transitioned to a diffusion model, fine-tuning the Stable Diffusion v1.5 UNet component on a custom dataset of approximately 500 text-image pairs. This approach produced sharper, semantically coherent, and higher-resolution images while maintaining training stability.

Beyond model development, the project incorporates deployment through a Flask-based web application featuring secure authentication, prompt submission, negative prompt filtering, and image saving. This integration bridges research and usability, providing a practical platform accessible to non-technical users. The results demonstrate that diffusion-based models significantly outperform GANs in realism and semantic alignment, and highlight the transformative potential of text-to-image generation in creative industries such as advertising, digital art, gaming, education, and design.

Keywords: Text-to-Image Generation, Deep Learning, Generative Adversarial Networks (GANs), Diffusion Models, Stable Diffusion, Natural Language Processing (NLP), Computer Vision, GRU Encoder, UNet, Flask Web Application, Creative Automation

