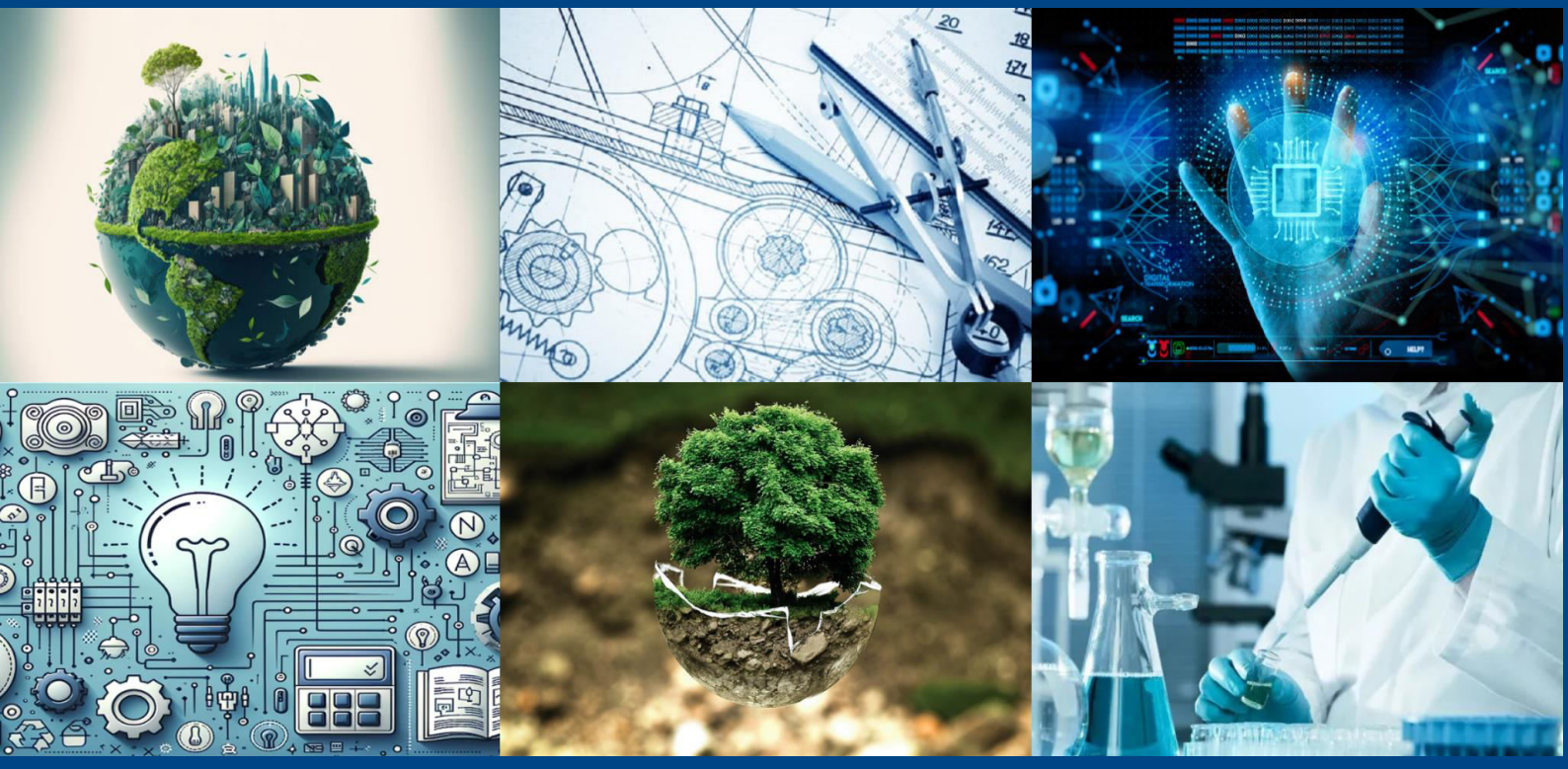
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 3, March 2025**

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI:10.15680/IJMRSET.2025.0803074

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI based Text to Image Generation

Dipali Kore, Vaishnavi Lobhe, Anjali Korde, Nikita Chavan, Ms.A.B.Jewalikar

Department of Computer Engineering, Jayawantrao Sawant Polytechnic, Pune, India

ABSTRACT: This project presents a novel AI-based text-to-image generation system that leverages the power of deep learning to create stunning visuals from textual descriptions. By harnessing the capabilities of natural language processing (NLP) and computer vision, our system generates high-quality images that accurately reflect the input text.

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has led to the development of innovative technologies that can generate visual content from textual descriptions. AI-based text-to-image generation is a cutting-edge technology that enables users to create stunning images from text inputs, revolutionizing the way we interact with visual content.Text-to-image generation is a type of AI-based image synthesis that involves generating images from textual descriptions.

II. ALGORITHM USED

1. **Generative Adversarial Networks (GANs):** GANs consist of two neural networks: a generator and a discriminator. The generator creates images from text, while the discriminator evaluates the generated images and provides feedback to the generator.

2. **Variational Autoencoders (VAEs):** VAEs are neural networks that learn to compress and reconstruct data. In text-to-image generation, VAEs can be used to encode text into a latent space and then generate images from this space.

3. **Transformers:** Transformers are a type of neural network architecture introduced in the paper "Attention is All You Need." They rely entirely on self-attention mechanisms to process input sequences. Transformers can be used for text-to-image generation by encoding text into a sequence of vectors and then generating images from these vectors.

4. **Auto-Regressive Models:** Auto-regressive models generate images pixel by pixel, conditioned on the previous pixels and the input text. These models can be computationally expensive but can produce high-quality images.

III. SCOPE

1. **Image Generation**: AI-based text-to-image generation can produce high-quality images from textual descriptions, including objects, scenes, and abstract concepts.

2. **Visual Content Creation:** This technology can automate visual content creation for various industries, such as advertising, graphic design, and entertainment

3. **Art and Design:** AI-based text-to-image generation can assist artists and designers in exploring new ideas, creating concept art, and automating repetitive tasks.

IV. PROPOSED SYSTEM

The proposed system for will integrate a natural language processing module with a state-of-the-art image generation framework.
First, the system will analyze the input text to extract key features and context.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Next, it will utilize Advanced Convolutional Neural Network (CNN) to create corresponding images that accurately reflect the description.

The system will also include a user-friendly interface that allows users to input text and receive high-quality images in real time, making it accessible for various applications in creative industries and education.

V. IMPLEMENTATION

A. SYSTEM REQUIREMENTS

Many implementations may exist for a given Specifications:

Hardware Requirements:

- Processor : Intel-V
- Speed : 1.2 Ghz
- RAM : 8 GB(min)
- Hard Disk : 512 GB (SSD)
- IO Devices : Key Board, Mouse
- Monitor : LCD/LED

Software Requirements:

- Operating system : Windows XP/7/LINUX.
- Front End : Python 3.8
- Back End : MySQL 5.5
- Tool/IDE : Python IDEL, VS Code
- Server : Web Server

B. TECHNOLOGY

FRONTEND:
Flask framework
Html.
CSS.

BACKEND:
Python 3.8+

Database:
MySQL 5.5

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI:10.15680/IJMRSET.2025.0803074

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

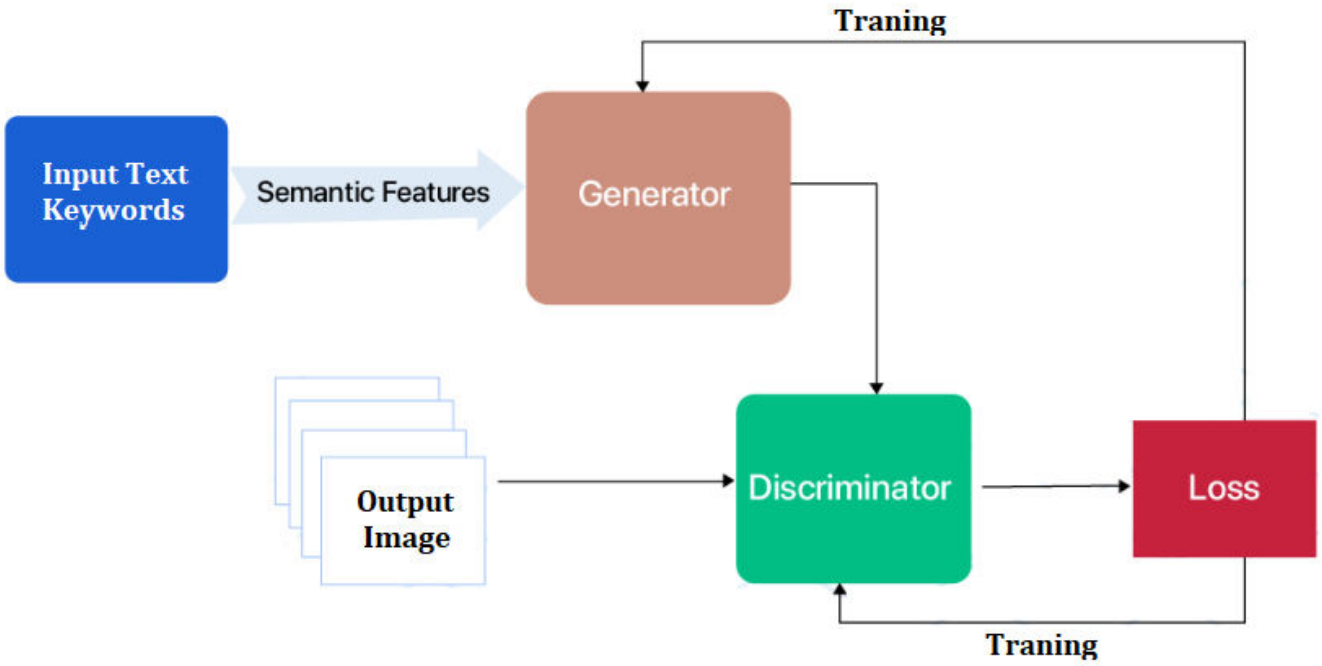


Fig. 1. Block diagram

1. Text Input: Textual description is input into the system.
2. Text Encoder: Text is encoded into a numerical representation using NLP techniques.
3. Text Embedding: Encoded text is embedded into a vector space.
4. Image Generator: Vectorized text is used to generate an image using a GAN or VAE.
5. Image Refiner: Generated image is refined using techniques such as super-resolution or style transfer.
6. Image Output: Final refined image is output.

VI. ACCURACY

The degree to which the generated images accurately reflect the input text, measured by metrics such as Inception Score, Fréchet Inception Distance, Structural Similarity Index Measure, and Peak Signal-to-Noise Ratio.

VII. METHODOLOGY

AI-based text-to-image generation involves creating images from textual descriptions using machine learning techniques. The process typically includes the following steps:

1. **Text Processing:** The input text is analyzed and converted into a numerical format, often using embedding, to capture semantic information.
2. **Image Generation Model:** Generative models, such as Generative Adversarial Networks (GANs) or Diffusion Models, are employed to synthesize images. These models learn from large datasets of image-text pairs to understand the relationship between textual descriptions and visual elements.
3. **Training:** The model is trained on extensive datasets containing images and their corresponding textual descriptions. Through this training, the model learns to generate images that align with given textual prompts.
4. **Image Synthesis:** During inference, the trained model generates images by interpreting the input text and producing visual content that matches the description.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VIII. IMPLEMENTATION

Set Up the Development Environment: Utilize programming languages like Python and install necessary libraries such as TensorFlow or PyTorch, depending on the chosen model's framework.
Generate Images: Input textual descriptions into the integrated model to produce corresponding images. Adjust parameters like resolution and style to refine outputs.

IX. CONCLUSION

AI-based text-to-image generation has revolutionized visual content creation by transforming textual descriptions into detailed images. This technology has found applications across various sectors, including art, design, marketing, and entertainment, enabling users to generate customized visuals efficiently.

We are thankful to our Project Guide Ms. A.B.Jewlikar and Project Coordinator Mrs. K.M.Shirole for their valuable guidance, genuine suggestion and constant encouragement during preparation of project paper work without which completion of this project would be a difficult task.

REFERENCES

[1] Ramesh, A., Pavlov, M., Goh, G., et al. (2021). "Zero-Shot Text-to-Image Generation." Proceedings of the 38th International Conference on Machine Learning, 139, 8821-8831. DOI: 10.5555/3495724.3495860.
[2] Tandjoura, M., & Ghods, M. (2022). "Generative Models for Text-to-Image Synthesis: A Review." Journal of Computer Vision and Image Understanding, 214, 103309. DOI: 10.1016/j.jcviu.2022.103309.
[3] Ding, Y., & Yu, Z. (2022). "Text to Image Generation with Deep Learning: A Comprehensive Survey." ACM Computing Surveys, 54(3), Article 52. DOI: 10.1145/3482247.
[4] Brock, A., Donahue, J., & Simonyan, K. (2021). "Large Scale GAN Training for High Fidelity Natural Image Synthesis." IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(8), 2634-2646. DOI: 10.1109/TPAMI.2021.3072442.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com