**A**

**SEMINAR REPORT ON**

# “Text-To-Image Models ”

SUBMITTED TO SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

Bachelor of Engineering

in

# Information Technology Class: T.E

## Submitted By Geetesh Karjavkar :25

**Under the guidance of**

Prof. **Jayashree Surpur**



**DEPARMENT OF INFORMATION TECHNOLOGY**

**RMD SINHGAD SCHOOL OF ENGINEERING**

**WARJE, PUNE-411058**

**A.Y: 2024 - 25**



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**RMD SINHGAD SCHOOL OF ENGINEERING**

WARJE, PUNE-411058

# CERTIFICATE

This is to certify that the Seminar Report entitled

**“Text-To-Image Model”**

Submitted by

## Geetesh Karjavkar : 25

is a bonafide work carried out by him/her under the supervision of **Prof. Jayashree Surpur** and it is submitted towards the partial fulfillment of the requirement for T.E (Information Technology) – 2019 Course of Savitribai Phule Pune University, Pune in the academic year 2024-2025.

**(Prof. Jayashree Surpur)** (**Prof. Saurabh Parhad)**

Guide Head,

Department of Information Technology Department of Information Technology

**(Dr. V.V. Dixit)**

Principal,

RMD Sinhgad School of Engineering Pune – 58

Place: Pune Date:

**ACKNOWLEDGEMENT**

I take this opportunity to thank our Seminar Guide Jayashree Surpur Mam and the Head of the Department Saurabh Parhad sir for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this seminar report. I am also thankful to all the staff members of the Department of Information Technology of RMDSSOE Warje, Pune for their valuable time, support, comments, suggestions, and persuasion. I would also like to thank the institute for providing the required facilities, Internet access, and important books.

**Geetesh Karjavkar**

**ABSTRACT**

Text-to-image models represent a remarkable advancement in the field of artificial intelligence (AI) and deep learning, enabling the generation of visual imagery from textual descriptions. These models harness the power of neural networks, specifically leveraging generative adversarial networks (GANs), diffusion models, or transformer-based architectures like DALL·E, to translate human language into coherent, detailed, and contextually accurate images. The rapid evolution of these models has significantly enhanced their ability to capture nuances in text, such as emotions, spatial relationships, and intricate details, leading to high-quality visual outputs that closely align with the input prompts.

This report delves into the working principles of text-to-image models, starting with an overview of natural language processing (NLP) techniques that convert textual data into representations that machines can process. It further explores the role of image generation techniques, including how models learn to associate words with visual features during training on vast datasets comprising paired text and images. A key focus is placed on the architecture and functioning of prominent text-to-image models like DALL·E, Imagen, and Stable Diffusion, detailing the underlying processes of encoding and decoding that facilitate accurate image synthesis.

The report also addresses the challenges and limitations associated with these models, such as biases in training data, the trade-off between image quality and diversity, and the complexity of generating abstract or highly specific content. Furthermore, ethical considerations, including issues related to copyright, the potential misuse of generated content, and bias in model outputs, are examined in depth.

This report concludes with a discussion of future directions, emphasizing ongoing research aimed at improving model accuracy, reducing biases, and expanding the creative potential of text-to-image systems.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Title Of Chapter** | **Page No.** |
| **1.** | **Introduction** | **1** |
| **2.** | **Literature Survey** | **2** |
| **3.** | **Motivation** | **5** |
| **4.** | **Objectives** | **7** |
| **5.** | **Purpose** | **8** |
| **6.** | **Scope** | **10** |
| **7.** | **Detail of Design** | **13** |
| **8.** | **Technology** | **16** |
| **9.** | **Conclusion** | **24** |
| **10.** | **References** | **25** |
| **11.** | **Plagiarism Check Report** | **26** |
| **12.** | **Base Paper** | **22** |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Title of Figure** | **Page No.** |
| 1. | The Overall Framework of VPD | 13 |
| 2. | Neural Networking Architecture | 17 |
| 3. | CLIP Architecture | 20 |
| 4. | Diffusion Model Architecture | 22 |
| 5. | Diffusion Model Architecture | 23 |