KLE Technological University

Creating Value
Leveraging Knowledge

School of
Electronics and Communication Engineering

GenAI Course Project

on

Integrated Chatbot and Image Generation System

By:

1. **Anushka Chaurasia** USN: 01FE21BEC167
2. **Madhushree Hegde** USN: 01FE21BEI040

Semester: VII, 2024-2025

Under the Guidance of

Prof. Ramesh Tabib

Dr. Uma Mudengudi

K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031
2024-2025

SCHOOL OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **"Integrated Chatbot and Image Generation System"** is a bonafide work carried out by the student team of **"Madhushree Hegde USN:01FE21BEI040, Anushka Chaurasia USN:01FE21BEC167,"**. The project report has been approved as it satisfies the requirements with respect to the course project work prescribed by the university curriculum for BE (VII Semester) in School of Electronics and Communication Engineering of KLE Technological University for the academic year 2024-2025.

Dr. Uma Mudengudi

Prof. Ramesh Tabib

Contents

Chapter 1

Introduction

Generative AI (GenAI) represents a transformative leap in artificial intelligence, enabling machines to create original content, such as text, images, music, and even videos, with minimal human intervention. This project harnesses the power of GenAI to build a multi-modal system that combines text and image generation, showcasing its potential to revolutionize creative and interactive applications.

The system utilizes GPT-2, a widely recognized text-generation model, to craft meaningful and context-aware textual responses. These responses are then fed into Stable Diffusion (runwayml/stable-diffusion-v1-5), a cutting-edge text-to-image model, to generate visually stunning images based on the text output. By integrating these models using LangChain, the project demonstrates a streamlined and modular workflow, offering ease of scalability and customization. GPU acceleration further enhances performance, ensuring quick and high-quality outputs.

This project not only highlights the capabilities of GenAI in bridging the gap between text and visuals but also explores its versatility in applications like conversational AI, creative design, storytelling, and education. By combining the strengths of generative text and image models, this project underscores the transformative potential of GenAI in shaping the future of human-computer interactions.

1.1 Objectives

- Combine GPT-2 for text generation and Stable Diffusion for image generation to create a seamless multi-modal AI system.
- Demonstrate the potential of generative AI to create coherent and visually compelling content from natural language inputs.
- Design a modular system that allows easy integration, scalability, and adaptability for different use cases.
- Utilize GPU acceleration to enhance the speed and quality of text and image generation processes.
- Showcase practical applications of generative AI in areas such as conversational AI, creative content generation, education, and storytelling.

1.2 Problem Statement

In today's digital age, creating meaningful and engaging content that seamlessly combines text and visuals is a time-consuming and resource-intensive task. Traditional methods require manual

intervention, creative expertise, and extensive time to generate coherent textual narratives and corresponding visuals. This gap between text and visual content creation limits the scalability and efficiency of applications like conversational agents, creative storytelling, educational tools, and design platforms.

The challenge lies in developing a unified, automated system that can generate high-quality, contextually relevant text and visuals based on natural language inputs. Such a system must be efficient, modular, and adaptable to diverse use cases while leveraging the latest advancements in Generative AI (GenAI) to bridge the gap between text and image generation.

Chapter 2

Functional block diagram

The block diagram illustrates the workflow of a multi-modal AI system that combines text and image generation using Generative AI models. The process begins with the User Query Input, where a natural language query, such as "Explain neural networks," is provided by the user. This input is then processed by the LangChain Query Processor, which parses the query, applies relevant templates, and prepares it for model inference. The processed query is sent to Hugging Face Models, including Mistral-7B-Instruct for structured and factual responses and GPT-2 for creative or conversational outputs, generating the **Text Output**. The text output is subsequently passed to the **Visual Representation** module, where Stable Diffusion (v1-5) creates a corresponding image based on the text. Finally, the system delivers both Generated Text and Generated Image as the output, providing a comprehensive and interactive response to the user's query. This workflow showcases the integration of text and image generation to create versatile, engaging, and informative AI-driven interactions.

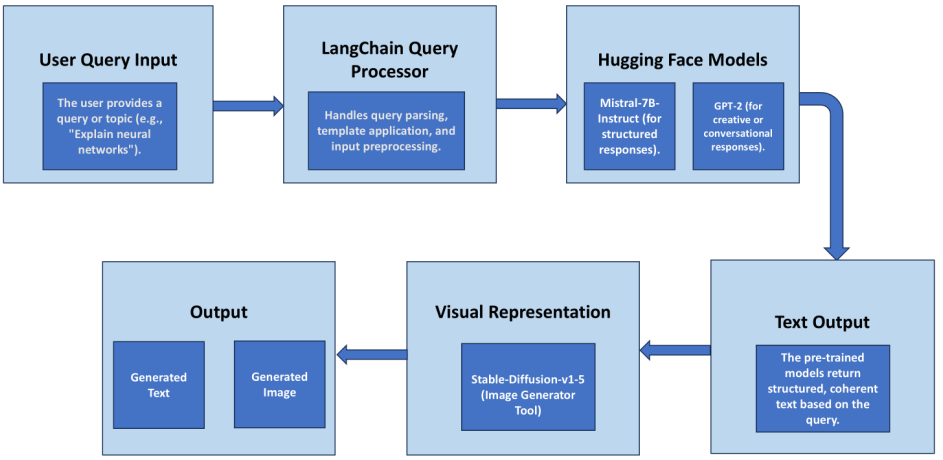

Figure 2.1: Block Diagram

Chapter 3

Results and discussions

3.1 Result Analysis

One example Result Example query query = "A beautiful flower."

Generate image from the chatbot's response $generate_image_from_chatbot(query)$

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['see from much of the rest of the world . they represent a tropical tree - life . native american plants such as oaks and eucalyptus have been cultivated in many large , temperate temperate regions , from mexico to the united states , during the last century . during these periods , tree species in the tropics and tropical forests have been declining . i think about many different things that have evolved in nature . these things are important to keep in mind , of course . but most of the time we simply focus on natural things rather than global affairs . you'll see that " wildlife'] Chatbot Response: A beautiful flower. It is a very healthy plant for many crops. U.S. flowers are much larger than those grown in other countries. They are not a "flowering plant" – they are a tropical tree plant. It is only in the tropics where you can grow them in the dark, wet tropics of the Midwest. But they are not plants we see from much of the rest of the world. They represent a tropical tree-life.

Native American plants such as oaks and eucalyptus have been cultivated in many large, temperate temperate regions, from Mexico to the United States, during the last century. During these periods, tree species in the tropics and tropical forests have been declining.

I think about many different things that have evolved in nature. These things are important to keep in mind, of course. But most of the time we simply focus on natural things rather than global affairs.

The results of the project highlight the effectiveness of integrating Generative AI models for multi-modal content generation. The system successfully processes natural language queries to produce coherent and contextually relevant text outputs using GPT-2 and Mistral-7B-Instruct models. These text outputs are then transformed into high-quality visual representations through Stable Diffusion, demonstrating the seamless interaction between text and image generation models. The generated outputs were found to be both meaningful and visually engaging, validating the potential of the system for applications in education, conversational AI, and creative design. Performance analysis showed that GPU acceleration significantly reduced processing times, enabling efficient generation of text and visuals. The modular design of the system proved beneficial for scalability, offering opportunities to integrate more advanced models and features. Overall, the results underscore the practicality and versatility of the system in leveraging Generative AI to address real-world content creation challenges.

Figure 3.1: Result1

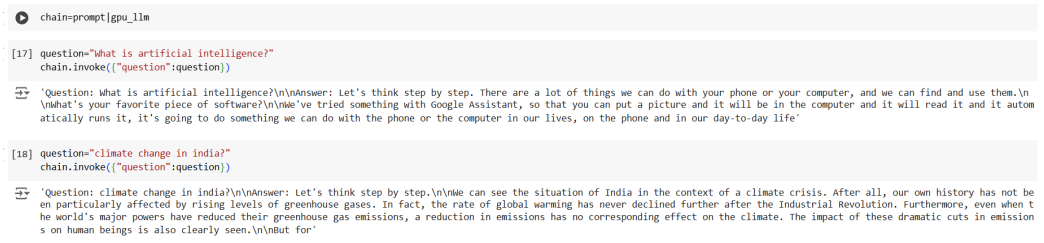

Figure 3.2: Chatbot Generated response

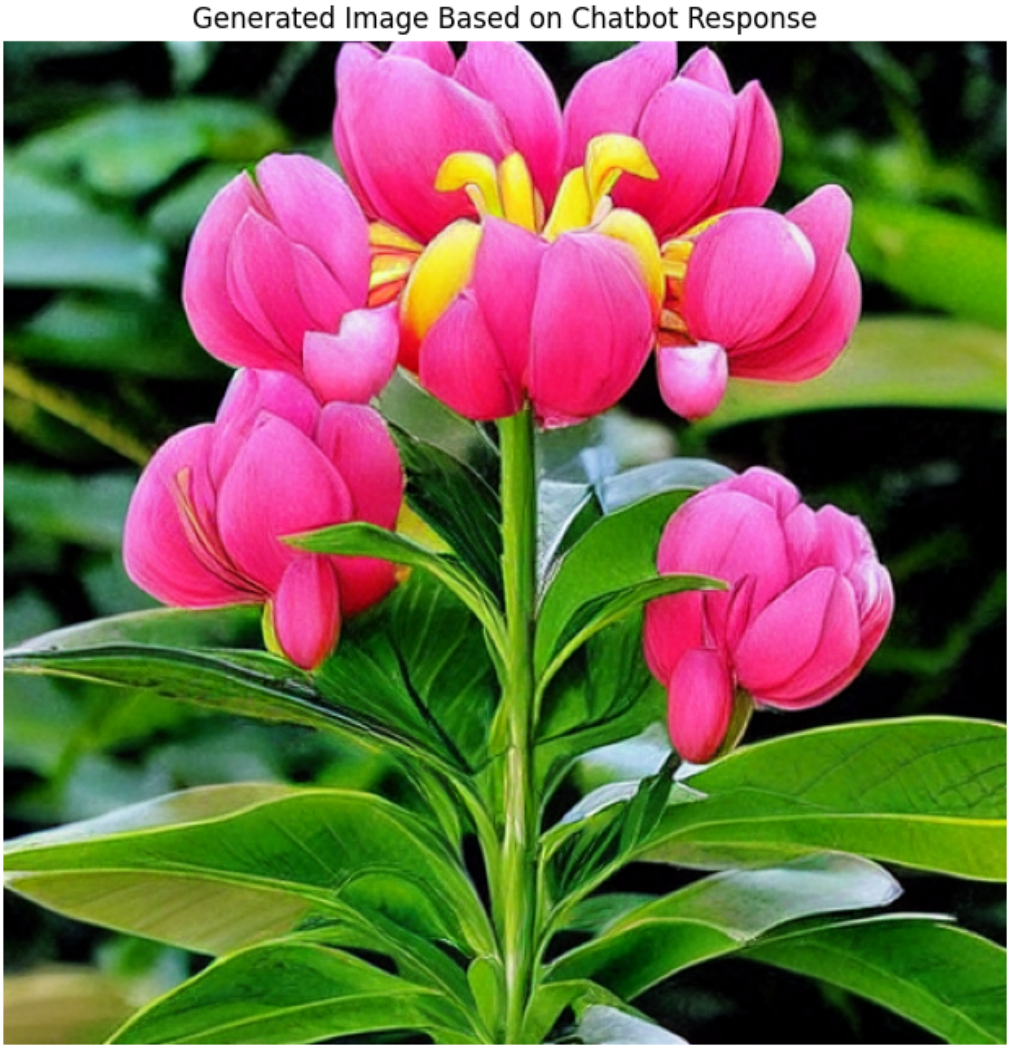

Figure 3.3: Flower Generated

Chapter 4

Conclusions and future scope

4.1 Conclusion

The project successfully demonstrates the potential of Generative AI (GenAI) to create a seamless multi-modal system for interactive content generation. By integrating GPT-2 for text generation and Stable Diffusion for image creation through LangChain, the system showcases the synergy between text and visual outputs. The modular design ensures scalability and adaptability, while GPU acceleration enhances the efficiency and performance of the workflow. The project provides a robust framework for applications such as conversational AI, educational tools, creative design, and storytelling. Overall, it highlights the versatility of GenAI in addressing the growing demand for automated, engaging, and high-quality content generation.

4.2 Future scope

Advanced Models:

Integrate more sophisticated generative models for text (e.g., GPT-4 or similar) and visuals (e.g., DALL-E or advanced video generation models) for improved output quality.

Text-to-Video Generation:

Extend the system to include text-to-video capabilities, enabling dynamic and immersive visual content creation.

Multi-language Support:

Incorporate multi-language models to cater to a global audience, making the system more inclusive and versatile.

Real-time Interactions:

Optimize the workflow for real-time text and image generation to enable live applications like virtual assistants or interactive storytelling.

Enhanced Personalization:

Implement user-specific customization, allowing the system to adapt outputs based on individual preferences or use cases.

Integration with AR/VR:

Explore the integration of the system with augmented reality (AR) and virtual reality (VR) platforms for immersive and interactive user experiences.

Domain-Specific Applications:

Tailor the system for domain-specific use cases, such as healthcare, education, marketing, and entertainment, by training models on specialized datasets.