AI Multimodal Communicator:

An AI streamlit Web Application for Conversations, Image Captioning, and PDF Interaction

A Project Report

submitted in partial fulfillment of the requirements

of

Industrial Artificial Intelligence with Cloud Computing

by

Suthar Komal Mahavir, 210841102058

Under the Esteemed Guidance of

Abdul Aziz Md, Master Trainer, Edunet Foundation.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have contributed to the completion of this project. Without their support, guidance, and encouragement, this endeavor would not have been possible.

First and foremost, I extend my heartfelt thanks to **Abdul Aziz Md, Master Trainer**, **Edunet Foundation**, whose expertise and mentorship provided invaluable insights throughout the duration of this project. Their unwavering support and commitment played a pivotal role in shaping its outcome.

I am also immensely grateful to **Dr. Latesh B. Chaudhari,** Principal, RNGPIT, and **Mr. V. C. Joshi,** Head of Department, and **Miss. Foram Sukla,** Faculty of Department, Bachelor of Vocation-Software Development for their dedication and contributions at various stages of the project. Their collaborative efforts, feedback, and constructive criticism significantly enhanced the quality and effectiveness of our work.

I would also like to acknowledge the support of and friends, who provided encouragement, understanding, and patience throughout this journey. Their unwavering belief in me served as a constant source of motivation, and for that, I am truly grateful.

Thank you to everyone who contributed in any way, no matter how big or small. Your support has been invaluable, and I am deeply appreciative of the opportunity to work alongside such talented and dedicated individuals.

Sincerely,

Suthar Komal Mahavir (210841102058)

ABSTRACT

This project presents a comprehensive Generative AI application that leverages advanced natural language processing, computer vision, and multimodal AI techniques to empower users with intuitive interactions and creative possibilities. It encompasses a range of features, including a conversational Chabot, image captioning, question answering, document analysis, and content generation capabilities. The application is designed to facilitate seamless communication, provide informative responses, and generate unique content based on user inputs. It seamlessly integrates with existing knowledge bases and external applications, offering a versatile platform for collaboration, problem-solving, and knowledge discovery. It offers a powerful tool for individuals and organizations to harness the capabilities of AI, creativity, and knowledge. By continuously exploring new advancements in generative AI, this project aims to push the boundaries of what is possible and empower users to unlock their full potential.

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1. Problem Statement:

Current content creation and human-computer interaction platforms lack a comprehensive solution that seamlessly integrates advanced AI techniques. Existing platforms often focus on specific tasks or domains, limiting their applicability and hindering users from unlocking their full creative potential. There is a need for a transformative Generative AI platform that provides a unified interface for various AI-powered functionalities, enabling users to engage in natural conversations, generate creative content, extract insights from complex documents, and enhance their interactions with technology in a seamless and efficient manner.

1.2. Problem Definition:

Data Acquisition and Management: Gathering and integrating diverse datasets encompassing text, images from various sources. Ensuring data quality, relevance, and accessibility for training and deployment.

Conversational AI and Natural Language Understanding: How to design a conversational AI interface that enables natural and intuitive interactions with users, leveraging advanced natural language processing techniques to interpret user intent and generate appropriate responses.

Model Content Generation and Quality Control: How to provide users with tools and features to generate unique and high-quality content, while addressing issues related to content diversity, originality, and potential biases in the generative models.

Model Evaluation: To establish a comprehensive evaluation framework that enables developers and users to accurately assess the performance and responsible use of generative AI models. Developing robust evaluation metrics and methodologies to assess the performance and quality of generative AI model.

Deployment and Integration: Implement the generative AI models into a user-friendly application or platform where users can input chat, question, pdf and receive answer through Chatbot or pdf chatbot. Ensure seamless integration with platforms for practical usability.

1.3. Expected Outcomes:

This platform empowers users with intuitive and creative capabilities through a comprehensive suite of AI techniques. By leveraging natural language processing, users can engage in seamless conversations with a chatbot, obtain detailed answers to their questions, and extract information from uploaded PDFs through a chat-based interface. Additionally, the platform provides advanced image captioning capabilities. This versatile platform enables users to unlock their creativity, enhance their interactions with technology, and drive innovation across diverse domains such as education, healthcare, entertainment, and business.

Keywords: Chatbot, Ask Me Anything, Chat with PDF, Image Captioning, Generative AI, Natural Language Processing, Machine Learning, Artificial Intelligence, Content Creation, Conversational AI, Document Understanding, Image Analysis, Creative Tools.

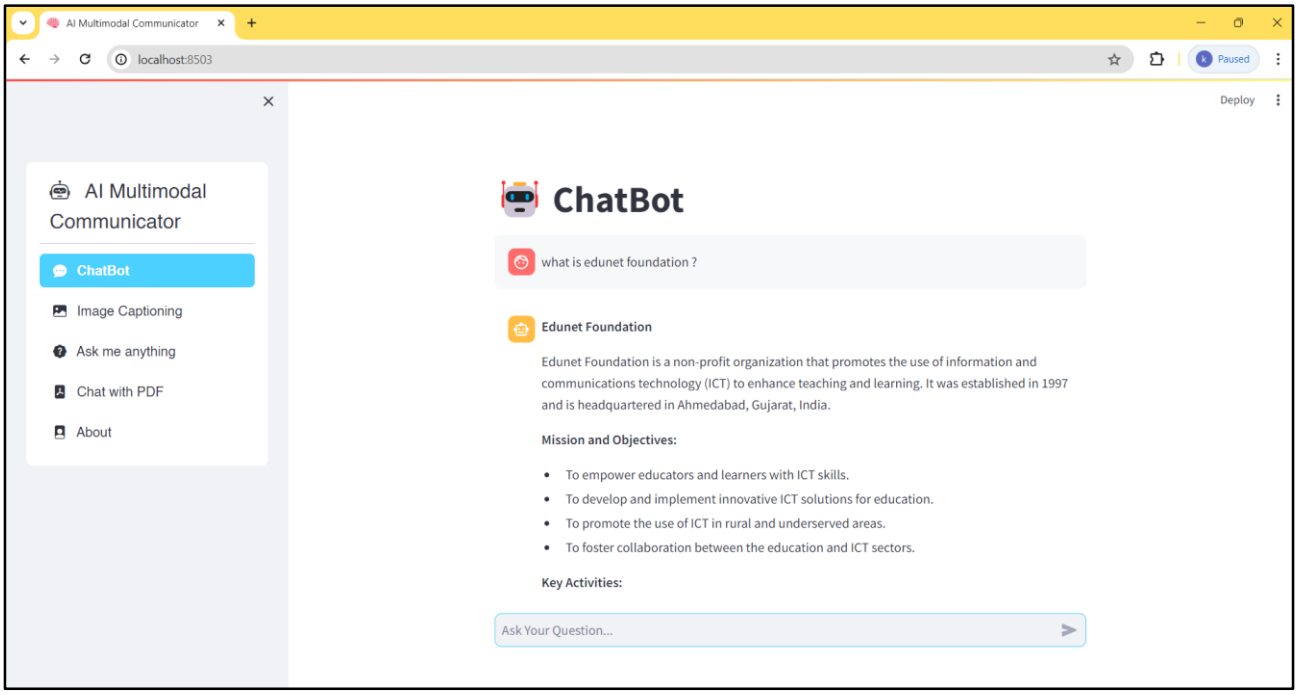


Figure 1.3.1 ChatBot

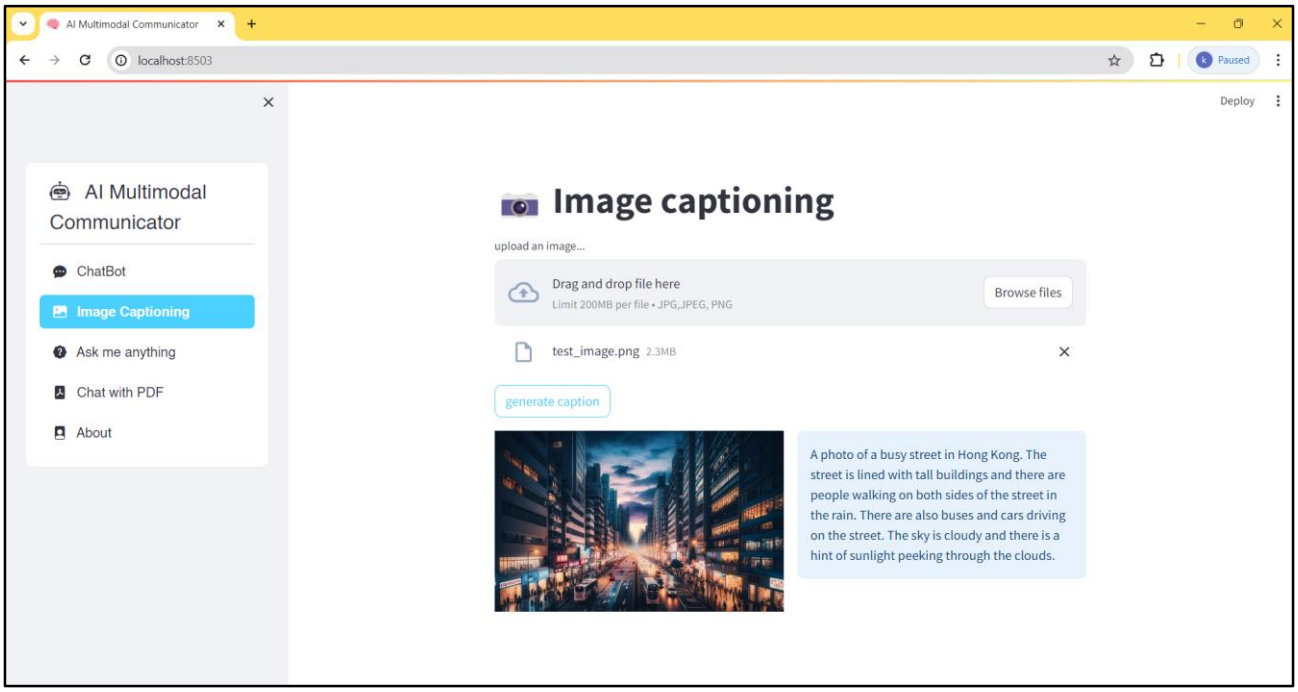


Figure 1.3.2 Image Captioning

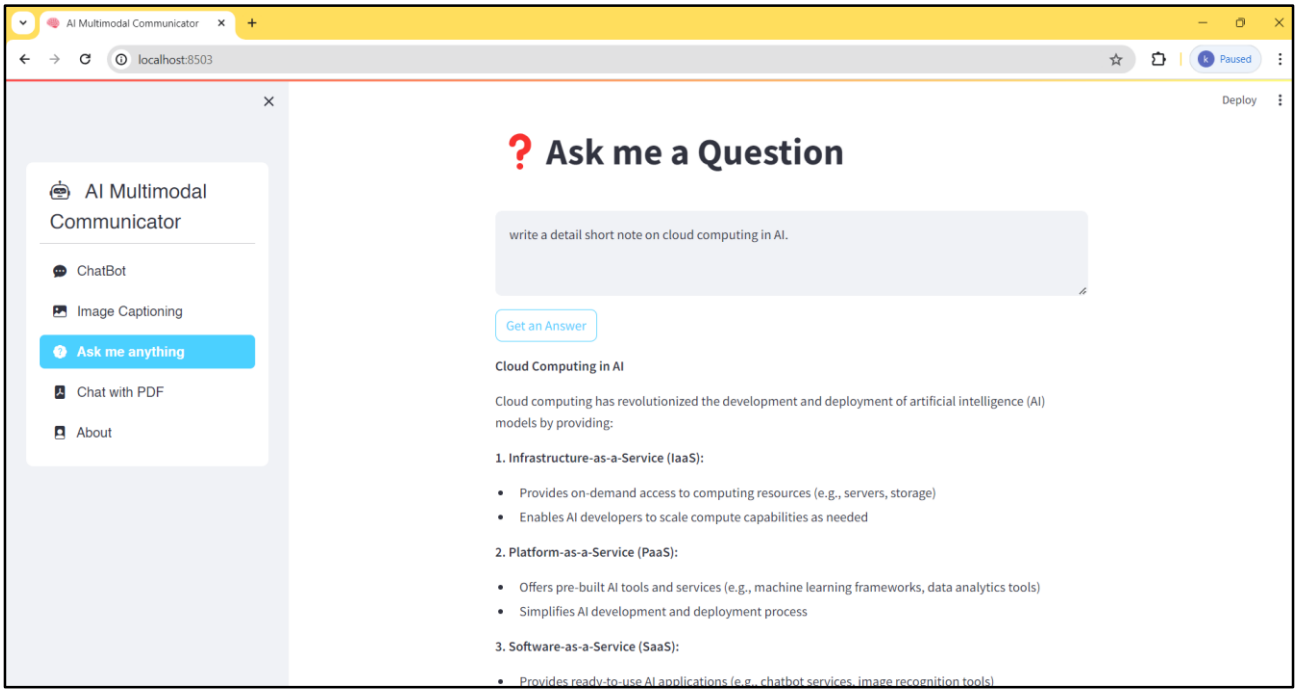


Figure 1.3.3 Ask me a Question

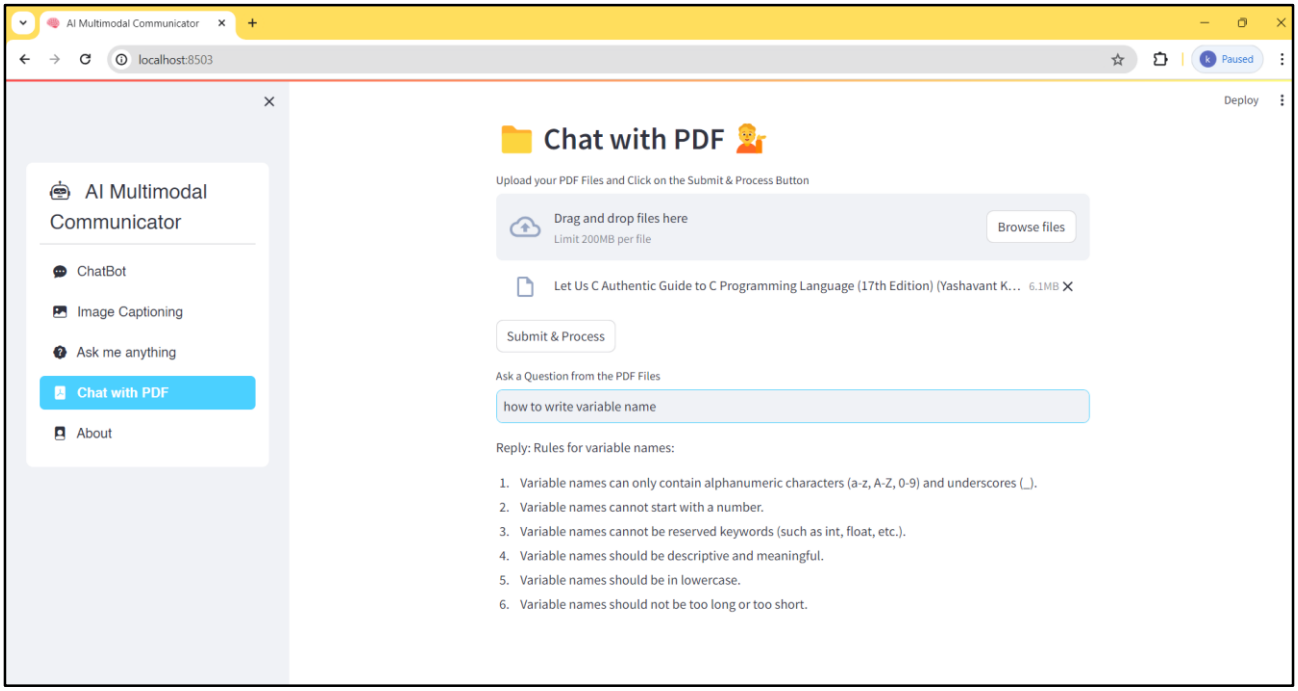


Figure 1.3.4 Chat with PDF

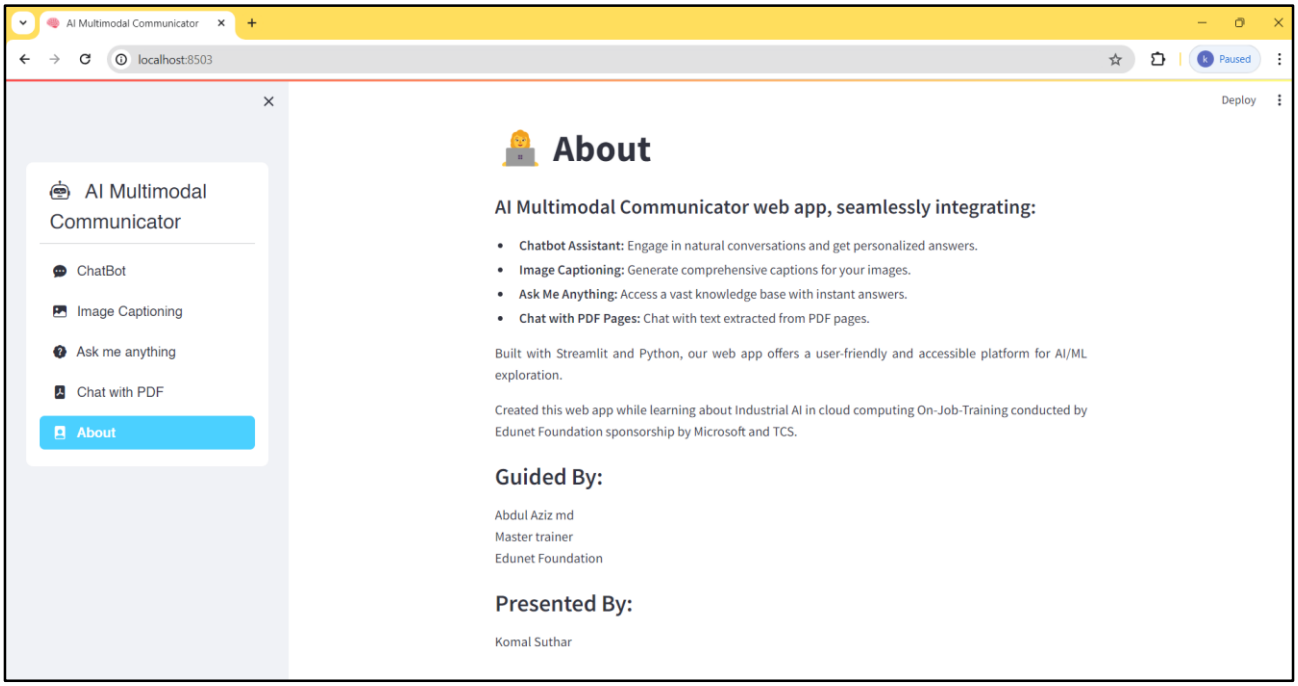


Figure 1.3.5 About

1.4. Organization of the Report

- The proposed methodology flow diagram illustrates the sequential steps involved in developing the generative AI web app. Each step in the flow diagram represents a key stage in the project's lifecycle, from collecting user input for data generation to model deployment. Advantages of techniques, this section describes the advantages of using generative AI techniques for developing the web app, such as python, streamlit, PyPDF2, Langchain, FAISS including improved customer engagement, personalized content, and increased efficiency.

- The implementation phase of the project involves translating the proposed methodology into executable code, including model development, and user interface design. This section documents the results of the evaluation, highlighting the web app's effectiveness in performing its functions and providing valuable insights for users.
- The conclusion summarizes the project's successful implementation of a generative AI web app, highlighting its effectiveness and potential implications for users in performing various task very easily. This section explores avenues for further development, such as expanding features, integrating with more AI model, and exploring advanced generative AI algorithms.
- This section acknowledges the sources consulted during the project, contributing to its completeness and providing resources for further exploration.

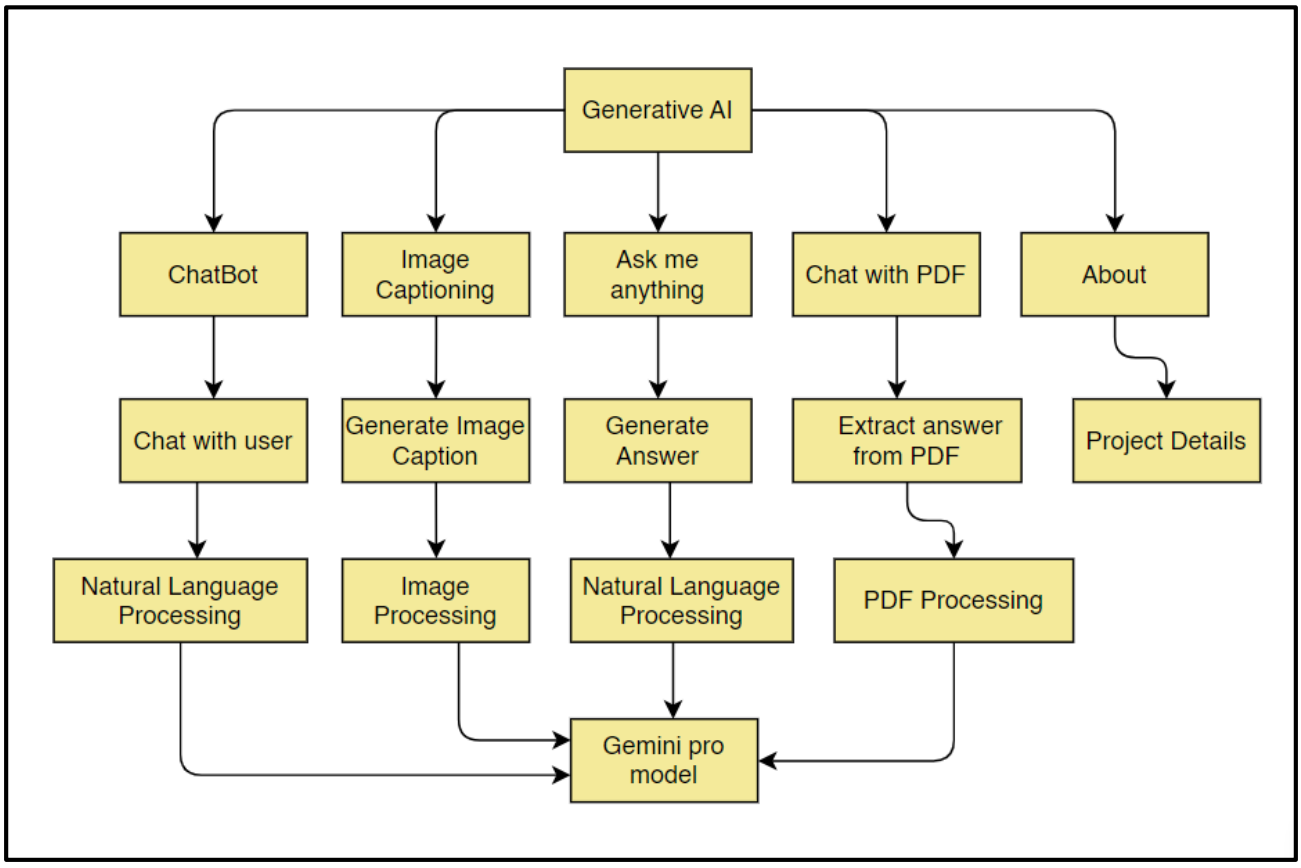
CHAPTER 2

PROPOSED METHODOLOGY

CHAPTER 2

PROPOSED METHODOLOGY

2.1 Data Flow Diagram



2.2 Advantages

- **Comprehensive Functionality:** The code combines multiple generative AI features into a single web app, offering a wide range of capabilities for users.
- **User-Friendly Interface:** The web app is built with Streamlit, providing a user-friendly and intuitive interface that is accessible to users with varying technical backgrounds.
- **State-of-the-Art Models:** The code utilizes state-of-the-art generative AI models from Google, ensuring accurate and high-quality results. Integrates with Google's knowledge base, enabling access to a vast amount of information and comprehensive answers. Utilizes the "gemini-pro" model, known for its advanced language generation capabilities and ability to engage in natural conversations.
- **Customizable:** The code allows for customization of the generative AI models and prompt templates, enabling users to tailor the web app to their specific needs. Leverages the "gemini-pro-vision" model, specifically designed for image captioning, providing accurate and descriptive captions. Employs the "stuff" chain type, which is optimized for extracting information from document-based contexts.
- **Community Support:** The project leverages open-source libraries and resources from the Langchain community, ensuring ongoing support and contributions.

2.3 Requirement Specification

2.3.1 Hardware Requirements:

Hardware Requirements	Description
Processor	Multi-core processor with a clock speed of 2.0 GHz or higher for efficient computation.
GPU	Dedicated graphics card with 4 GB or more of VRAM recommended
RAM	Minimum 8GB RAM for basic AIML tasks; 16GB or more recommended for complex models and larger datasets.
Storage	Solid State Drive (SSD) with at least 256GB storage capacity recommended.
Network Connection	High-speed internet connection with low latency for accessing cloud platforms, downloading datasets, and collaborating on projects.

2.3.2 Software Requirements:

Software Requirements	Description
Operating System	Windows 10 or 11, macOS 10.15 or later, or Ubuntu 18.04 or later
Programming Language	Python or R for data analysis and machine learning tasks.
Integrated Development Environment (IDE)	PyCharm, Jupyter Notebook, VS Code
Machine Learning Libraries	Langchain, PyPDF2, FAISS, google-generativeai, streamlit-option-menu, streamlit, pillow, python-dotenv, langchain_google_genai
Web Development Frameworks (Optional)	Flask or Streamlit /HTML/CSS/JavaScript for frontend. Langchain, PyPDF2, FAISS for backend.
Deployment Platforms	Microsoft Azure for deploying machine learning models.
Version Control	Git for version control; GitHub, GitLab, Bitbucket for hosting repositories.

CHAPTER 3

IMPLEMENTATION AND RESULT

CHAPTER 3

IMPLEMENTATION AND RESULT

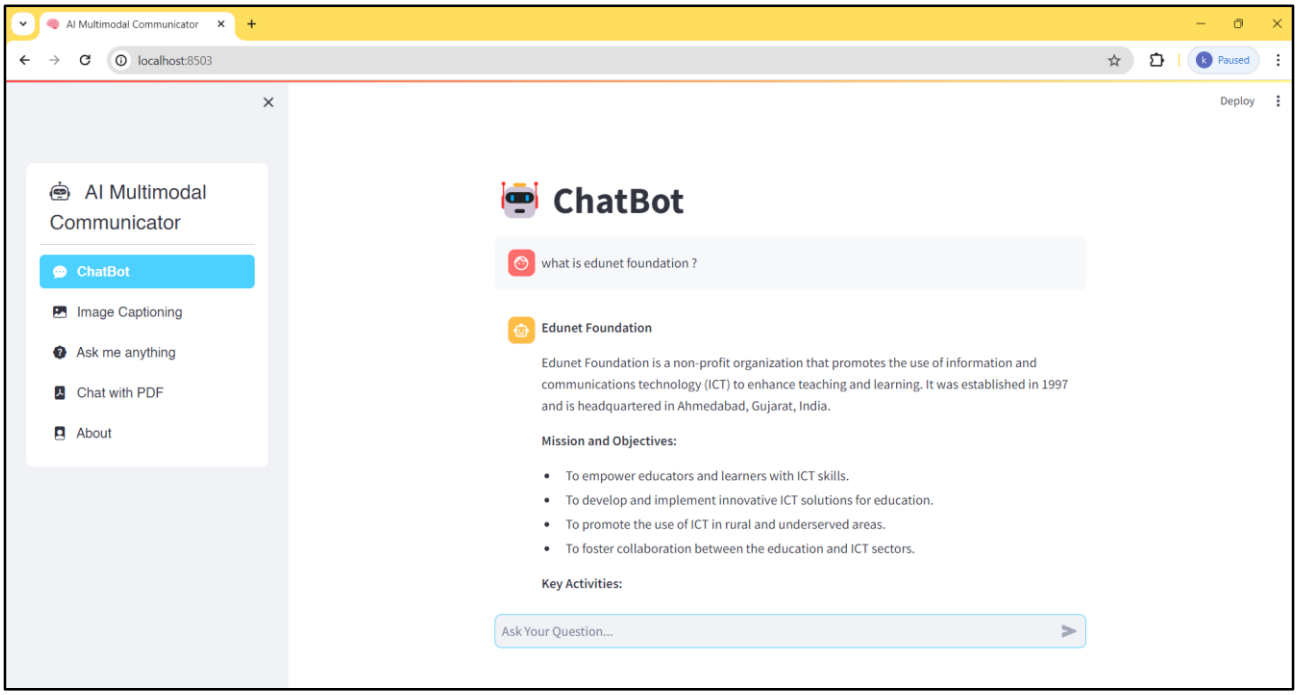


Figure 4.1 ChatBot

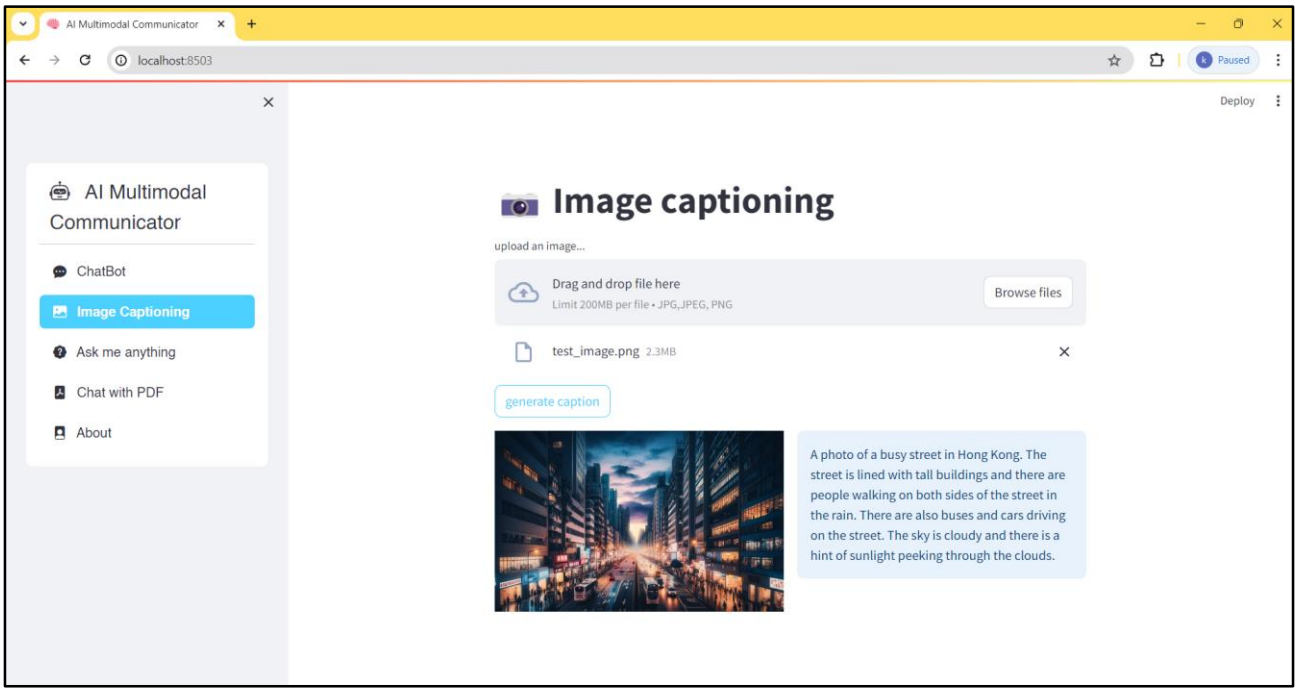


Figure 4.2 Image Captioning

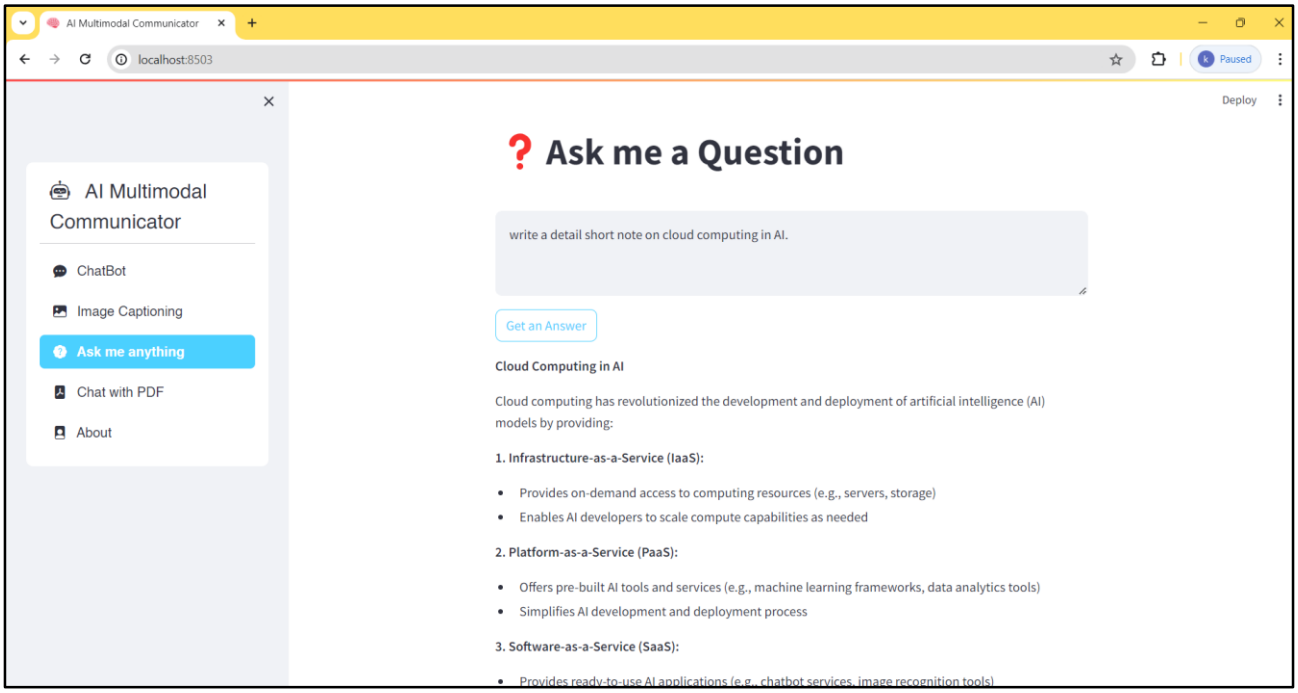


Figure 4.3 Ask me a Question

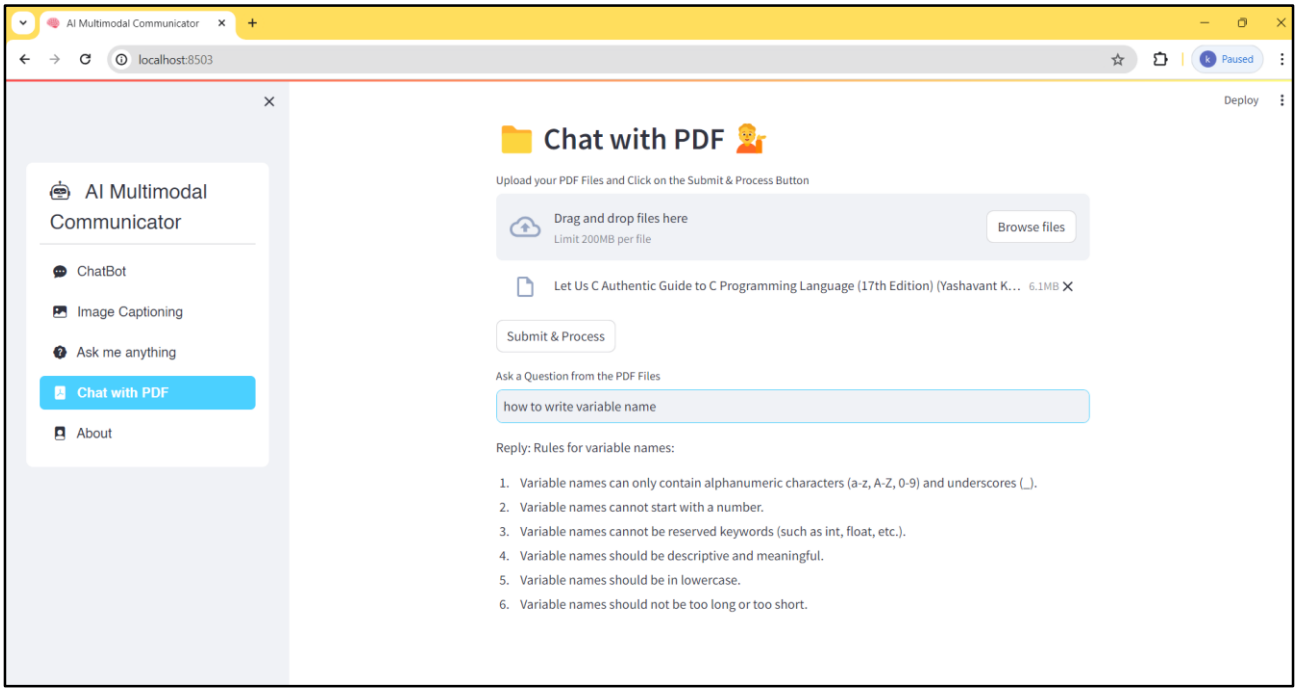


Figure 4.4 Chat with PDF

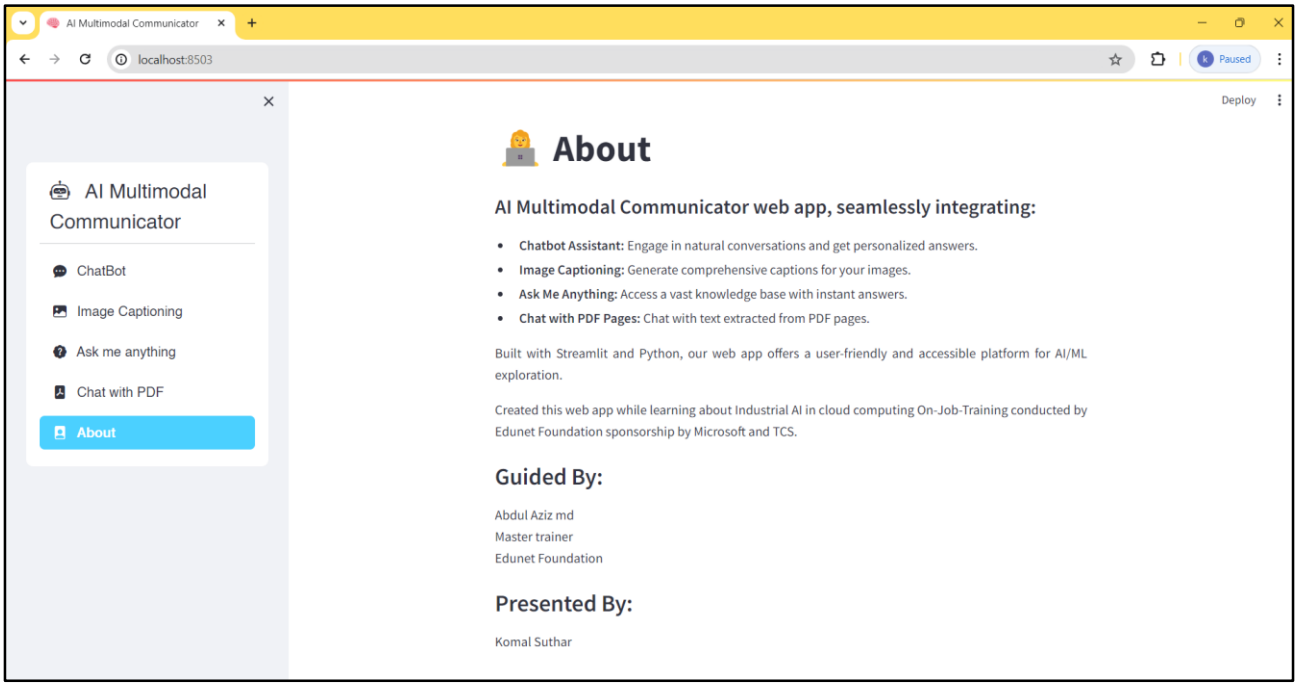


Figure 4.5 About

CHAPTER 4

CONCLUSION

CHAPTER 4

CONCLUSION

In conclusion, the AI Multimodal Communicator web app, which seamlessly integrates chatbot, image captioning, ask me anything, and chat with PDF capabilities, has been successfully developed and deployed. The web app leverages state-of-the-art AI models and provides a user-friendly interface for users to interact. The web app combines multiple generative AI features into a single platform, offering a wide range of capabilities for users. The generative AI models used in the web app provide accurate and high-quality results, ensuring reliable information and engaging interactions. The web app is designed with a user-centric approach, providing an intuitive and accessible interface for users of all technical backgrounds. The code is designed to be extensible, allowing for the easy integration of additional AI models or features in the future. The project serves as a valuable resource for learning about AI and its practical applications. The project's success demonstrates the potential of AI to transform various industries and empower users with powerful tools for communication, information discovery, and content creation.

CHAPTER 5

FUTURE SCOPE

CHAPTER 5

FUTURE SCOPE

- Integrating additional generative AI models.
- Adding new categories features to perform various other task.
- Image and Video Generation.
- Explore advanced NLP techniques like named entity recognition, conference resolution, and question classification.
- Implement safeguards and guidelines to ensure responsible and ethical use of the application.
- Developing a mobile application version of the system.

GitHub Link

https://github.com/sutharkomal16/Generative-AI.git

Video Link

https://github.com/sutharkomal16/AI_Multimodal_Communicator/blob/main/AI_multimodal_communicator.mp4

REFERENCES

1.] https://www.youtube.com/watch?v=sf5MrM0AIiU&list=PLfFghEzKVmju5URmh5TdiiSdk01G2EL25
2.] https://github.com/siddhardhan23/gemini-pro-streamlit-chatbot
3.] https://streamlit.io/components