

# **AI ASSISTANCE CHATBOT FOR GOVERNMENT SCHEMES**

**A PROJECT REPORT**

*Submitted by*

**A AAKASH (210701002)  
P ARVIND (210701034)**

*in partial fulfillment for the course*

**GE19612 - PROFESSIONAL READINESS FOR INNOVATION,  
EMPLOYABILITY AND ENTREPRENEURSHIP**

*for the degree of*

**BACHELOR OF ENGINEERING**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**RAJALAKSHMI ENGINEERING COLLEGE**

**RAJALAKSHMI NAGAR**

**THANDALAM**

**CHENNAI – 602 105**

**MAY 2024**

# **RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI**

## **BONAFIDE CERTIFICATE**

Certified that this Thesis titled “**AI ASSISTANCE CHATBOT FOR GOVERNMENT SCHEMES**” is the bonafide work of “**A AAKASH (210701002), P ARVIND (210701034)**” who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

### **SIGNATURE**

Dr.T.Kumaragurubaran., M.Tech., Ph.D

### **PROJECT COORDINATOR**

Assistant Professor

Department of Computer Science and Engineering

Rajalakshmi Engineering College

Chennai – 602 105

Submitted to project Viva-Voce Examination held on\_\_\_\_\_

**Internal Examiner**

**External Examiner**

## **ACKNOWLEDGMENT**

First, we thank the almighty god for the successful completion of the project. Our sincere thanks to our chairman **Mr. S. Meganathan B.E., F.I.E.**, for his sincere endeavor in educating us in his premier institution. We would like to express our deep gratitude to our beloved Chairperson **Dr. Thangam Meganathan Ph.d.**, for her enthusiastic motivation which inspired us a lot in completing this project and Vice Chairman Mr. Abhay Shankar Meganathan B.E., M.S., for providing us with the requisite infrastructure.

We also express our sincere gratitude to our college Principal, **Dr. S. N. Murugesan M.E., PhD.**, and **Dr. P. KUMAR M.E., PhD, Director computing and information science , and Head Of Department of Computer Science and Engineering** and our project **Dr.T.Kumaragurubaran., M.Tech., Ph.D** for his encouragement and guiding us throughout the project towards successful completion of this project and to our parents, friends, all faculty members and supporting staffs for their direct and indirect involvement in successful completion of the project for their encouragement and support.

**A AAKASH(210701002)**

**P ARVIND(210701034)**

## **ABSTRACT**

In today's digital era, navigating through the labyrinth of government schemes poses a formidable challenge for many citizens, often deterred by bureaucratic intricacies and the fragmented nature of information dissemination. In response, we present an innovative AI chatbot solution designed on the foundation of the MERN stack alongside Python-based web scraping scripts. Our solution harnesses the power of open-source transformer models, specifically tailored for natural language processing (NLP), to offer personalized guidance on comprehending, evaluating eligibility for, and engaging with various government schemes. Departing from traditional manual search methods, our chatbot automates the data retrieval process, facilitating seamless access to pertinent information. With features such as user authentication, eligibility assessment tools, and proactive notifications, our solution transcends mere information provision to foster dynamic user interaction and transparency. By amalgamating state-of-the-art AI technology, exemplified by the GPT-3.5 turbo model, with an intuitive React.js interface, we endeavor to redefine citizen engagement with public services, cultivating a culture of inclusivity and accessibility. Continual updates and refinement ensure that our chatbot remains at the forefront of delivering real-time, actionable insights, thereby empowering individuals to make informed decisions and maximize their entitlements from government schemes.

## TABLE OF CONTENTS

<b>CHAPTER NO</b>	<b>TITLE</b>	<b>PAGE NO</b>
	<b>ACKNOWLEDGEMENT</b>	<b>3</b>
	<b>ABSTRACT</b>	<b>4</b>
	<b>LIST OF FIGURES</b>	<b>7</b>
	<b>LIST OF ABBREVIATIONS</b>	<b>8</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>8</b>
1.1	RESEARCH PROBLEM	8
1.2	PROBLEM STATEMENT	9
1.3	SCOPE OF THE WORK	10
1.4	AIM AND OBJECTIVE	11
1.5	MOTIVATION	11
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>12</b>
2.1	EXISTING SYSTEM	15
2.2	PROPOSED SYSTEM	15
<b>3</b>	<b>SYSTEM DESIGN</b>	<b>16</b>
3.1	GENERAL	16
3.2	SYSTEM ARCHITECTURE DIAGRAM	17
3.3	DEVELOPMENT ENVIRONMENT	17
3.3.1	HARDWARE REQUIREMENT	17
3.3.2	SOFTWARE REQUIREMENT	17
3.4	SEQUENCE DIAGRAM	18
<b>4</b>	<b>PROJECT DESCRIPTION</b>	<b>19</b>
4.1	MODULES	19

4.1.1	DATA SCRAPING AND PREPROCESSING	19
4.1.2	SPLITTING RAW DATA INTO CHUNKS	19
4.1.3	FEEDING CONTENT TO TRANSFORMER	20
4.1.4	GETTING QUERIES FROM USER	21
<b>5</b>	<b>RESULT AND DISCUSSION</b>	<b>22</b>
5.1	FINAL OUTPUT	22
5.2	RESULT	24
<b>6</b>	<b>CONCLUSION AND SCOPE FOR FUTURE ENHANCEMENT</b>	<b>25</b>
6.1	CONCLUSION	25
6.2	FUTURE ENHANCEMENT	25
	<b>APPENDIX</b>	26
	REFERENCES	33

### **LIST OF FIGURES**

<b>FIGURE NO.</b>	<b>NAME OF FIGURES</b>	<b>PAGE NO.</b>
3.1	ARCHITECTURE DIAGRAM	15
3.2	SEQUENCE DIAGRAM	17
5.1.1	BASIC QUERIES BY THE USER	21
5.1.2	LIST QUERIES BY THE USER	21
5.1.3	SAMPLE DATASET	

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 RESEARCH PROBLEM**

The complexity and inaccessibility of information regarding government schemes often hinder citizens from fully understanding and utilizing available benefits. This research aims to address this issue by developing an intelligent chatbot that leverages natural language processing (NLP) and web scraping techniques to provide accurate, timely, and comprehensible information about government schemes. The primary research problem is to evaluate the effectiveness of an AI-driven chatbot in improving user engagement and comprehension of government schemes. Key objectives include developing a sophisticated web scraping mechanism capable of collecting comprehensive data from official government websites, meticulously preprocessing and structuring this data to optimize its utilization by an advanced AI model. Moreover, the integration of a state-of-the-art natural language processing model (GPT-4) will empower the chatbot to deliver human-like responses to user queries, enhancing the user experience. Additionally, a meticulously crafted user-friendly frontend interface will be created using React.js, ensuring seamless interactions with the chatbot. The study will measure the chatbot's performance across multiple dimensions, including accuracy, user engagement, comprehension, and efficiency, thereby providing valuable insights into its effectiveness compared to traditional methods of information retrieval. Through this research, we aim to illuminate the chatbot's potential to revolutionize user interaction with government schemes and significantly improve the overall accessibility of this critical information.



## **1.2 PROBLEM STATEMENT**

Traditional methods of disseminating information about government schemes often prove inadequate, resulting in low user engagement and limited comprehension among citizens. As the government schemes of various states are split across various sites, it's tough for a single user to meticulously collect all the schemes and compare the ups and downs themselves. Apart from that, most of the existing chatbots are not driven by natural language processing (NLP), and they make use of primitive neural networks to fetch pre-written responses, which is highly inefficient. This fragmentation of information leads to user frustration and confusion, as individuals struggle to navigate multiple sources and decipher complex eligibility criteria and application processes on their own. Consequently, there is a pressing need for a more streamlined and user-centric approach to disseminating information about government schemes, one that leverages advanced NLP techniques to deliver personalized and easily understandable guidance to citizens. This research aims to address these shortcomings by developing an intelligent chatbot that not only consolidates information from diverse sources but also interprets user queries in natural language, providing tailored and informative responses in real-time. Through this initiative, we seek to enhance user engagement, comprehension, and accessibility to government schemes, ultimately empowering citizens to make informed decisions about their entitlements and benefits.

### **1.3 SCOPE OF THE WORK**

The scope of this work encompasses the comprehensive development and meticulous evaluation of an intelligent chatbot meticulously crafted to revolutionize user interaction with vital information about government schemes. The project will embark on several pivotal phases, commencing with the meticulous collection of data through sophisticated web scraping techniques meticulously executed on official government websites. This collected data will undergo meticulous processing and structuring, ensuring its optimal utilization by a state-of-the-art natural language processing (NLP) model, specifically GPT-4. The integration of this cutting-edge NLP model will empower the chatbot to proficiently interpret user queries and provide informative responses in real-time. Moreover, the chatbot will seamlessly integrate into a user-friendly frontend interface meticulously constructed using React.js, ensuring a frictionless and intuitive user experience. Rigorous assessment of the chatbot's performance will be conducted, focusing on critical metrics such as accuracy, response time, and its capacity to adeptly adapt to new and updated information. Additionally, the project will meticulously measure the tangible impact of the chatbot on user engagement and comprehension of government schemes, providing valuable insights into its effectiveness compared to traditional methods. Ultimately, the overarching objective is to showcase the chatbot's unparalleled potential in significantly enhancing accessibility and understanding of government schemes for the average citizen. This ambitious endeavor will not only serve as a catalyst for societal empowerment but will also lay the foundation for future enhancements and widespread applications of AI-driven tools in the domain of public information dissemination.

## **1.4 AIM AND OBJECTIVE**

The primary objective of this project is to develop an AI chatbot capable of assisting users in understanding and accessing government schemes efficiently.

Specifically, the objectives include:

- Providing users with relevant information about available government schemes based on their queries.
- Determining user eligibility for specific schemes through intelligent conversation and data analysis.
- Guiding users through the application process, including necessary documents and procedures.

## **1.5 MOTIVATION**

The motivation for this project arises from the barriers citizens face in accessing and understanding government scheme information. These schemes are often complex, dispersed across various sources, and written in bureaucratic language that is difficult to comprehend. This project aims to bridge that gap using natural language processing and AI, specifically GPT-4, to develop an intelligent chatbot. This chatbot will provide users with an intuitive way to access relevant information quickly and accurately. Additionally, the project aims to show AI's potential in enhancing public access to information, fostering inclusivity and engagement with government resources.

## **CHAPTER 2**

### **LITERATURE REVIEW**

Upon development of chatbots has evolved significantly, with advancements in artificial intelligence (AI) and natural language processing (NLP) leading to more sophisticated and user-friendly systems. Early chatbots, such as ELIZA and PARRY, were rudimentary and relied on simple pattern matching techniques. These early systems highlighted the potential of chatbots but also underscored their limitations in understanding and generating natural language.

In recent years, transformer models, particularly those based on the Transformer architecture introduced by Vaswani et al. (2017), have revolutionized NLP. Models such as BERT (Bidirectional Encoder Representations from Transformers) by Devlin et al. (2018) and GPT (Generative Pre-trained Transformer) by OpenAI have set new benchmarks in language understanding and generation. GPT-3, with its 175 billion parameters, demonstrated the capability to generate coherent and contextually relevant text, paving the way for more interactive and intelligent chatbots.

Web scraping, an essential technique for data collection, has also seen significant advancements. Tools like BeautifulSoup, Scrapy, and Selenium have made it easier to extract data from websites, enabling the aggregation of vast amounts of information from diverse sources. This capability is crucial for developing chatbots that provide updated and comprehensive responses based on real-time data.

In the context of government schemes, several challenges have been identified. These include the fragmented nature of information, the use of bureaucratic language, and the lack of user-friendly platforms for accessing this information. Studies have shown that citizens often struggle to understand their eligibility for various schemes and the application processes involved. The introduction of AI-driven chatbots in this domain aims to address these challenges by providing a more intuitive and accessible interface for users.

The integration of NLP with web scraping in chatbot development offers a promising solution. By continuously updating the chatbot's knowledge base with the latest data from government websites, the system can provide accurate and timely information. This approach not only enhances user engagement but also ensures that the information provided is relevant and up-to-date.

Upon accessing the AI chatbot, users are greeted with a user-friendly interface that invites them to interact with the system. The first step involves initiating a conversation by typing or speaking their query into the chat interface. The chatbot employs NLP algorithms to analyze the user's input and generate a relevant response. Through intelligent conversation, the chatbot guides users through a series of inquiries to understand their specific needs and circumstances regarding government schemes.

Next, the chatbot leverages its access to up-to-date information gathered through web scraping from government websites to provide tailored recommendations and guidance. This includes informing users about the availability of relevant schemes, assessing their eligibility based on personal details provided, and explaining the application process step by step.

Additionally, the chatbot may prompt users for additional information or documentation required for the application.

Throughout the interaction, users have the option to ask questions, seek clarification, or request further assistance as needed. The chatbot strives to provide clear and concise responses, utilizing natural language to ensure a seamless and intuitive user experience. Upon completing the interaction, users may receive a summary of the information discussed or relevant links to government resources for further reference. This user-centric approach ensures that individuals can easily access and understand information about government schemes, empowering them to make informed decisions about their entitlements.

Moreover, the use of machine learning models like GPT-3.5 turbo for conversational capabilities significantly enhances the chatbot's ability to handle diverse queries and provide contextually accurate responses. The incorporation of user feedback into the system allows for continuous improvement of the chatbot's performance, making it more adept at handling complex queries over time.

Overall, the literature indicates that AI-driven chatbots, enhanced by advanced NLP and web scraping techniques, hold great potential in transforming how citizens interact with government schemes. By making information more accessible and understandable, these chatbots can play a crucial role in increasing public engagement and ensuring that more individuals benefit from available government resources.

## 2.1 EXISTING SYSTEM

Currently, accessing information about government schemes often involves manual browsing of government websites, which can be time-consuming and confusing for users. Moreover, determining eligibility criteria and application procedures typically requires extensive research and understanding of complex bureaucratic language. Existing chatbots may lack the sophistication to handle nuanced queries or provide accurate guidance.

## 2.2 PROPOSED SYSTEM

The proposed system aims to overcome the limitations of the existing system by leveraging AI technology to provide personalized assistance and streamline the process of accessing government schemes. Key features of the proposed system include:

**Web Scraping:** Python scripts will be used to periodically scrape data from government websites, ensuring that the chatbot has access to up-to-date information on available schemes.

**Natural Language Processing:** An Open Source Transformer model will be employed to enable the chatbot to understand and respond to user queries in natural language, providing a more intuitive and user-friendly experience.

**User Authentication:** Users may be required to authenticate themselves to access personalized information or submit applications securely.

**Eligibility Assessment:** The chatbot will analyze user inputs to determine eligibility for specific schemes, providing tailored recommendations based on individual circumstances.

## CHAPTER 3

### SYSTEM DESIGN

#### 3.1 GENERAL

In this section, we would like to show how the general outline of how all the components end up working when organized and arranged together. It is further represented in the form of a flow chart below.

#### 3.2 SYSTEM ARCHITECTURE DIAGRAM

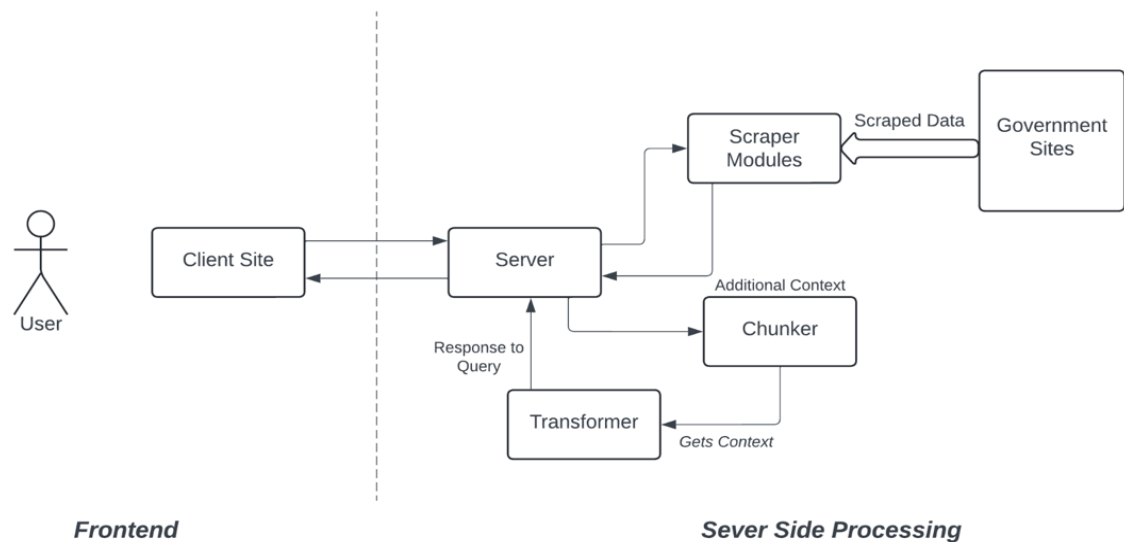


Fig 3.1: Architecture Diagram



### **3.3 DEVELOPMENT ENVIRONMENT**

#### **3.3.1 HARDWARE REQUIREMENT**

The hardware requirements may serve as the basis for a contract for the system's implementation. It should therefore be a complete and consistent specification of the entire system. It is generally used by software engineers as the starting point for the system design.

<b>COMPONENT</b>	<b>SPECIFICATION</b>
PROCESSOR	Intel Core i5
RAM	8 GB RAM
MONITOR	15" COLOR
HARD DISK	512 GB
PROCESSOR SPEED	MINIMUM 1.1 GHz

#### **3.3.2 SOFTWARE REQUIREMENT**

The software requirements document is the specifications of the system. It should include both a definition and a specification of requirements. It is a set of what the system should rather be doing than focus on how it should be done. As this project is a web-based project, all it needs is a working browser with a modern V8 engine. Additionally, this project needs Python 3.10 to be installed if the project is to be self-hosted because of the server-side scraping process done. Proper configuration of the backend environment and dependencies is also crucial for the system's seamless operation. Regular updates and maintenance will ensure its long-term functionality and security.

### 3.4 SEQUENCE DIAGRAM

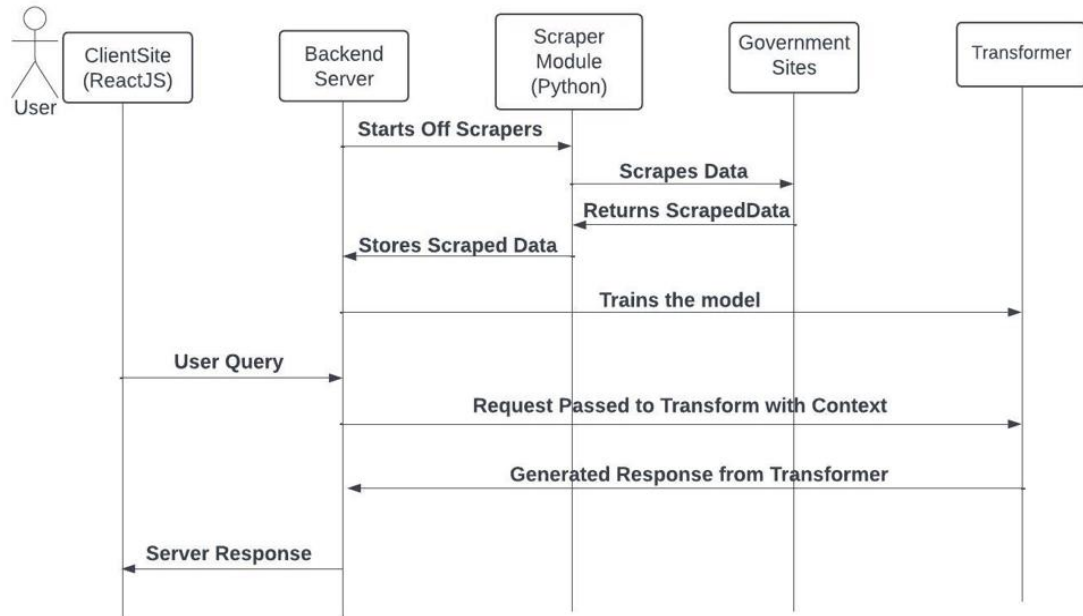


Fig 3.2: Sequence Diagram

## **CHAPTER 4**

### **PROJECT DESCRIPTION**

#### **4.1 MODULES**

##### **4.1.1 DATA SCRAPING AND PREPROCESSING**

The Scraper module is designed to automate the process of data collection from various government websites. Given the absence of a unified API endpoint for accessing information about government schemes, the Scraper module provides a solution by systematically navigating through the HTML structures of target websites using Python and the BeautifulSoup 4 library. It extracts relevant data such as eligibility criteria, benefits, and application procedures, which are then stored in a structured format such as JSON for further processing

##### **4.1.2 SPLITTING RAW DATA INTO CHUNKS**

The Chunker module is designed to process the raw JSON data generated by the Scraper module and split it into several manageable data blocks. As the data from government websites can be extensive and complex, the Chunker module ensures that it is organized into smaller, more manageable. Implemented using Python with the JSON module, this module parses the raw JSON data and divides it into discrete sections or subsets based on predefined criteria such as the number of records or the size of the data blocks. By breaking down the data into smaller chunks, the Chunker module optimizes resource utilization and enhances system performance

### **4.1.3 FEEDING CONTEXT TO TRANSFORMER:**

The Transformer module is a pivotal component of our project, leveraging the state-of-the-art capabilities of the GPT-3.5 Turbo model imported through the G4F library. This module plays a crucial role in processing the chunked data generated by the Chunker module and training the transformer model to comprehend and respond to queries about government schemes effectively. Implemented using Python, this module orchestrates the training process, feeding the chunked data as context to the pre-trained transformer model. Through iterative training iterations, the transformer model learns to analyze and interpret the nuances of government scheme-related information, thereby enhancing its proficiency in answering user queries accurately and comprehensively. Furthermore, to optimize the model's performance and specialization in government scheme-related queries, we implement fine-tuning techniques. By constraining the model's training data to exclusively focus on government schemes, we tailor its learning process to prioritize and excel in this specific domain. This fine-tuning process further enhances the model's accuracy, relevance, and responsiveness to user queries, ensuring that it becomes a reliable and valuable resource for users seeking information about government schemes. Overall, the Transformer module constitutes a pivotal component in our project, harnessing the power of advanced AI technology to facilitate seamless interaction and engagement between users and government schemes.

#### **4.1.4 GETTING REQUESTS FROM USER**

The Frontend module serves as the user interface for our project, providing a seamless and intuitive platform for users to interact with the system. Built on top of React.js, this module facilitates the exchange of information between users and the backend system.

At its core, the Frontend module features a chat interface where users can input queries related to government schemes. This chat interface is designed to be user-friendly and engaging, resembling a typical messaging platform. Alongside the chat area, a text box and a submit button are provided, enabling users to input their queries conveniently and submit them for processing.

Upon receiving a query from the user, the Frontend module sends a request to the backend, requesting the relevant response from the Transformer module. This request is typically made using asynchronous HTTP requests, allowing for seamless communication between the frontend and backend components of the system.

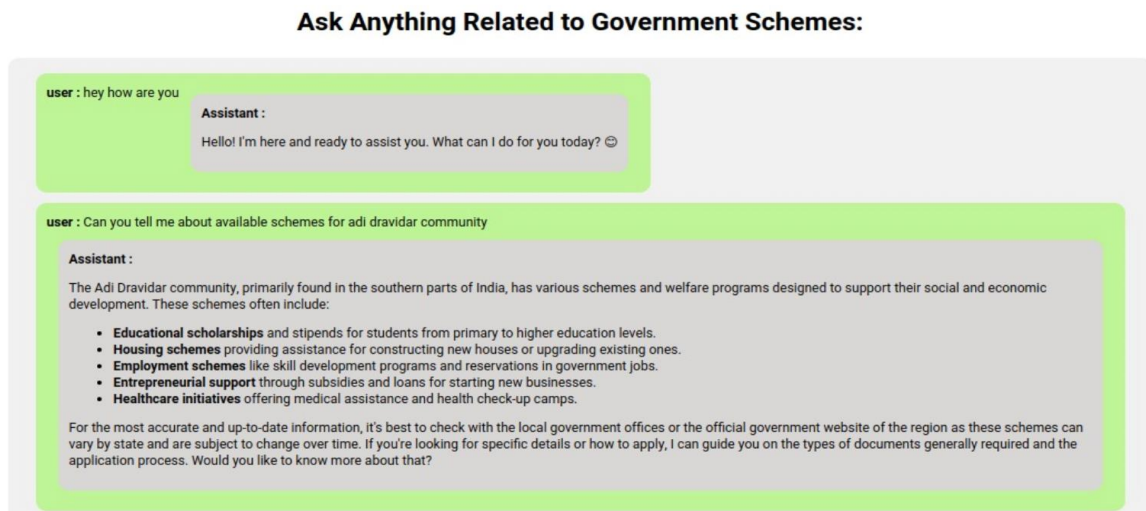
Once the response is received from the backend, the Frontend module displays it in the chat area, allowing users to view the information provided by the system. This response is presented in a conversational format, mimicking a natural dialogue between the user and the system.

## CHAPTER 5

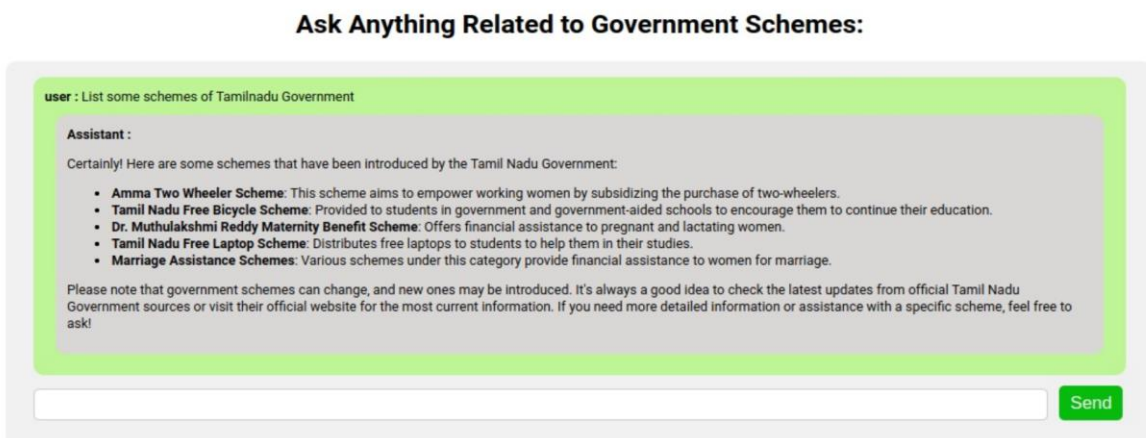
### RESULTS AND DISCUSSIONS

#### 5.1 OUTPUT

The following images contain the images of the chatbot responding user queries:



**Fig 5.1.1: Basic Querying by User**



**Fig 5.1.2: Listing Queries by User**

```

{
  "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
  "concerned_dist": "All Districts",
  "org_name": "Adi Dravidar Welfare Directorate",
  "name": "Adi Dravidar and Tribal Welfare Department -Hostels - Special Guides",
  "associated": "",
  "sponsored_by": "State",
  "funding_pattern": "",
  "beneficiaries": "Students",
  "benefits": "Hostel Facilities",
  "income": "0",
  "age": "--",
  "community": "--",
  "other": "--",
  "how": "Warden/ Matron of the hostel\r\n",
  "introduced_on": "",
  "valid_upto": "--",
  "desc": "All Boys/ Girls Boarders studying upto XII Standard\r\n"
},
{
  "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
  "concerned_dist": "All Districts",
  "org_name": "Adi Dravidar Welfare Directorate",
  "name": "Adi Dravidar and Tribal Welfare Department -Incentive / Award of Prizes - District Level Priz
  "associated": "",
  "sponsored_by": "State",
  "funding_pattern": "",
  "beneficiaries": "Students",
  "benefits": "Incentives",
  "income": "0",
  "age": "--",
  "community": "--",
  "other": "--",
  "how": "District Adi Dravidar and Tribal Welfare Officer",
  "introduced_on": "",

```

**Fig 5.1.3: Sample Dataset**

## 5.2. RESULT

The result of our project is a transformative solution that significantly enhances citizen access to government schemes. Through the integration of cutting-edge technologies such as AI-driven chatbots, web scraping, and natural language processing, we have developed a robust and user-friendly system that streamlines the process of accessing information about government programs. By automating data collection and organization, our solution eliminates the need for manual browsing of multiple websites and documents, improving accessibility and efficiency. The personalized assistance offered by our chatbot empowers users to understand eligibility criteria, benefits, and application procedures for various government programs, fostering transparency and inclusivity in citizen-government interactions. Moreover, with modular design and fine-tuning capabilities, our system is scalable and adaptable to evolving user needs and government policies. Overall, our project delivers a comprehensive and empowering solution that empowers citizens to navigate the complex landscape of government schemes with confidence and ease, ultimately contributing to a more informed and inclusive society. Furthermore, the implementation of our project not only enhances citizen access to government schemes but also facilitates efficient resource utilization and system performance. By automating data collection and organization through web scraping and chunking techniques, we optimize the handling of extensive and complex datasets, ensuring timely and accurate information retrieval. This streamlined approach not only improves user experience but also enhances the scalability and adaptability of our system to meet the evolving needs of users and government policies.



## **CHAPTER 6**

### **CONCLUSION AND FUTURE ENHANCEMENT**

#### **6.1 CONCLUSION**

In conclusion, our project successfully addresses the challenges faced by citizens in accessing and navigating government schemes through the development of a sophisticated AI-driven chatbot system. By leveraging advanced technologies such as the MERN Stack, Python-based web scraping, and pretrained transformer models for natural language processing, we have created a comprehensive and user-friendly solution that enhances accessibility, efficiency, and transparency. The system's ability to provide personalized assistance, coupled with its robust data management and processing capabilities, ensures that users receive accurate and relevant information in a timely manner. Ultimately, this project not only empowers citizens by simplifying their interaction with government schemes but also contributes to fostering a more informed and inclusive society.

#### **6.2 FUTURE ENHANCEMENT**

For future enhancements, our project aims to further expand its capabilities and impact. We plan to incorporate advanced machine learning algorithms to continuously improve the accuracy and relevance of responses provided by the chatbot. Additionally, we can bring multilingual support to cater to a broader audience. Enhancing the user interface with more interactive features and accessibility options will further improve user experience. We also aim to establish partnerships with government agencies to ensure real-time updates and comprehensive coverage of new schemes and policy changes. Moreover, by leveraging user feedback and analytics, we will refine the system to better meet the evolving needs of citizens.

## APPENDIX

### Scraper.py:

```
from bs4 import BeautifulSoup;
import requests;
import json
import re;
import os;

main_text = requests.get('https://www.tn.gov.in/scheme/departement_wise/',
verify=False).text;
main_soup = BeautifulSoup(main_text, 'lxml');
maine = main_soup.find_all('div', class_='scheme_list');
main_links = [];
for div in maine:
    p_tag = div.find('p')
    if p_tag:
        a_tag = p_tag.find('a')
        if a_tag and 'href' in a_tag.attrs:
            href_value = a_tag['href']
            main_links.append(href_value)
            print(f"Found href: {href_value}")

for m_link in main_links:
    html_text = requests.get(m_link, verify=False).text;
    soup = BeautifulSoup(html_text, 'lxml');

    full = soup.find_all('div', class_='scheme_lst');
    links = [];
    for div in full:
        p_tag = div.find('p')
        if p_tag:
            a_tag = p_tag.find('a')
            if a_tag and 'href' in a_tag.attrs:
                links.append(href_value)
                print(f"Found href: {href_value}")

    for slink in links:
        spage_text = requests.get(slink, verify=False).text;
        soup = BeautifulSoup(spage_text, 'lxml');
        full = soup.find_all('span', class_='right_column');

        ext_data = [];
        for f in full:
            ext_data.append(f.text);
```

```

extracted_data = {
    'concerned_dept': ext_data[0],
    'concerned_dist': ext_data[1],
    'org_name'      : ext_data[2],
    'name'          : ext_data[3],
    'associated'    : ext_data[4],
    'sponsered_by'  : ext_data[5],
    'funding_pattern': ext_data[6],
    'beneficiaries' : ext_data[7],
    'benefits'      : ext_data[8],
    'income'        : ext_data[9],
    'age'           : ext_data[10],
    'community'     : ext_data[11],
    'other'         : ext_data[12],
    'how'           : ext_data[13],
    'introduced_on' : ext_data[14],
    'valid_upto'    : ext_data[15],
    'desc'          : ext_data[17],
}
try:
    with open(file_path, 'r') as file:
        existing_data = json.load(file)

    if not isinstance(existing_data, list):
        existing_data = [existing_data]

except FileNotFoundError:
    existing_data = []

existing_data.append(extracted_data)
with open(file_path, 'w') as file:
    json.dump(existing_data, file, indent=2)

```

### **Chunker.py:**

```

import json
import os

with open('./Datasets/raw_data.json', 'r') as file:
    data = json.load(file)

chunk_size = 50
chunks = [data[i:i+chunk_size] for i in range(0, len(data), chunk_size)]

```

```

for i, chunk in enumerate(chunks):
    new_filename = f'../Datasets/preprompt_ds_{i + 1}.json'
    with open(new_filename, 'w') as new_file:
        json.dump(chunk, new_file, indent=2)

print(f"split into {len(chunks)} chunks.")

os.remove(f'../Datasets/raw_data.json')

```

## Api.py

```

import g4f
import sys
import json
import os

def read_file_content(file_path):
    try:
        with open(file_path, 'r') as file:
            return file.read()
    except FileNotFoundError:
        print(f"Error: File not found at {file_path}")
        sys.exit(1)
    except Exception as e:
        print(f"Error reading file: {e}")
        sys.exit(1)

# Hardcoded path to the folder containing external text files
external_files_folder = './preprompts'

# User-provided JSON data as a command line argument
user_json_data = sys.argv[1]

try:
    # Parse the user-provided JSON string into a Python data structure
    user_data = json.loads(user_json_data)
except json.JSONDecodeError as e:
    print(f"Error decoding user-provided JSON: {e}")
    sys.exit(1)

g4f.debug.logging = True # Enable debug logging
g4f.debug.version_check = False # Disable automatic version checking
print(g4f.Provider.Bing.params) # Print supported args for Bing
print(type(user_data)) # Ensure user_data is of the correct type

# Iterate over each external file in the folder

```

```

for file_name in os.listdir(external_files_folder):
    file_path = os.path.join(external_files_folder, file_name)

    # Read the content from the external file
    external_content = read_file_content(file_path)

    # Create a new element to prepend to the user-provided list
    external_element = {"role": "user", "content": external_content}

    # Add the external element to the user_data
    user_data.append(external_element)

user_data.append({"role": "user", "content": user_json_data})
# Combine the user-provided data with the additional external elements
data = user_data

# Using automatic a provider for the given model
## Streamed completion
response = g4f.ChatCompletion.create(
    # model="gpt-3.5-turbo",
    model = "gpt-4",
    provider = g4f.Provider.Bing,
    messages=data,
    stream=True,
)

for message in response:
    print(message, flush=True, end="")

```

### **Sample Dataset:**

```

[
  {
    "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
    "concerned_dist": "All Districts",
    "org_name": "Adi Dravidar Welfare Directorate",
    "name": "Adi Dravidar and Tribal Welfare Department -Hostels - Special Guides",
    "associated": "",
    "sponsored_by": "State",
    "funding_pattern": "",
    "beneficiaries": "Students",
    "benefits": "Hostel Facilities",
    "income": "0",
    "age": "--",
    "community": "--",

```

```

"other": "--",
"how": "Warden/ Matron of the hostel\r\n",
"introduced_on": "",
"valid_upto": "--",
"desc": "All Boys/ Girls Boarders studying upto XII Standard\r\n"
},
{
  "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
  "concerned_dist": "All Districts",
  "org_name": "Adi Dravidar Welfare Directorate",
  "name": "Adi Dravidar and Tribal Welfare Department -Incentive / Award of Prizes - District Level Prize plus2 Examination Rs.3000/- 10th Std First prize Rs.1000/- Second Prize Rs.500/- Third Prize Rs.300/.",
  "associated": "",
  "sponsored_by": "State",
  "funding_pattern": "",
  "beneficiaries": "Students",
  "benefits": "Incentives",
  "income": "0",
  "age": "--",
  "community": "--",
  "other": "--",
  "how": "District Adi Dravidar and Tribal Welfare Officer",
  "introduced_on": "",
  "valid_upto": "--",
  "desc": "Adi Dravidar and Tribal Welfare Department -Incentive / Award of Prizes - District Level Prize plus2 Examination Rs.3000/- 10th Std First prize Rs.1000/- Second Prize Rs.500/- Third Prize Rs.300/.One Boy and One Girl each from Adi Dravidar/Tribal / Adi Dravidar Converted to Christianity who have secured highest mark"
},
{
  "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
  "concerned_dist": "All Districts",
  "org_name": "Adi Dravidar Welfare Directorate",
  "name": "Adi Dravidar and Tribal Welfare Department -Incentive / Award of Prizes - For each subjects (c) plus2 Examination Rs.2000/- (d)10th Std Examination Rs.1000/-",
  "associated": "",
  "sponsored_by": "State",
  "funding_pattern": "",
  "beneficiaries": "Students",
  "benefits": "Incentives",
  "income": "0",
  "age": "--",
  "community": "--",

```

```

    "other": "--",
    "how": "Director of Adi Dravidar welfare, Chennai-5\r\nThrough the District
Adi Dravidar and Tribal welfare officer",
    "introduced_on": "",
    "valid_upto": "--",
    "desc": "One Student each from Adi Dravidar/Tribal/Adi Dravider converted
to Christianity for each subject"
  },
  {
    "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
    "concerned_dist": "All Districts",
    "org_name": "Adi Dravidar Welfare Directorate",
    "name": "Adi Dravidar and Tribal Welfare Department -Scholarship - Free
Education upto 12th Std. to all i.e. tuition fee will not be collected and the
amount will be reimbursed by Government.",
    "associated": "",
    "sponsored_by": "State",
    "funding_pattern": "",
    "beneficiaries": "Students",
    "benefits": "Education Benefits",
    "income": "0",
    "age": "--",
    "community": "--",
    "other": "--",
    "how": "Headmaster of the School.",
    "introduced_on": "",
    "valid_upto": "--",
    "desc": "All Adi Dravidar/Tribal / Adi Dravidar Converted to Christianity
students are eligible with no income limit\r\n"
  },
  {
    "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
    "concerned_dist": "All Districts",
    "org_name": "Adi Dravidar Welfare Directorate",
    "name": "Adi Dravidar and Tribal Welfare Department -Scholarship - Public
Examination fee for 10th and 12th Std.",
    "associated": "",
    "sponsored_by": "State",
    "funding_pattern": "",
    "beneficiaries": "Students",
    "benefits": "Education Benefits",
    "income": "0",
    "age": "--",
    "community": "--",
    "other": "--",
    "how": "Headmaster of the school\r\n",

```

```

    "introduced_on": "",
    "valid_upto": "--",
    "desc": "Public Examination fee for 10th and 12th Std. These amounts will be
reimbursed to the Directorate of Government Examinations directly by Govt..
All Adi Dravidar/Tribal / Adi Dravidar Converted to Christianity students are
eligible with no income limit"
  },
  {
    "concerned_dept": "Adi Dravidar and Tribal Welfare Department",
    "concerned_dist": "All Districts",
    "org_name": "Adi Dravidar Welfare Directorate",
    "name": "Adi Dravidar and Tribal Welfare Department - Stationary - Text
Books",
    "associated": "",
    "sponsored_by": "State",
    "funding_pattern": "",
    "beneficiaries": "Students",
    "benefits": "Education Benefits",
    "income": "0",
    "age": "--",
    "community": "--",
    "other": "--",
    "how": "Headmaster of the school\r\n",
    "introduced_on": "",
    "valid_upto": "--",
    "desc": "All student studying in 1st standard to 12th standard in Adi Dravidar
and Tribal welfare schools are eligible for free text books\r\n"
  }
}]

```



## REFERENCES

- [1] "Deep Learning" by Ian Goodfellow, Yoshua Bengio( published in 2018).
- [2] "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron (Published in 2019)
- [3] "Python Deep Learning" by Ivan Vasilev and Daniel Slater in 2017.
- [4] "Deep Learning with Python" by Francois Chollet (Published in 2017).
- [5] "Practical Python and OpenCV" by Adrian Rosebrock (Published in 2019)
- [6] "Machine Learning Yearning" by Andrew Ng (Published in 2018)
- [7] "Python Deep Learning Projects" by Matthew Lamons, Rahul Kumar, and Abhishek Nagaraja (Published in 2019)
- [8] "Blockchain and the Supply Chain Concepts, Strategies and Practical Applications" by Tiana Laurence (Published in 2019)