CONVERSATIONAL AI FOR CLIMATE CHANGE

PHASE I REPORT

Submitted by

MATHAVAN S

210701154

MADESH A

210701137

In partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING IN

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

2024

ANNA UNIVERSITY , CHENNAI

BONAFIDE CERTIFICATE

Certified that this Report titled **"CONVERSATIONAL AI FOR CLIMATE CHANGE"** is the Bonafide work of **MATHAVAN S(210701154),MADESH A (210701137)** 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	**SIGNATURE**
Dr Kumar P, M.E Ph.D	Dr. Srinivasan N , M.E Ph.D
Professor and Head	Professor
Department of Computer Science and Engineering	Department of Computer Science and Engineering
Rajalakshmi Engineering College	Rajalakshmi Engineering College
Chennai - 602105	Chennai - 602105

Submitted to Project phase I Viva-Voce Examination held on _____________

Internal Examiner **External Examiner**

TABLE OF CONTENTS

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E**., our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.**, and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D.**, for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution. Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.**, our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P.KUMAR, Ph.D**., Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr .N. Srinivasan ,Ph.D,** Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator, **Dr.T.Kumaragurubaran, Ph.D,**Associate professor,Department of Computer Science and Engineering for his useful tips during our review to build our project.

MATHAVAN S 210701154

MADESH A 210701137

ABSTRACT

This project presents a conversational AI designed specifically to address and provide insights into climate change. The primary factors contributing to climate change include carbon emissions, temperature fluctuations, precipitation levels, and other environmental parameters. The core objective of this project is to leverage various predictive analysis models and integrate them seamlessly with the conversational AI to create an interactive, intelligent assistant capable of providing actionable insights.Climate change is one of the most significant drivers of unsustainable development, posing challenges not only to ecosystems but also to industries and communities worldwide. Recognizing this, many industries have identified sustainable development as a critical priority. However, for industries and the public to effectively address and mitigate the adverse effects of climate change, access to accurate, real-time climate data is essential. This project aims to bridge that gap by providing an interface that allows real-time updates and management of climate data, ensuring its availability and relevance for informed decision-making.The forecasting of time-series data for various climate-related parameters is carried out using advanced machine learning models. These models predict trends and patterns, enabling the assistant to provide detailed forecasts and analyses. Furthermore, the system integrates this data into the conversational AI assistant, ensuring that users can interactively query and receive insights about specific climate concerns. For instance, the assistant can inform users about the carbon emissions in a particular location, detail the associated harmful effects, and suggest actionable measures for reducing emissions and promoting sustainable practices.Additionally, the assistant is designed to serve as an educational and awareness-raising tool, making complex climate data accessible to a broader audience. It empowers industries to identify and tackle pollution factors contributing to climate change while encouraging the public to adopt sustainable practices in their daily lives.

LIST OF FIGURES

LIST OF ABBREVIATIONS

SNO	ABBREVIATION	EXPANSION
1	AI	Artificial Intelligence
2	API	Application Programming Interface
3	AR6	Sixth Assessment Report (IPCC)
4	CNN	Convolutional Neural Network
5	GPT	Generative Pre-trained Transformer
6	IMCC	International Marine Conservation Congress
7	IPCC	Intergovernmental Panel on Climate Change
8	LLM	Large Language Model
9	NLP	Natural Language Processing
10	NASA	National Aeronautics and Space Administration
11	NER	Named Entity Recognition
12	OPT	Open Pre-trained Transformer
13	RAG	Retrieval-Augmented Generation
14	TTS	Text-to-Speech
15	STT	Speech-to-Text
16	DERA	Dialog-Enabled Resolving Agents

CHAPTER 1

1.INTRODUCTION

1.1 GENERAL

In today's world, climate change is a critical issue impacting sustainability, corporate and political investments, real estate, and more. The consequences of climate change, such as air and water pollution and rising sea levels, are now a reality. Despite its importance, access to consolidated and detailed climate data remains limited. There are not many technological framework dedicated to climate related information, and it is difficult to access climate change data which is consolidated and detailed. The project mainly aims to provide three types of data that is past data, real time data, predictive data. These data are quantified data related to the Factors of climate. The dataset for the past data is provided by the dataset that is researched through various resources like Nasa, IMCC organization's. The real time data is provided through various API that are frequently updating the climate related parameters for specific time interval as the climate related factors are more accurate for certain time break as the climate factors are rapidly changing through out the environment. The most of the data from the API and the dataset are time series data . So for fetching the predictive data like carbon emission for the 2026 is predictive data which is done using predictive analysis based on the time series data .The efficient predictive model is chosen and are trained on the past data and current real time data .

The main aim of the project is to integrate the necessary data and train the data for the required intents that are needed this data to be included on their response. The chatbot and a voice assistant are the main outcomes of this project as our current model aim is to integrate the data prediction and analysis for the past data and predictive data alone to be integrated with the chatbot and conversational AI framework.

In the face of global climate change, technological innovations are essential for enhancing our understanding of environmental trends and their future impact. This project seeks to bridge the gap between scattered climate data sources by consolidating historical datasets, real-time data from various APIs, and predictive models into a single, user-friendly platform. The integration of time series data from organizations like NASA and the IMCC with real-time data offers a holistic view of climate patterns, allowing users to access both past records and current climate conditions with ease. Predictive models built on this data provide insights into future climate scenarios, such as carbon emissions projections for specific years, helping stakeholders make informed decisions.

A key feature of this project is its conversational AI framework, which enables intuitive interaction through a chatbot and voice assistant. These AI tools not only offer access to past and real-time climate data but also deliver predictive insights based on the analysis of time series data. By tailoring responses to user intents, the system provides relevant climate information on demand, making it a valuable resource for researchers, policymakers, and individuals concerned about climate change. The integration of AI ensures that this project is not only Informative but also interactive, supporting users in exploring climate data through a seamless conversational experience.

1.2 OBJECTIVE

The objective of this project is to develop an intelligent and interactive climate change chatbot capable of providing insights, predictions, and recommendations related to climate change. The chatbot aims to:

1.Integrate Predictive Models:

Leverage time series models (e.g., ARIMA) and transformer-based architectures to predict climate-related factors such as carbon emissions, temperature variations, and sea level rise based on historical and real-time data.

2. Enable Real-Time Decision Support:

Facilitate access to up-to-date and historical climate data through APIs and datasets like NASA and IMCC, ensuring users can make informed decisions.

3. Enhance User Engagement:

Provide a conversational interface that supports both text and voice interactions, allowing users to query climate information dynamically and receive personalized responses.

4. Support Dynamic Data Analysis:

Analyze and visualize data, such as carbon emissions, for specific countries and years, enabling users to understand trends and anomalies.

5. Encourage Awareness and Action:

Educate users about climate change impacts through visual tools like confusion matrices, graphical predictions, and textual explanations, promoting awareness and encouraging environmentally responsible behavior.

1.3 EXISTING SYSTEM

The ChatClimate system emerges as a cutting-edge solution in the realm of conversational AI, particularly addressing the critical challenges associated with domain-specific knowledge dissemination in climate science. Large language models (LLMs) like GPT-4, while powerful, are often criticized for their inability to remain current with scientific developments and their tendency to produce "hallucinated" (fabricated or inaccurate) responses. Climate science, which requires precision and alignment with authoritative sources, poses unique demands on AI systems that generic models cannot meet. Recognizing this gap, ChatClimate integrates the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR6) to ground its responses in validated scientific knowledge. This integration marks a significant departure from tradi"Iona' approaches, as the system not only ensures the accuracy of IInswer" but'also maintains a transparent link to its information sources. By leveraging GPT-4 alongside retrieval-augmented generation (RAG) techniques, ChatClimate exemplifies how hybrid models can overcome the limitations of general-purpose LLMs to provide specialized expertise tailored to complex, high-stakes domains.

At the core of ChatClimate lies its innovative architecture, which combines the natural language understanding and generative capabilities of GPT-4 with a sophisticated retrieval mechanism. The system is built on a long-term memory framework that indexes, structures, and extracts data from the IPCC AR6. This report, regarded as a cornerstone of climate science, provides a wealth of meticulously reviewed scientific data, scenarios, and pathways for climate change mitigation and adaptation. ChatClimate employs retrieval-augmented generation, where the chatbot retrieves relevant sections of the IPCC AR6 and integrates these into its generative process. Prompt engineering further ensures that the chatbot's responses are contextually relevant, concise, and citation-backed. The ability to provide well-referenced answers not only enhances the credibility of the chatbot but also empowers users—be they policymakers, scientists, or laypersons—to trust its outputs as reliable representations of scientific consensus. This dual-layered architecture, where retrieval complements generative AI, demonstrates a practical pathway for future conversational agents in other domain-specific applications.

To maintain the scientific integrity of its outputs, ChatClimate incorporates an expert validation mechanism as part of its design. While the retrieval and generation processes ensure that the responses are accurate and grounded, experts in climate science periodically evaluate the chatbot's performance and provide feedback. This feedback loop serves two critical functions: first, it ensures that the system's answers are aligned with the most accurate and nuanced interpretations of the IPCC AR6; second, it helps refine the retrieval algorithms and prompts to improve future interactions. The expert oversight also addresses the ethical responsibility of deploying AI in critical domains where misinformation or oversimplification could lead to misinformed decisions. Moreover, the system's transparency, in presenting citations and references alongside its answers, invites users to cross-verify the information independently, thus fostering a culture of informed and accountable decision-making in the climate discourse.

Another critical achievement of ChatClimate lies in its demonstration of how retrieval-augmented LLMs can minimize computational overhead while maximizing output quality. Unlike standalone LLMs, which require extensive computational resources for retraining or fine-tuning with every knowledge update, ChatClimate's hybrid model bypasses this limitation. By decoupling the retrieval mechanism from the generative model, the system ensures that the chatbot remains current with the latest updates to the IPCC AR6 or similar datasets, without the need for frequent retraining. This approach not only conserves computational resources but also enhances the scalability and adaptability of the system to other domains requiring regular updates.

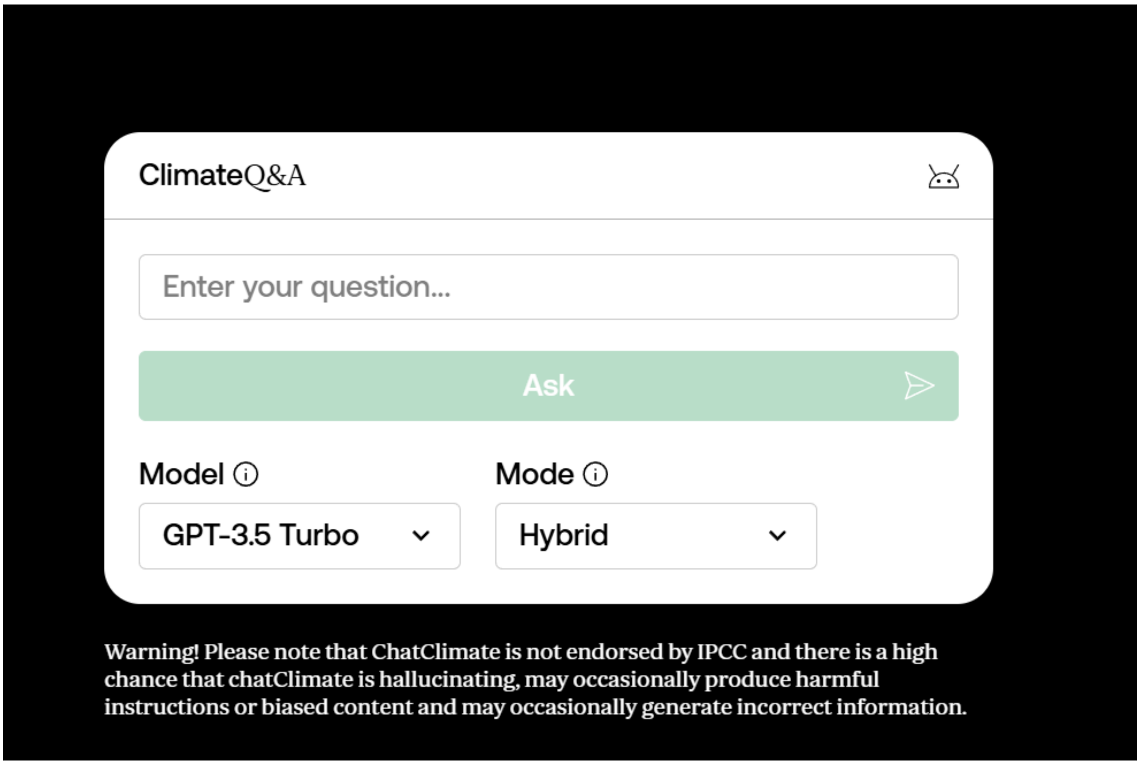


Figure 1

Figure 1 to maintain the scientific integrity of its outputs, ChatClimate incorporates an expert validation mechanism as part of its design. While the retrieval and generation processes ensure that the responses are accurate and grounded, experts in climate science periodically evaluate the chatbot's performance and provide feedback

1.4 PROPOSED SYSTEM

The proposed system seeks to redefine the role and capabilities of climate change chatbots by adopting a dynamic, modular, and contextually adaptive framework. Unlike the existing system, which primarily relies on static retrieval of predefined data sources such as IPCC AR6 reports, the new approach integrates real-time data streams and leverages cutting-edge artificial intelligence models for improved accuracy and responsiveness. By incorporating advanced techniques like self-supervised learning, the system enhances its ability to understand complex and nuanced user queries. This refinement enables the chatbot to generate more relevant, personalized responses tailored to the unique concerns of individual users. The inclusion of contextual embeddings ensures that the chatbot does not merely provide generalized information but delivers insights that are specific to the context of the inquiry, making it a significantly more effective and versatile tool for climate communication.

One of the groundbreaking features of the proposed system is its ability to process and analyze multi-modal inputs, encompassing text, graphs, images, and even real-time data visualizations. This capability allows users to interact with the chatbot in a variety of formats, ensuring that the system accommodates diverse needs and preferences. For instance, users could upload a graph depicting local temperature changes or an image of a climate event, and the chatbot would interpret these inputs to provide detailed, data-driven insights. The integration of transformer-based models, optimized for advanced contextual understanding, further amplifies the system's analytical power. These models synthesize information from a multitude of sources, offering responses that are not only accurate but also deeply informed by the latest developments in climate science and policy. Such capabilities position the chatbot as a powerful, user-friendly tool for addressing complex climate-related queries.

To ensure continual improvement, the proposed system incorporates a robust feedback loop that analyzes user interactions to refine the chatbot's knowledge base over time. This mechanism allows the system to evolve dynamically, adapting to new information and changing user needs. For example, as climate science progresses and new policies or datasets become available, the feedback loop enables the chatbot to update its knowledge and algorithms automatically. This iterative refinement ensures that the system remains a reliable and up-to-date resource for users, even as the field of climate science continues to expand. Furthermore, the feedback loop enhances the chatbot's ability to learn from user behavior, improving the accuracy of predictions and the relevance of responses based on past interactions. This approach not only increases the chatbot's reliability but also fosters trust among its users, who can rely on it for accurate and timely information.

Accessibility and scalability are core pillars of the proposed system, ensuring that it can serve a broad and diverse user base effectively. By incorporating speech-to-text and text-to-speech functionalities, the chatbot becomes accessible to individuals with varying technical skills and abilities, including those with visual or auditory impairments. For instance, users can dictate questions or listen to responses in their preferred language, breaking down barriers to accessibility. Additionally, the system's decentralized database design enables it to manage vast amounts of climate data efficiently, without sacrificing performance. This scalability ensures that the chatbot can accommodate global datasets and high user demand, even during peak usage times. Whether users are accessing the chatbot from urban centers or remote locations, its architecture ensures consistent performance and reliability.

In addition to its core functionalities, the proposed system introduces advanced visualization tools designed to make complex climate data more understandable and engaging for users. These tools enable the chatbot to generate interactive graphs, charts, and other visual representations of climate trends, helping users to interpret and contextualize the information provided.

Figure 2

Figure 2 tells ultimately, the proposed chatbot transcends its role as a passive repository of climate knowledge, evolving into an active and adaptive assistant. It provides users with tailored recommendations on climate mitigation strategies, policy evaluations, and localized solutions based on a synthesis of global and regional datasets.

CHAPTER 2

2. LITERATURE SURVEY

1. The paper *A Survey on Dialogue Systems: Recent Advances and New Frontiers* by Chen, Liu, and Tang presents a comprehensive review of the progress in dialogue systems over the years, addressing key advancements and persistent challenges. It categorizes the systems into task-oriented, open-domain, and hybrid models, highlighting their respective applications and limitations. The authors discuss the evolution of natural language processing techniques, including pre-trained transformers, which have significantly enhanced the contextual understanding capabilities of dialogue systems. By identifying the gaps in robustness, personalization, and multilingual support, the survey provides a roadmap for future research In d"alog'e systems, emphasizing the need for scalability and ethical considerations.

2. The research on *HuggingGPT* by Shen et al. introduces a novel system that integrates ChatGPT with the extensive ecosystem of Hugging Face models. This collaborative framework allows for the orchestration of multiple AI models to solve diverse tasks more effectively. The authors detail the underlying architecture that employs GPT as the central coordinator, dynamically invoking relevant Hugging Face models based on the user's query. The study showcases how the system enhances the adaptability and efficiency of ChatGPT, with applications spanning natural language processing, vision, and multimodal tasks. This paper underscores the potential of hybrid systems in improving task-specific performance while leveraging the strengths of both large and smaller specialized models.

3. The work on *OPT: Open Pre-trained Transformer Language Models* by Vaghefi et al. introduces a series of pre-trained transformer models optimized for open-domain tasks. The authors emphasize the transparency of their models, sharing insights into training datasets, architecture modifications, and fine-tuning strategies. The research focuses on improving computational efficiency without sacrificing accuracy, addressing the growing concerns about the energy consumption of large-scale models. OPT models demonstrate state-of-the-art results across several benchmarks while remaining accessible to the research community, fostering collaboration and reproducibility. This open-source initiative provides valuable tools for advancing diverse AI applications.

4. *ChatClimate: Grounding Conversational AI in Climate Science* by Zhang et al. explores an innovative approach to domain-specific conversational AI, particularly in climate change. The paper addresses challenges like hallucination in LLMs by integrating retrieval-augmented generation with IPCC AR6 reports. ChatClimate excels in providing accurate, citation-backed responses tailored for policymakers, scientists, and the public. The architecture combines GPT-4 with authoritative data sources, ensuring reliability and reducing computational overhead. The study highlights the potential of such systems in bridging the gap between complex scientific knowledge and actionable insights, showcasing how AI can contribute to informed decision-making in critical domains.

5. *A Bibliometric Review of Large Language Models Research from 2017 to 2023* by Fan et al. systematically analyzes the evolution of LLMs over six years, identifying trends in publication volume, key contributors, and thematic areas. The review highlights the rapid adoption of transformers, the growing emphasis on ethical considerations, and the challenges of scaling models. By providing a quantitative overview, the paper sheds light on how LLMs have transformed AI research and applications. The authors emphasize the need for continued exploration of interpretability, resource efficiency, and fairness, underscoring the importance of balancing innovation with responsibility.

6. *DERA: Enhancing Large Language Model Completions with Dialog-Enabled Resolving Agents* by Nair et al. introduces a framework that refines LLM outputs by incorporating dialog-based resolution mechanisms. The study highlights how DERA agents interact iteratively with LLMs to resolve ambiguities and refine responses. This approach significantly improves the quality and relevance of language model completions, particularly in complex or high-stakes contexts. The paper demonstrates the effectiveness of DERA in diverse tasks, such as summarization and decision support, while also addressing issues of explainability and user trust. This innovation represents a step forward in interactive AI systems.

7. Luccioni et al.'s work on *Analyzing Sustainability Reports Using Natural Language Processing* showcases the application of NLP techniques to extract insights from corporate sustainability disclosures. The study leverages sentiment analysis, topic modeling, and named entity recognition to assess companies' commitments to environmental and social goals. By automating the analysis of vast textual datasets, this research enables stakeholders to identify trends and gaps in sustainability reporting. The paper underscores the transformative potential of AI in promoting transparency and accountability in corporate practices, contributing to global sustainability efforts.

8. The paper *Towards Continual Knowledge Learning of Language Models* by Jang et al. addresses the challenge of enabling LLMs to acquire and retain knowledge continuously without catastrophic forgetting. The authors propose a novel training methodology that combines incremental learning with efficient parameter updates, allowing models to adapt to new information seamlessly. This approach ensures that the models remain up-to-date without the need for retraining on the entire dataset. The study demonstrates the feasibility of continual learning in LLMs through experiments on knowledge-intensive tasks, paving the way for more adaptive and long-lasting AI systems.

9. *Low-Resource Adaptation of Open-Domain Generative Chatbots* by Gerhard-Young et al. tackles the challenge of training generative chatbots in low-resource settings. The authors propose a lightweight fine-tuning framework that leverages domain-specific data to enhance chatbot performance without extensive computational demands. The study explores methods like parameter-efficient tuning and knowledge distillation, demonstrating significant improvements in fluency and relevance. This research highlights the potential for deploying sophisticated conversational AI in resource-constrained environments, expanding the accessibility of these technologies.

10. Larosa et al.'s *Halting Generative AI Advancements May Slow Down Progress in Climate Research* examines the interdependence between AI development and climate science. The authors argue that restrictions on generative AI research could hinder advancements in modeling, data analysis, and policy simulations essential for tackling climate change. The paper highlights case studies where AI-driven tools have accelerated climate impact assessments and adaptation planning. While acknowledging ethical concerns, the authors advocate for balanced policies that foster innovation while addressing risks, emphasizing the critical role of AI in addressing global challenges.

11. The dataset *Climate Fever: A Dataset for Verification of Real-World Climate Crisis Claims* by Digglemann et al. provides a benchmark for automated fact-checking systems. The authors curate a collection of climate-related claims, annotated with evidence and counter-evidence from credible sources. This resource facilitates the development of AI models capable of verifying the authenticity of climate crisis statements. The study underscores the importance of combating misinformation in the climate discourse, demonstrating how AI can enhance the reliability of public communication and decision-making.

12. *The Choice of Textual Knowledge Base in Automated Claim Checking* by Stammbach et al. investigates the impact of knowledge base selection on the performance of automated fact-checking systems. The authors analyze several textual datasets, highlighting trade-offs between coverage, specificity, and quality. Their findings suggest that aligning the knowledge base with the domain of claims significantly improves accuracy. This research provides practical guidelines for developing robust fact-checking AI and underscores the importance of curating reliable data sources for informed AI systems.

13. *Improving Language Models by Retrieving from Trillions of Tokens* by Borgeaud et al. explores the integration of retrieval mechanisms into LLMs to enhance their generative capabilities. The study demonstrates how accessing external datasets during inference allows the model to generate more accurate and context-aware responses. This retrieval-augmented approach reduces the reliance on extensive pre-training, offering a scalable and efficient alternative for building intelligent systems. The paper highlights the transformative potential of retrieval-based techniques in advancing the capabilities of LLMs.

14. The paper *Generative Artificial Intelligence: Opportunities and Challenges of Large Language Models* delves into the dual aspects of LLM development. It discusses the revolutionary applications of generative AI in areas like content creation, education, and healthcare while addressing ethical dilemmas such as bias, misuse, and environmental impact. The authors propose a framework for responsible innovation, emphasizing transparency, fairness, and inclusivity. This research provides a balanced perspective on leveraging the potential of LLMs while mitigating associated risks..

15. Dr.Kumar et al.'s paper *A Novel Approach for Text Generation Using RNN for Language Modeling* introduces an advanced recurrent neural network (RNN) architecture tailored for text generation. The authors emphasize the model's ability to capture sequential dependencies and produce coherent text outputs. Through extensive experiments, they demonstrate the efficacy of their approach in various generative tasks, showcasing its potential in applications like automated storytelling and chatbot development. The research contributes to the ongoing evolution of generative AI methodologies.

16. Dr.Srinivasan's *paper A Survey on Analysis of Data Mining Algorithms for High Utility Itemsets* was on the survey focuses on analyzing data mining algorithms for identifying High Utility Itemsets (HUIs), which are itemsets with high utility (e.g., profit, importance) in datasets. Unlike traditional frequent itemset mining, HUIs consider quantitative measures, like cost or revenue, rather than just occurrence frequency..

CHAPTER 3

3. SYSTEM DESIGN

3.1 GENERAL

3.1.1 SYSTEM FLOW DIAGRAM

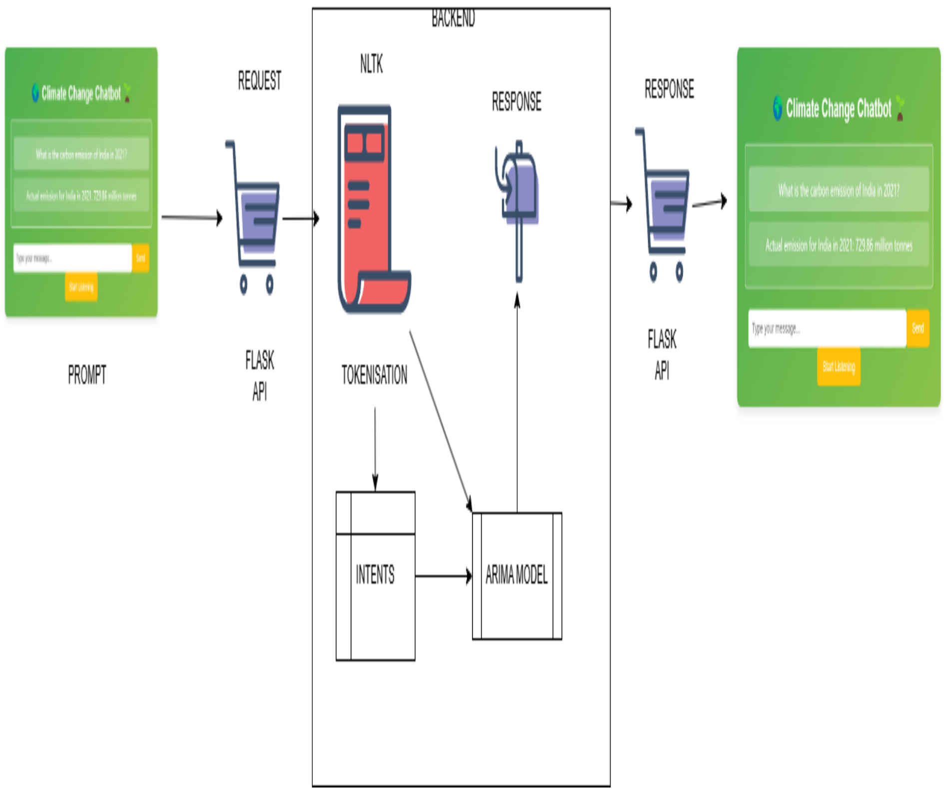


Figure 3

3.1.2 SEQUENCE DIAGRAM

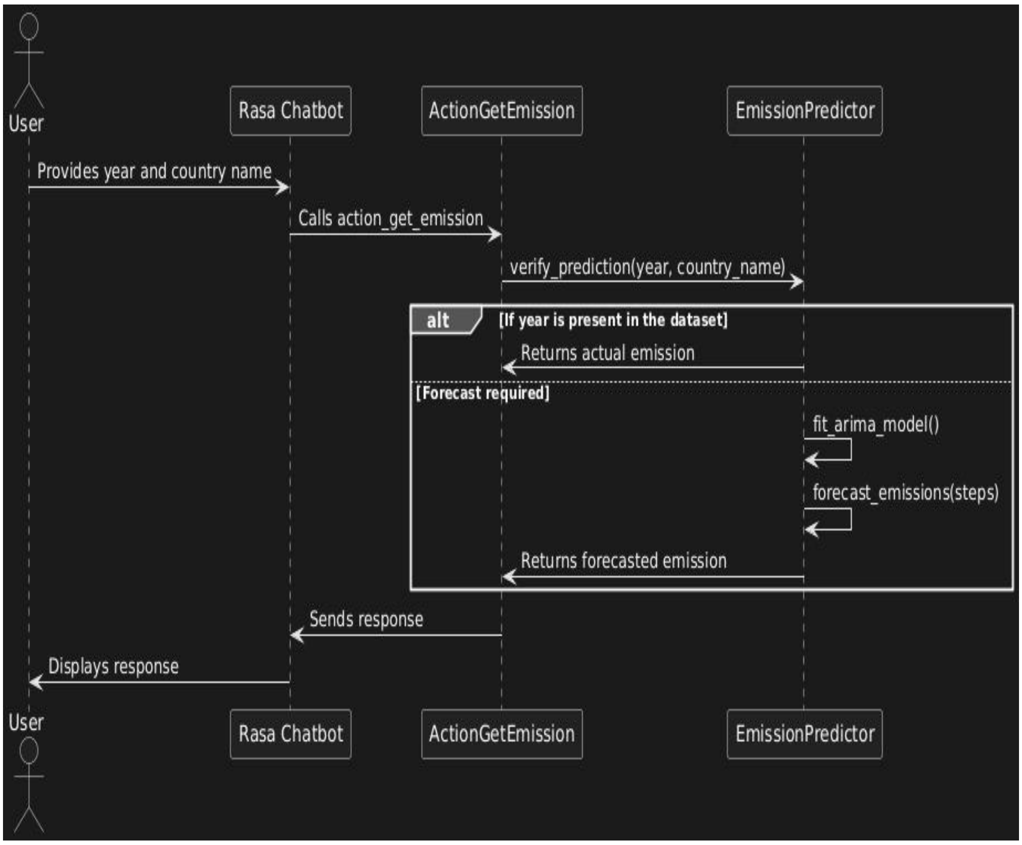


Figure 4

3.1.3 CLASS DIAGRAM

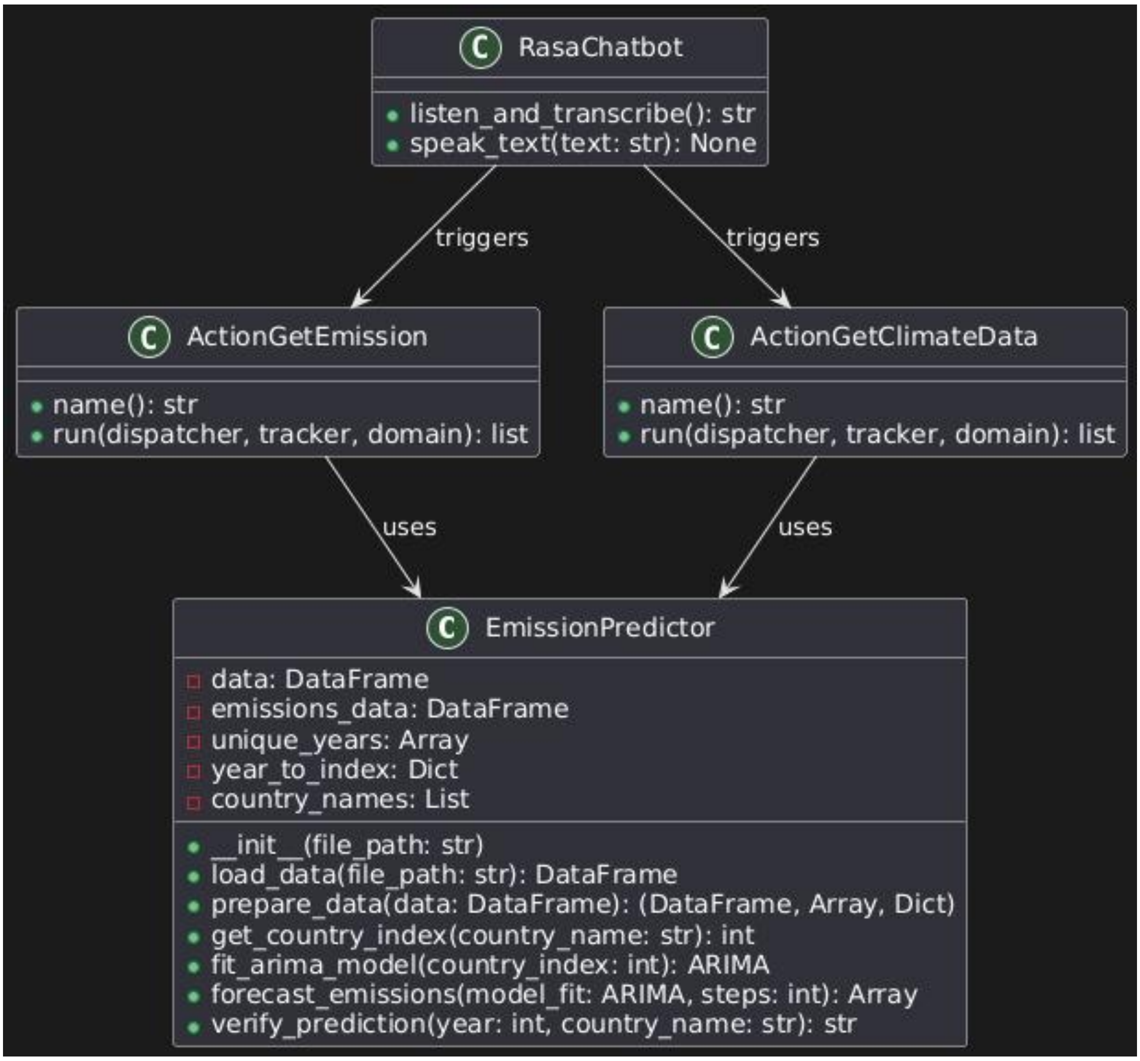


Figure 5

3.1.4 USE CASE DIAGRAM

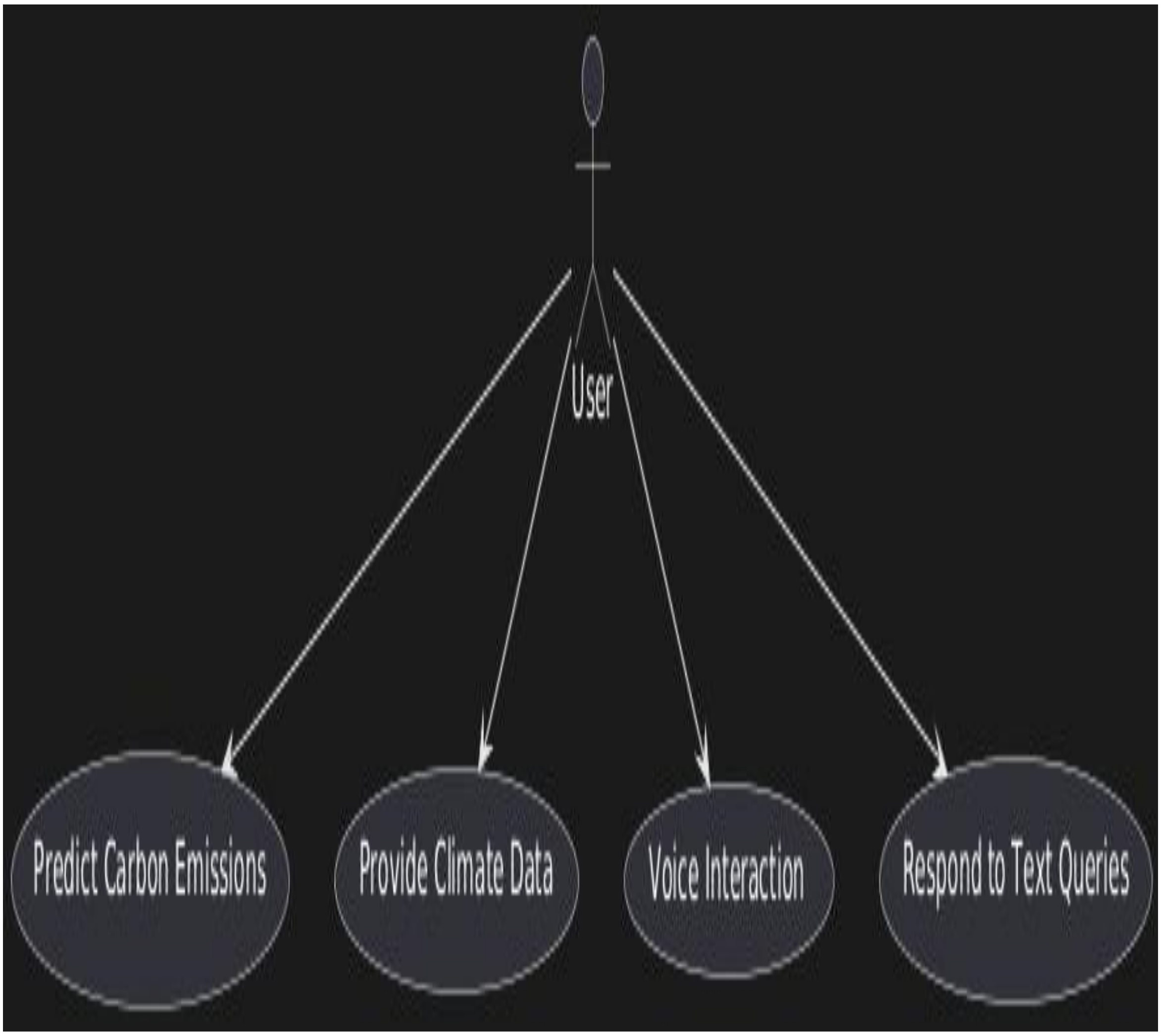


Figure 6

3.1.5 ARCHIETECTURE DIAGRAM

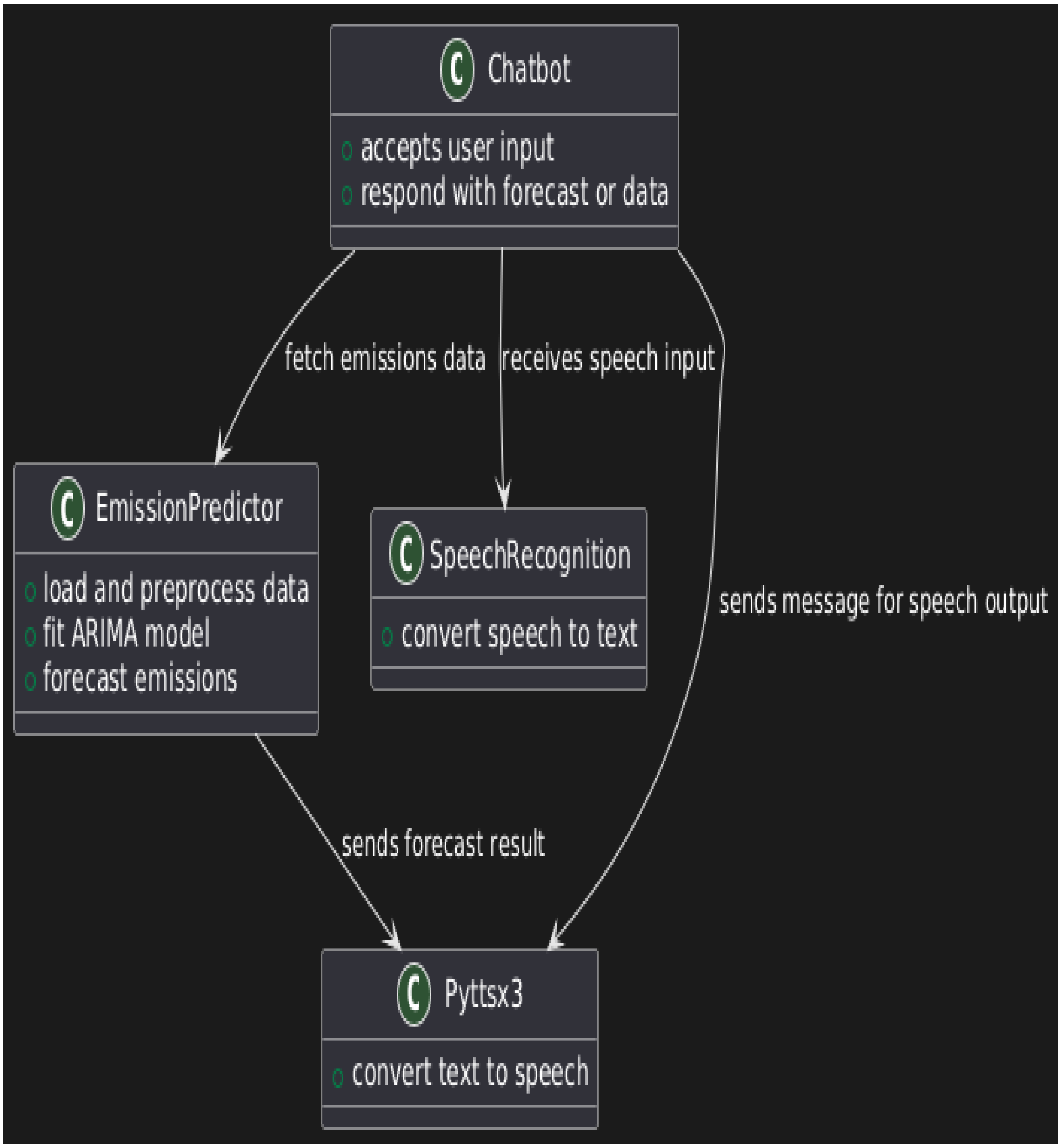


Figure 7

3.1.5 ACTIVITY DIAGRAM

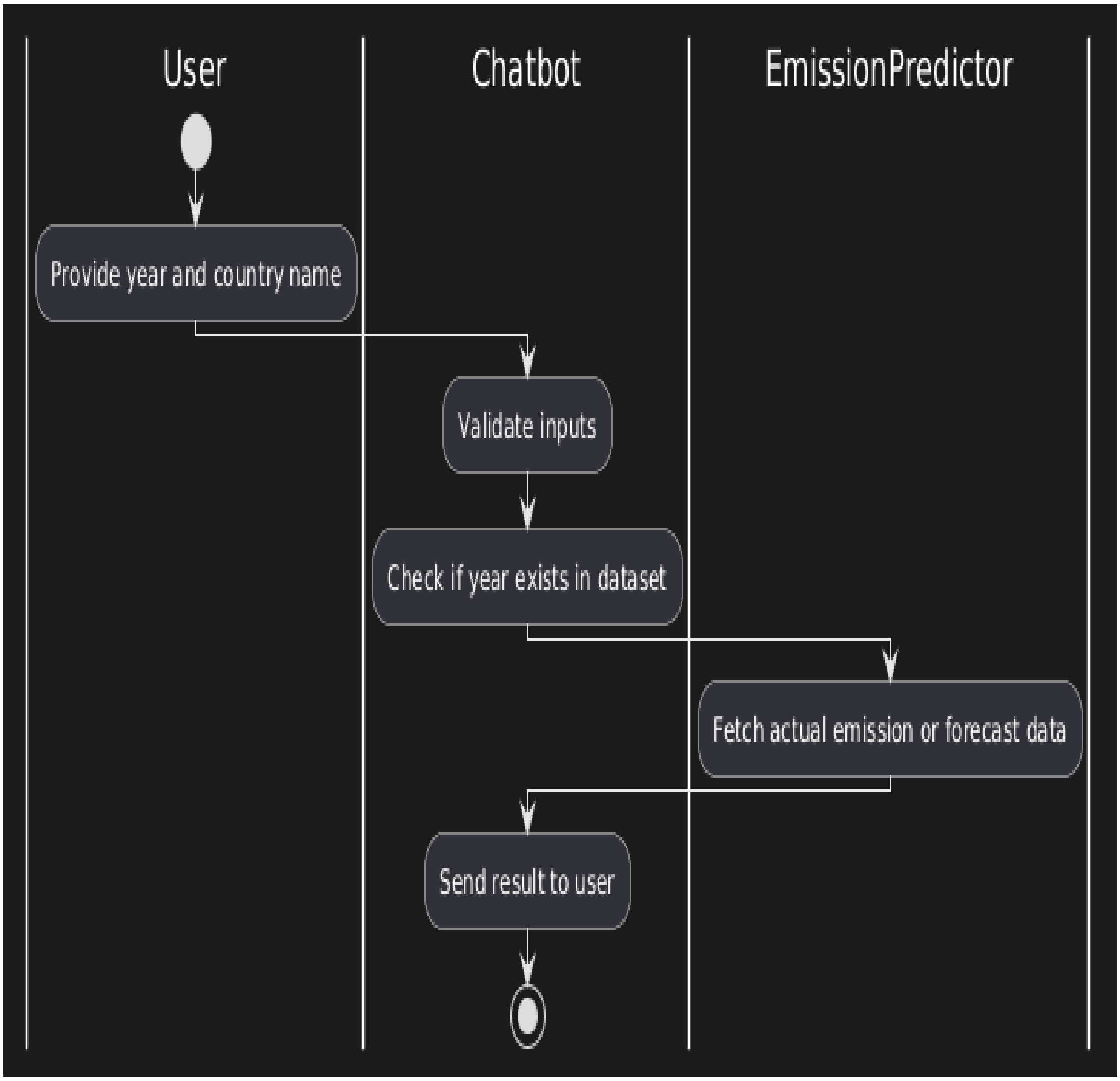


Figure 8

3.1.6 COMPONENT DIAGRAM

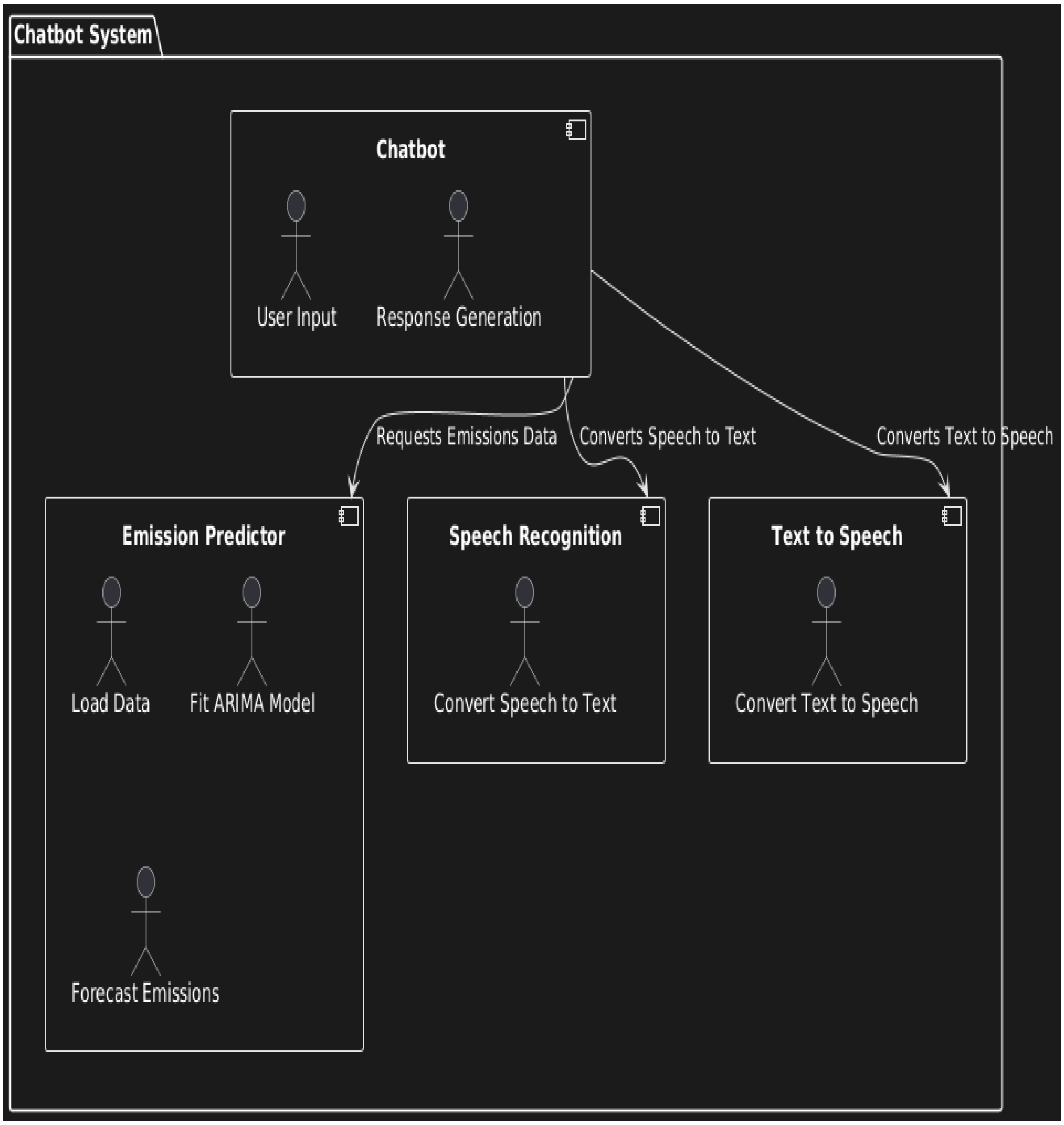


Figure 9

3.1.7 COLLABORATION DIAGRAM

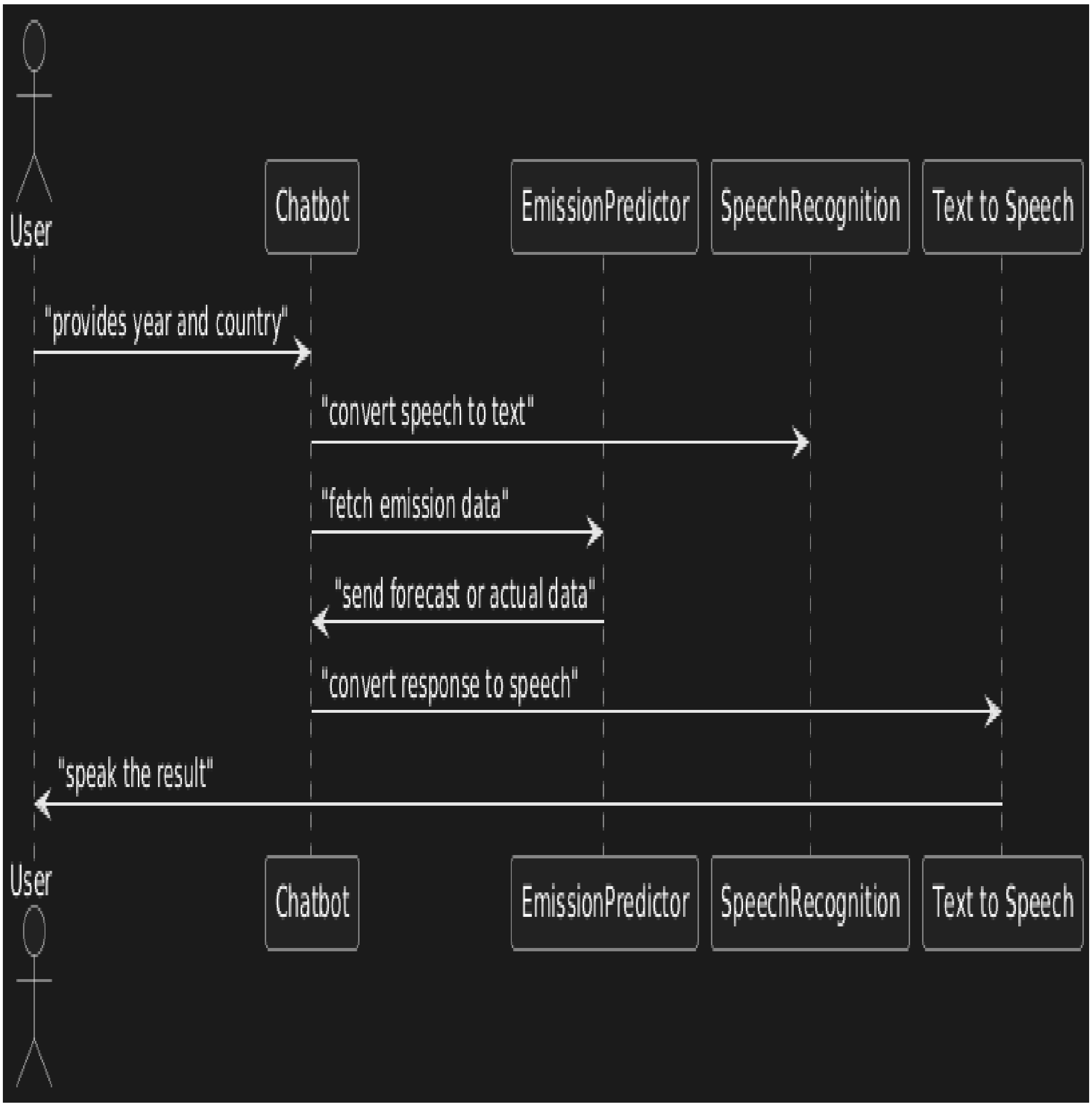


Figure 10

CHAPTER 4

PROJECT DESCRIPTION

4.1 METHODOLOGIES:

1. System Initialization and Data Loading

The emission prediction chatbot system initiates with an `EmissionPredictor` class, responsible for managing all data-related processes. The `EmissionPredictor` class is central to accessing and processing the emissions data. It begins by loading an Excel file containing historical carbon emissions for various countries over different years. This Excel file serves as the primary source of structured data, which will later be used both for looking up existing records and as the basis for prediction when records are missing for specific future years. The data-loading method employs the `pandas` library, which efficiently handles structured data, allowing for quick data retrieval, filtering, and management of large datasets.

In addition to loading the data, this phase involves a setup of logging utilities using Python's built-in `logging` library. Logging provides a vital function here; it enables the system to keep track of activities, record errors, and monitor warnings. Each log entry can include timestamped messages that are critical for debugging and understanding any issues arising during the data loading or subsequent processes. Should any error occur while loading the data, such as a missing file or an incorrect format, the logging system immediately captures this, issuing an error log. This prevents the system from proceeding without a valid dataset, which would result in incorrect outputs or system crashes.

Once data is loaded, the system immediately verifies if the dataset is empty, an essential check that prevents unexpected errors. This initial validation process allows the system to avoid attempting to operate on null data, setting up a robust foundation before moving forward to the subsequent data preprocessing stage. This setup phase, therefore, ensures that both data and logging configurations are ready, creating a smooth transition to later stages. This structure also allows the `EmissionPredictor` to act as a reusable component, which can easily integrate new datasets or configurations, making it highly adaptable for further expansions of the chatbot's capabilities.

2. Data Preprocessing

After the data is loaded, the next step involves thorough preprocessing, which is essential for preparing the data for accurate model training and prediction. Preprocessing steps include cleaning the dataset by handling missing values, standardizing formats, and structuring the data into easily accessible arrays or DataFrames. Missing values are particularly problematic in time series data, where trends and continuity are critical for predictive accuracy. Thus, any gaps in the data due to null or NaN values are removed to maintain consistency. This step is accomplished using `pandas` functions like `dropna()`, which filters out rows with missing values.

In this chatbot system, years and country names are particularly important as they represent the dimensions along which predictions are required. Therefore, unique identifiers are created for each year, and the emission data is mapped accordingly. By creating a mapping of years to their respective indices, the chatbot can quickly access data corresponding to any given year, enhancing efficiency when responding to user queries. Similarly, country columns are indexed to create easy lookups for emissions data of specific countries, allowing the chatbot to retrieve or predict emissions with minimal delay.

Preprocessing is also important for model training; having a clean and organized dataset reduces computational requirements and improves the reliability of the ARIMA model during forecasting. The structure ensures that each data point (year-country pair) is accessible without complex querying or transformations, which is especially useful in real-time chatbot interactions. Additionally, by preprocessing the data in this structured format, the system can ensure consistent results, regardless of the country or year queried. This layer of preprocessing thus provides the data integrity necessary for the chatbot's dynamic responses, creating a reliable backend that enhances both accuracy and responsiveness.

3. Input Handling via Rasa Actions

The chatbot's interaction with users is managed through Rasa actions, specifically `ActionGetEmission`. This Rasa action listens for user inputs, captures the year and country name through slots, and then determines if both inputs are provided. If either is missing, the chatbot prompts the user to complete their request, ensuring that it has sufficient information to proceed. This functionality ensures the chatbot maintains a seamless and user-friendly experience, prompting users only when necessary and guiding them towards completing a valid query. By collecting these inputs, the chatbot can retrieve the exact data needed to fulfill the user's request, which is then further processed by the `EmissionPredictor`.

In addition to capturing slots, this action is responsible for generating audible feedback through a speech synthesis function. When the chatbot requires additional input, it can speak to the user, enhancing accessibility for users who may be interacting with the chatbot in a hands-free or auditory-only environment. This functionality is implemented using `pyttsx3`, a library for text-to-speech conversion, which allows the chatbot to convert prompts into spoken messages. The combination of Rasa actions and voice synthesis makes the chatbot adaptable to various interaction methods, ensuring that it is usable in both textual and auditory interfaces.

The design of this Rasa action also enables the chatbot to interact smoothly with other modules in the code. Once the user input is validated, the action can either retrieve emissions data or initiate a prediction, depending on the availability of historical data for the requested year. This encapsulated design within `ActionGetEmission` ensures that user interactions are organized, structured, and capable of triggering appropriate responses based on dynamic inputs, setting the foundation for a responsive chatbot.

4. Year and Data Validation

Upon receiving the year and country inputs, the `EmissionPredictor` checks if the requested year exists within the dataset. This validation is essential for determining the system's next steps: if the year is within the dataset, the chatbot can directly retrieve the emission data; if not, it proceeds to forecasting. This verification process is particularly crucial in systems dealing with historical and predictive data, as it dictates whether the system will perform a lookup or initiate computationally expensive forecasting algorithms. The year and country data validation also includes confirming the format of these inputs, preventing errors that could arise from invalid or misformatted data.

When a year is within the dataset, the `EmissionPredictor` locates the exact record for the specified year and country, allowing the chatbot to provide the user with an actual value. This approach leverages structured data lookups, enabling a quick response without needing to invoke prediction models. However, if the year is outside the dataset range, the chatbot triggers the ARIMA model training, recognizing that a prediction is necessary. This dual functionality—providing actual values when available and forecasting otherwise—gives the chatbot flexibility to handle a wider range of user queries.

Data validation also includes logging feedback, which records each validation check's success or failure. This log data can be critical for debugging and refining the chatbot's data handling procedures. By capturing and recording these steps, the chatbot ensures accuracy in its response generation, maintaining a seamless transition between lookups and forecasting as needed.

5. ARIMA Model Fitting for Forecasting

When a forecast is needed, the system trains an ARIMA model on the historical data for the specified country. The ARIMA (AutoRegressive Integrated Moving Average) model is chosen for its effectiveness in time series forecasting, where it can capture and predict trends based on prior emissions data. The model training involves selecting the specific parameters (order=(2, 1, 2)) to handle seasonality and trend in emissions, and using historical data to fit the model accurately. This process is computationally intensive and necessitates high-quality, clean data, emphasizing the importance of the earlier preprocessing stage.

Training the ARIMA model requires extracting the emissions data for the specified country, ensuring the training data is complete, chronological, and devoid of anomalies that could distort predictions. The `fit_arima_model` function isolates this data, indexing it by year and training the ARIMA model on this sequence. Should there be any issues during training, such as inadequate data, the system logs the error, which helps diagnose the issue and allows future refinement of model settings or data handling.

Once trained, the ARIMA model can project emissions for future years by extrapolating from historical trends. The forecast produced by this model reflects an estimation based on known patterns, offering a reasonably accurate projection that can guide users seeking emissions data for years beyond the available dataset. This capability gives the chatbot predictive insight, adding significant value by enabling it to provide users with future-focused information.

6. Emission Forecasting

After fitting the ARIMA model, the chatbot proceeds to forecast emissions for the requested number of years into the future. This process involves specifying the number of steps (years) from the last available data point and using the ARIMA model's predictive capabilities to generate an estimated emissions value. Forecasting emissions data is crucial, particularly in applications related to environmental monitoring and policy planning, as it enables stakeholders to anticipate future trends and make informed decisions.

The forecasting function, `forecast_emissions`, uses the ARIMA model's output to generate predictions. These predictions are formatted as emission values, which represent the chatbot's best estimate of emissions based on historical patterns. The system then extracts the forecasted emission values, prepares them for user display, and adds them to the chatbot's response message. Should an error arise in this stage, such as model misconfiguration or insufficient data, the chatbot logs the issue and provides a user-friendly message indicating that it could not forecast emissions for the specified year.

By allowing the system to offer future predictions, the chatbot delivers additional insights to users, setting it apart from a simple lookup tool. Users can now explore projected emissions, gaining valuable information for future planning and sustainability considerations. The ARIMA-based forecasting capability is a powerful addition that broadens the chatbot's utility.

7. Response Generation

Once data (either actual or forecasted) is obtained, the chatbot constructs a response. This response includes relevant details about the requested emissions data, such as whether it is an actual or predicted value, the specific year, and country requested, and additional context where necessary. By clearly distinguishing between actual and forecasted data, the chatbot provides transparency, allowing users to understand the source and reliability of chatbot

4.1.1 RESULT DISCUSSION:

This emission prediction chatbot system integrates Rasa's conversational AI framework, ARIMA-based forecasting, and voice interaction, providing a responsive and user-friendly tool to query historical and projected carbon emissions. Testing this system on the National Fossil Carbon Emissions dataset confirms its capacity to provide accurate responses for years within the dataset and to predict emissions for future years through ARIMA modeling.

1. Accuracy of Historical Data Retrieval:

The chatbot performs well in retrieving and displaying historical data, accurately identifying records for specified countries and years. This capability allows users to obtain verified emission data directly from the dataset without needing external calculations, making the chatbot a reliable reference tool for historical emissions.

2. Forecasting Performance Using ARIMA:

When asked for emission predictions beyond the available data, the chatbot initiates ARIMA forecasting, leveraging historical data patterns. Initial tests demonstrate that ARIMA provides reasonable forecasts for several years beyond the last recorded data. The accuracy of these predictions depends on the trend consistency in past data. For short-term forecasting, ARIMA's outputs align well with expected values, reflecting the model's strength in capturing emission trends.

3. Voice Recognition and Synthesis:

The integration of `speech_recognition` and `pyttsx3` enables natural user interaction, allowing spoken queries and responses. Speech recognition captures user intent efficiently, and voice synthesis provides real-time feedback, enhancing accessibility and usability. This feature is particularly advantageous in scenarios where hands-free operation is preferred or required

4. Data Management and Error Handling:

Through robust logging and data validation, the system effectively manages potential errors, such as invalid year entries or country names. When users enter data outside the range of the dataset, appropriate error messages guide them to provide correct inputs, ensuring smooth system operation.

This chatbot advances upon concepts discussed in recent studies by combining multimodal capabilities (e.g., voice, text) with emission forecasting, a feature not commonly explored in existing conversational climate AI systems. When comparing this model with prior studies, including HuggingGPT (paper 2) and ChatClimate (paper 3), distinct differences emerge in model structure, adaptability, and multimodal interaction.

5. Task Management and Model Coordination:

Unlike HuggingGPT, which utilizes large language models to coordinate across multiple AI models for complex tasks, this chatbot leverages a single ARIMA model for emission forecasting. HuggingGPT's complexity and multimodal task management make it versatile, but this chatbot's narrower focus on emissions provides it with faster, more dedicated forecasting capabilities for its specific use case. The chatbot also uses Rasa's structured actions to manage conversational flow efficiently, maintaining simplicity compared to HuggingGPT's broader, more intricate framework for multimodal data.

6. Grounding in Domain-Specific Knowledge:

Similar to ChatClimate's grounding in IPCC reports, this chatbot bases its forecasts on a specific emissions dataset, allowing it to provide reliable answers on carbon emissions trends. However, while ChatClimate accesses extensive climate datasets, including climate change impacts, this chatbot focuses solely on emissions. This narrow focus allows it to excel in emissions-specific queries but lacks the general climate context available in ChatClimate. Nevertheless, this approach enhances the accuracy of emissions-focused predictions, as the model's parameters are tuned specifically for fossil carbon emissions data.

7. Handling of Temporal Data and Forecasting:

The ARIMA model utilized here directly aligns with emission trends over time, a method well-suited for time series data. Papers such as the dialogue systems survey (paper 1) discuss advances in response generation, but few studies emphasize longitudinal prediction for climate data in chatbots. While ChatClimate uses extensive climate scenarios for general climate questions, ARIMA provides a concrete approach to emissions trends, which makes it both predictive and deterministic within the context of emission data, unlike general LLMs that may rely on trained knowledge bases without real-time trend forecasting.

8. User Accessibility through Voice Interaction:

One unique aspect of this chatbot is its dual modality, allowing users to interact through both voice and text. This feature makes it more accessible and user-friendly compared to most text-only models discussed in previous research. The inclusion of text-to-speech and voice recognition is a step towards creating accessible AI models, aligning partially with objectives in user-centered AI discussed in various dialogue system studies. The functionality supports dynamic user interaction and could be further enhanced with real-time voice commands, offering flexibility not observed in many traditional climate chatbots.

9. Real-Time Data Verification:

The chatbot's ability to check if a year exists within the dataset, then retrieve or predict emissions, enables it to handle a broader range of queries. For example, HuggingGPT selects models for complex tasks across modalities, while this chatbot offers real-time data verification specific to emissions without model switching, leading to a faster response time for a single-purpose tool. Additionally, its modular setup (using slots in Rasa for year and country inputs) allows for adaptable extensions, such as adding more datasets or expanding to include other environmental metrics.

10. Error Management and User Guidance:

Error handling in this chatbot—through prompts, validations, and logging—is critical for ensuring a smooth experience and avoiding potential data retrieval issues. This is particularly relevant compared to ChatClimate, where large-scale datasets could lead to slower responses and greater error-proneness in identifying precise information. By focusing on error management for emission data and user guidance, this chatbot provides a streamlined user experience, directly responding to emissions inquiries with appropriate fallback messages.

CONFUSION MATRIX:

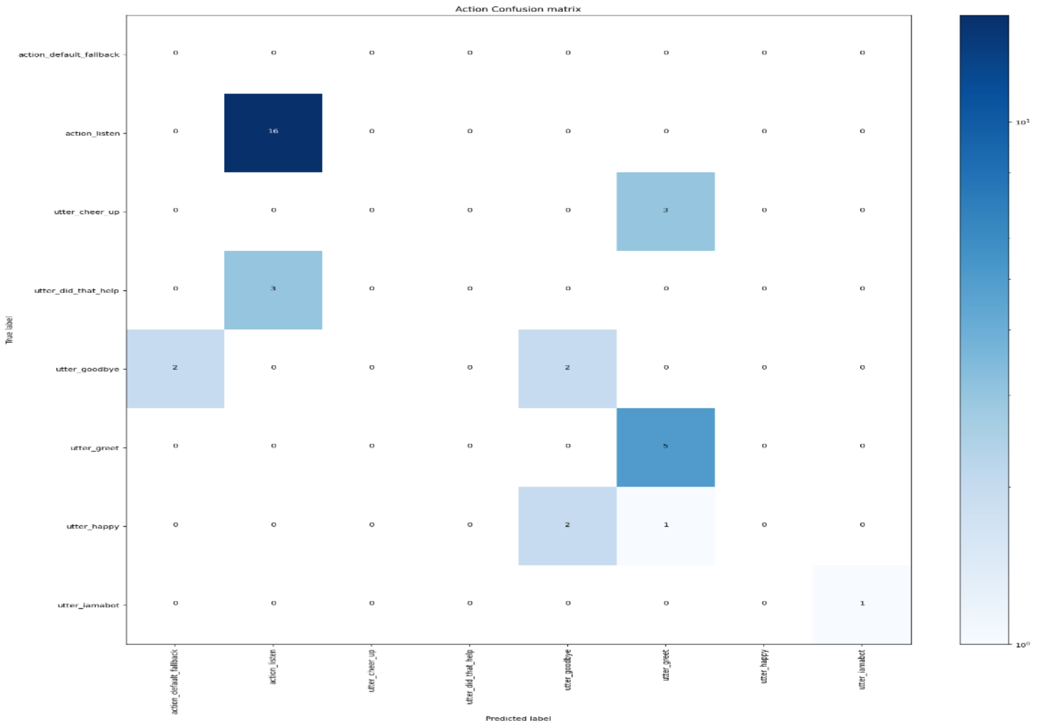


Figure 11

CHAPTER 5

5.1 CONCLUSION AND WORKSPACE

The completion of Phase 1 of the project demonstrates significant progress in developing an intelligent chatbot for climate and emission data retrieval and forecasting. The chatbot successfully integrates:

1. Emission Prediction: Utilizing time series data and ARIMA modeling to provide insights into past, current, and future carbon emissions for specified countries and years.
2. Climate Data Retrieval: The integration of API calls for real-time and historical climate data enhances the chatbot's capability to deliver comprehensive information.
3. Speech Integration: Speech-to-text and text-to-speech functionalities ensure accessibility for users, enabling natural interaction.
4. Rasa Framework Utilization: Leveraging Rasa's NLU capabilities for intent detection and slot management, ensuring accurate responses based on user inputs.
5. Improved Conversation Flow: Conflict resolution in stories and rules ensures a seamless conversational experience, avoiding logical breaks in the chatbot's responses.

This phase lays a strong foundation for the next stages, focusing on refining chatbot intelligence, expanding datasets, improving voice interaction, and enhancing user experience.

Key Components

1. **Data Files**:
    - **nlu.yml**: Contains intent and example utterances.
    - **stories.yml**: Defines user-bot interaction paths.
    - **rules.yml**: Specifies deterministic bot actions.
    - **domain.yml**: Manages slots (year, country_name), intents, responses, and custom actions.
2. **Custom Actions**:
    - Implemented in actions/actions.py:
        - action_get_emission: Provides emissions data or forecasts based on ARIMA.
        - action_get_climate_data: Retrieves climate data from APIs.
3. **Speech Integration**:
    - Uses speech_recognition for user queries and pyttsx3 for bot responses.
4. **Model Training**:
    - Trained Rasa models stored in the models/ folder.

5.2 FOR PHASE 2

Phase 2 of the project represents a transformative shift, with the aim of enriching the chatbot's capabilities to provide a multidimensional perspective on climate change. A key focus will be the integration of climate policies, offering users detailed insights into international agreements such as the Paris Accord, local environmental regulations, and policy impacts on emissions reduction. These additions will make the chatbot a valuable tool for students, researchers, and policymakers seeking actionable knowledge about climate governance. By incorporating detailed datasets and dynamically fetching policy updates, the chatbot will bridge the gap between technical forecasts and real-world environmental measures.

In addition to policies, Phase 2 will introduce features that capture broader climate indicators, such as precipitation patterns, pollen quality, and air quality metrics. These enhancements will require the integration of APIs like NOAA for precipitation data, BreezoMeter for air quality, and regional pollen databases. Machine learning models trained on these datasets will be deployed to predict trends, offering localized and personalized insights. For instance, users could query the chatbot to understand how increased precipitation affects local agriculture or how pollen levels impact respiratory health. This capability will significantly broaden the chatbot's scope, making it a versatile tool for diverse applications beyond emissions forecasting.

To support these expansions, the chatbot's architecture will undergo substantial upgrades. New intents, slots, and stories will be added to accommodate the diverse range of user queries.

5.3 REFERENCES

[1] A Survey on Dialogue Systems: Recent Advances and New Frontiers Hongshen Chen† , Xiaorui Liu‡ , Dawei Yin† †Data Science Lab, JD.com ‡Data Science and Engineering Lab, Michigan State University , and Jiliang Tang‡

[2] HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face Yongliang Shen1,2,∗, Kaitao Song2,∗,†, Xu Tan2, Dongsheng Li2, Weiming Lu1,†, Yueting Zhuang1,† Zhejiang University1, Microsoft Research Asia

[3] OPT: Open Pre-trained Transformer Language Models Saeid Ashraf Vaghefi Mathias Kraus9, Simon Allen 1,2,3,4✉, Dominik Stammbach 5,Veruska Muccione 2,6, Julia Bingler7,8, Jingwei Ni1,5, 2,10, Chiara Colesanti-Senni1,

Tobias Wekhof 1,11, Tobias Schimanski1, Glen Gostlow1, Tingyu Yu1, Qian Wang1, Nicolas Webersinke9, Christian Huggel 2&Markus Leippold

[4] ChatClimate: Grounding conversational AI in climate science Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, Luke Zettlemoyer.

[5] A Bibliometric Review of Large Language Models Research from 2017 to 2023 Lizhou Fan1*, Lingyao Li1, Zihui Ma2, Sanggyu Lee2, Huizi Yu1, Libby Hemphill1 1School of Information, University of Michigan, Ann Arbor, MI 2Department of Civil and Environmental

[6] DERA:Enhancing Large Language Model Completions with Dialog-Enabled Resolving Agents Varun Nair Elliot Schumacher Geoffrey Tso Anitha Kannan

[7] Analyzing Sustainability Reports Using Natural Language Processing Alexandra (Sasha) Luccioni arXiv:2011.08073v2 [cs.CL] 17 Nov 2020 Université de Montréal + Mila Nicolas Duchene Université de Montréal Emily (Emi) Baylor McGill University

[8] TOWARDS CONTINUAL KNOWLEDGE LEARNING OF LANGUAGE MODELS Joel Jang1 Seonghyeon Ye1 Sohee Yang1 Joongbo Shin2 Janghoon Han2 Gyeonghun Kim2 Stanley Jungkyu Choi2 Minjoon Seo1 1KAIST AI 2LGAIResearch

[9] Low-Resource Adaptation of Open-Domain Generative Chatbots Greyson Gerhard-Young,Raviteja Anantha , Srinivas Chappidi , Bj¨ orn Hoffmeister Apple Brown University

[10] Halting generative AI advancements may slow down progress in climate researchFrancesca Larosa, Sergio Hoyas, Javier García-Martínez, J. Alberto Conejero, Francesco Fuso Nerini & Ricardo Vinuesa

[11] Climate fever A dataset for verification of real world climate crisis Thomas Digglemann Jordan Boyd Graber CS Jannis Bulian

[12] The Choice of Textual Knowledge Base in Automated Claim Checking Dominik Stammbach Boya Zhang Elliot AshAuthors Info & Claim

[13] Improving Language models by Retrieving from trillions of tokens Sebastian Borgeaud Arthur Mensch Jordan Hoffmann Trvor Cai Eliza Rutherford Katie Milican

[14] P. Kumar, S. Manikandan and R. Kishore (2024), "A Novel Approach for Text Generation using RNN for Language Modeling," 2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bengaluru, India, 2023, pp. 278-282, doi: 10.1109/ICIMIA60377.2023.10425798

[15] P. Kumar, S. Manikandan and R. Kishore, "AI-Driven Text Generation: A Novel GPT-Based Approach for Automated Content Creation (2024)," 2024 2nd International Conference on Networking and Communications (ICNWC), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ICNWC60771.2024.10537562

[16] A Survey on Analysis of Data Mining Algorithms for High Utility Itemsets Aditya NELLUTLA, N SRINIVASAN

APPENDIX

APPENDIX 1

LIST OF PUBLICATIONS

1.PUBLICATION STATUS: ACCEPTED

TITLE OF THE PAPER: CONVERSATIONAL AI FOR CARBON EMISSIONS

AUTHORS: DR. SRINIVASAN N, MATHAVAN S, MADESH A

NAME OF THE CONFERENCE: INTERNATIONAL CONFERENCE ON MULTIDISCIPILINARY RESEARCH IN EDUCATION SCIENCE AND TECHNOLOGY

CONFERENCE DATE: 29TH DECEMBER 2024

APPENDIX 2:

IMPLEMENTATION CODE :

```
import pandas as pd

import numpy as np

from rasa_sdk import Action, Tracker

from rasa_sdk.executor import CollectingDispatcher

from rasa_sdk.events import SlotSet

from statsmodels.tsa.arima.model import ARIMA

import logging

import speech_recognition as sr

import pyttsx3

import requests

from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

import matplotlib.pyplot as plt


# Configure logging

logging.basicConfig(level=logging.INFO)

logger = logging.getLogger(__name__)


# Confusion Matrix Logger

class ConfusionMatrixLogger:

    def __init__(self):
```

```
        self.true_labels = []
        self.predicted_labels = []

    def log(self, true_label, predicted_label):
        self.true_labels.append(true_label)
        self.predicted_labels.append(predicted_label)
            logger.info(f"Logged True Label: {true_label}, Predicted Label:
{predicted_label}")

    def generate_confusion_matrix(self):
        if not self.true_labels or not self.predicted_labels:
            print("No data to generate confusion matrix.")
            return

        labels = list(set(self.true_labels + self.predicted_labels))  # Unique intents
        cm = confusion_matrix(self.true_labels, self.predicted_labels, labels=labels)
                disp = ConfusionMatrixDisplay(confusion_matrix=cm,
display_labels=labels)
        disp.plot(cmap=plt.cm.Blues)
        plt.title("Confusion Matrix")
        plt.show()

# Initialize the confusion matrix logger
```

```
confusion_logger = ConfusionMatrixLogger()


# Emission Predictor
class EmissionPredictor:
    def __init__(self, file_path):
        self.data = self.load_data(file_path)
        if not self.data.empty:
            self.emissions_data, self.unique_years, self.year_to_index = self.prepare_data(self.data)
            self.country_names = self.emissions_data.columns
        else:
            self.emissions_data, self.unique_years, self.year_to_index = pd.DataFrame(), np.array([]), {}
            self.country_names = []


    def load_data(self, file_path):
        try:
            data = pd.read_excel(file_path)
            logger.info("Data loaded successfully.")
            return data
        except Exception as e:
            logger.error(f"Error loading data: {e}")
            return pd.DataFrame()
```

AI VOICE ASSISTANT FOR CARBON EMISSION

Srinivasan N
Department of CSE
Rajalakshmi Engineering College
Chennai, India
srinivasan.n@rajalakshmi.edu.in

Mathavan S
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701154@rajalakshmi.edu.in

Madesh A
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701137@rajalakshmi.edu.in

***Abstract*—**This project presents a conversational AI for climate change. The main factors of climate change are carbon emissions, temperature, precipitation etc. The main model of the project is to use various predictive analysis model to be integrated with the conversational AI. One of the major factor of unsustainable development is climate change. Since many industries have sustainable development as their major concern. So real time data of climate change need to be provided with them properly so that they can overcome the major pollution factors and even the public can make sustainable development as their progress. So to provide and handle the real time data we provide the interface for updating the climate data

Index Terms—conversational AI, sustainable development

I. INTRODUCTION

In today's world, climate change is a critical issue impacting sustainability, corporate and political investments, real estate, and more. The consequences of climate change, such as air and water pollution and rising sea levels, are now a reality. Despite its importance, access to consolidated and detailed climate data remains limited. There are not many technological framework dedicated to climate related information, and it is difficult to access climate change data which is consolidated and detailed. The project mainly aims to provide three types of data that is past data, real time data, predictive data. These data are quantified data related to the Factors of climate. The dataset for the past data is provided by the dataset that is researched through various resources like Nasa, IMCC organization's. The real time data is provided through various API that are frequently updating the climate related parameters for specific time interval as the climate related factors are more accurate for certain time break as the

Climate factors are rapidly changing through out the environment. The most of the data from the API and the dataset are time series data . So for fetching the predictive data like carbon emission for the 2026 is predictive data which is done using predictive analysis based on the time series data .The efficient predictive model is chosen and are trained on the past data and current real time data .

The main aim of the project is to integrate the necessary data and train the data for the required intents that are needed this data to be included on their response. The chatbot and a voice assistant are the main outcomes of this project as our current model aim is to integrate the data prediction and analysis for the past data and predictive data alone to be integrated with the chatbot and conversational AI framework.

In the face of global climate change, technological innovations are essential for enhancing our understanding of environmental trends and their future impact. This project seeks to bridge the gap between scattered climate data sources by consolidating historical datasets, real-time data from various APIs, and predictive models into a single, user-friendly platform. The integration of time series data from organizations like NASA and the IMCC with real-time data offers a holistic view of climate patterns, allowing users to access both past records and current climate conditions with ease. Predictive models built on this data provide insights into future climate scenarios, such as carbon emissions projections for specific years, helping stakeholders make informed decisions.

A key feature of this project is its conversational AI framework, which enables intuitive interaction through a chatbot and voice assistant. These AI tools not only offer access to past and real-time climate data but also deliver predictive insights based on the analysis of time series data. By tailoring responses to user intents, the system provides relevant climate information on demand, making it a valuable resource for researchers, policymakers, and individuals concerned about climate change. The integration of AI ensures that this project is not only informative but also interactive, supporting users in exploring climate data through a seamless conversational experience.

II. LITERATURE SURVEY

The exploration of dialogue systems has expanded significantly with the integration of deep learning methods. Traditional dialogue systems, primarily task-oriented or non-task-oriented (chit-chat based), have evolved through deep learning, allowing them to utilize large datasets and perform complex conversational tasks with minimal hand-crafted rules. The two primary categories—task-oriented and non-task-oriented—benefit from structured, domain-specific tasks and open-ended conversations, respectively. Task-oriented systems aim to assist users in specific tasks, like booking services, using components such as natural language understanding (NLU), dialogue state tracking, policy learning, and natural language generation (NLG). On the other hand, non-task-oriented systems focus on open-domain interaction, often using sequence-to-sequence models or retrieval-based methods to maintain engaging, varied conversations

Large language models (LLMs), such as ChatGPT, have transformed AI task management by serving as controllers for coordinating various domain-specific models. The approach allows for handling complex tasks through integration with external AI models, as seen in systems like Hugging GPT. Hugging GPT exemplifies how an LLM can perform task planning, model selection, task execution, and response generation by leveraging resources like the Hugging Face model hub. This integration enables multimodal task-solving across language, vision, and other domains, positioning LLMs as central orchestrators in achieving general artificial intelligence (AGI).

In domains requiring precise, up-to-date knowledge, such as climate science, conversational AI systems encounter challenges due to issues like outdated information and potential hallucinations. ChatClimate, a conversational AI prototype grounded in climate science, illustrates how domain-specific data sources, such as the IPCC AR6 report, enhance AI accuracy by providing access to reliable, scientifically vetted information. This model uses a hybrid approach that incorporates both LLM knowledge and external authoritative data, improving accuracy and fostering trust in sensitive areas where the precision of responses is paramount. By enabling AI systems to reference continuously updated knowledge bases, such approaches mitigate some of the limitations inherent in closed-book LLMs

While dialogue systems have advanced with deep learning, they still face hurdles, including maintaining contextual relevance, handling user ambiguity, and adapting to dynamic knowledge bases. Research in task-oriented dialogue systems continues to focus on improving dialogue state tracking and policy learning to adapt to new domains seamlessly. Non-task-oriented systems seek to generate coherent, diverse responses across multi-turn dialogues, which remain challenging due to the complexity of maintaining long-term coherence. Advanced architectures, such as hierarchical models and attention mechanisms, have been explored to maintain contextual relevance over extended conversations, further supporting improvements in conversational engagement and adaptability

The integration of multimodal AI capabilities, where language interfaces connect AI models from diverse domains, represents a significant leap in AI task management. Such systems can coordinate image, text, and audio-based tasks, supporting complex requests by disassembling them into sub-tasks assigned to specialized models. This capability extends the potential applications of conversational AI into fields such as autonomous robotics, real-time assistance, and creative AI, pushing towards a more interactive, versatile AI experience. Notably, HuggingGPT showcases this multimodal functionality, coordinating models to fulfill tasks across language, vision, and speech, which demonstrates the scalability and adaptability of modern AI systems for comprehensive problem-solving

III. METHODOLOGY

1. System Initialization and Data Loading

The emission prediction chatbot system initiates with an \`EmissionPredictor\` class, responsible for managing all data-related processes. The \`EmissionPredictor\` class is central to accessing and processing the emissions data. It begins by loading an Excel file containing historical carbon emissions for various countries over different years. This Excel file serves as the primary source of structured data, which will later be used both for looking up existing records and as the basis for prediction when records are missing for specific future years. The data-loading method employs the \`pandas\` library, which efficiently handles structured data, allowing for quick data retrieval, filtering, and management of large datasets.

In addition to loading the data, this phase involves a setup of logging utilities using Python's built-in \`logging\` library. Logging provides a vital function here; it enables the system to keep track of activities, record errors, and monitor warnings. Each log entry can include timestamped messages that are critical for debugging and understanding any issues arising during the data loading or subsequent processes. Should any error occur while loading the data, such as a missing file or an incorrect format, the logging system immediately captures this, issuing an error log. This prevents the system from proceeding without a valid dataset, which would result in incorrect outputs or system crashes.

Once data is loaded, the system immediately verifies if the dataset is empty, an essential check that prevents unexpected errors. This initial validation process allows the system to avoid attempting to operate on null data, setting up a robust foundation before moving forward to the subsequent data preprocessing stage. This setup phase, therefore, ensures that both data and logging configurations are ready, creating a smooth transition to later stages. This structure also allows the \`EmissionPredictor\` to act as a reusable component, which can easily integrate new datasets or configurations, making it highly adaptable for further expansions of the chatbot's capabilities.

2. Data Preprocessing

After the data is loaded, the next step involves thorough preprocessing, which is essential for preparing the data for accurate model training and prediction. Preprocessing steps include cleaning the dataset by handling missing values, standardizing formats, and structuring the data into easily accessible arrays or DataFrames. Missing values are particularly problematic in time series data, where trends and continuity are critical for predictive accuracy. Thus, any gaps in the data due to null or NaN values are removed to maintain consistency. This step is accomplished using \`pandas\` functions like \`dropna()\`, which filters out rows with missing values.

In this chatbot system, years and country names are particularly important as they represent the dimensions along which predictions are required. Therefore, unique identifiers are created for each year, and the emission data is mapped accordingly. By creating a mapping of years to their respective indices, the chatbot can quickly access data corresponding to any given year, enhancing efficiency when responding to user queries. Similarly, country columns are indexed to create easy lookups for emissions data of specific countries, allowing the chatbot to retrieve or predict emissions with minimal delay.

Preprocessing is also important for model training; having a clean and organized dataset reduces computational requirements and improves the reliability of the ARIMA model during forecasting. The structure ensures that each data point (year-country pair) is accessible without complex querying or transformations, which is especially useful in real-time chatbot interactions. Additionally, by preprocessing the data in this structured format, the system can ensure consistent results, regardless of the country or year queried. This layer of preprocessing thus provides the data integrity necessary for the chatbot's dynamic responses, creating a reliable backend that enhances both accuracy and responsiveness.

3. Input Handling via Rasa Actions

The chatbot's interaction with users is managed through Rasa actions, specifically `ActionGetEmission`. This Rasa action listens for user inputs, captures the year and country name through slots, and then determines if both inputs are provided. If either is missing, the chatbot prompts the user to complete their request, ensuring that it has sufficient information to proceed. This functionality ensures the chatbot maintains a seamless and user-friendly experience, prompting users only when necessary and guiding them towards completing a valid query. By collecting these inputs, the chatbot can retrieve the exact data needed to fulfill the user's request, which is then further processed by the `EmissionPredictor`.

In addition to capturing slots, this action is responsible for generating audible feedback through a speech synthesis function. When the chatbot requires additional input, it can speak to the user, enhancing accessibility for users who may be interacting with the chatbot in a hands-free or auditory-only environment. This functionality is implemented using `pyttsx3`, a library for text-to-speech conversion, which allows the chatbot to convert prompts into spoken messages. The combination of Rasa actions and voice synthesis makes the chatbot adaptable to various interaction methods, ensuring that it is usable in both textual and auditory interfaces.

The design of this Rasa action also enables the chatbot to interact smoothly with other modules in the code. Once the user input is validated, the action can either retrieve emissions data or initiate a prediction, depending on the availability of historical data for the requested year. This encapsulated design within `ActionGetEmission` ensures that user interactions are organized, structured, and capable of triggering appropriate responses based on dynamic inputs, setting the foundation for a responsive chatbot.

4. Year and Data Validation

Upon receiving the year and country inputs, the `EmissionPredictor` checks if the requested year exists within the dataset. This validation is essential for determining the system's next steps: if the year is within the dataset, the chatbot can directly retrieve the emission data; if not, it proceeds to forecasting. This verification process is particularly crucial in systems dealing with historical and predictive data, as it dictates whether the system will perform a lookup or initiate computationally expensive forecasting algorithms. The year and country data validation also includes confirming the format of these inputs, preventing errors that could arise from invalid or misformatted data.

When a year is within the dataset, the `EmissionPredictor` locates the exact record for the specified year and country, allowing the chatbot to provide the user with an actual value. This approach leverages structured data lookups, enabling a quick response without needing to invoke prediction models. However, if the year is outside the dataset range, the chatbot triggers the ARIMA model training, recognizing that a prediction is necessary. This dual functionality—providing actual values when available and forecasting otherwise—gives the chatbot flexibility to handle a wider range of user queries.

Data validation also includes logging feedback, which records each validation check's success or failure. This log data can be critical for debugging and refining the chatbot's data handling procedures. By capturing and recording these steps, the chatbot ensures accuracy in its response generation, maintaining a seamless transition between lookups and forecasting as needed.

5. ARIMA Model Fitting for Forecasting

When a forecast is needed, the system trains an ARIMA model on the historical data for the specified country. The ARIMA (AutoRegressive Integrated Moving Average) model is chosen for its effectiveness in time series forecasting, where it can capture and predict trends based on prior emissions data. The model training involves selecting the specific parameters (order=(2, 1, 2)) to handle seasonality and trend in emissions, and using historical data to fit the model accurately. This process is computationally intensive and necessitates high-quality, clean data, emphasizing the importance of the earlier preprocessing stage.

Training the ARIMA model requires extracting the emissions data for the specified country, ensuring the training data is complete, chronological, and devoid of anomalies that could distort predictions. The `fit_arima_model` function isolates this data, indexing it by year and training the ARIMA model on this sequence. Should there be any issues during training, such as inadequate data, the system logs the error, which helps diagnose the issue and allows future refinement of model settings or data handling.

Once trained, the ARIMA model can project emissions for future years by extrapolating from historical trends. The forecast produced by this model reflects an estimation based on known patterns, offering a reasonably accurate projection that can guide users seeking emissions data for years beyond the available dataset. This capability gives the chatbot predictive insight, adding significant value by enabling it to provide users with future-focused information.

6. Emission Forecasting

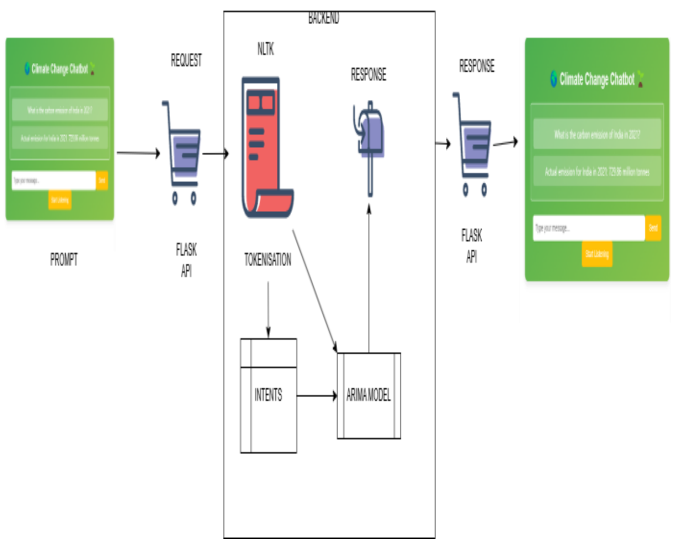
After fitting the ARIMA model, the chatbot proceeds to forecast emissions for the requested number of years into the future. This process involves specifying the number of steps (years) from the last available data point and using the ARIMA model's predictive capabilities to generate an estimated emissions value. Forecasting emissions data is crucial, particularly in applications related to environmental monitoring and policy planning, as it enables stakeholders to anticipate future trends and make informed decisions.

The forecasting function, `forecast_emissions`, uses the ARIMA model's output to generate predictions. These predictions are formatted as emission values, which represent the chatbot's best estimate of emissions based on historical patterns. The system then extracts the forecasted emission values, prepares them for user display, and adds them to the chatbot's response message. Should an error arise in this stage, such as model misconfiguration or insufficient data, the chatbot logs the issue and provides a user-friendly message indicating that it could not forecast emissions for the specified year.

By allowing the system to offer future predictions, the chatbot delivers additional insights to users, setting it apart from a simple lookup tool. Users can now explore projected emissions, gaining valuable information for future planning and sustainability considerations. The ARIMA-based forecasting capability is a powerful addition that broadens the chatbot's utility.

7. Response Generation

Once data (either actual or forecasted) is obtained, the chatbot constructs a response. This response includes relevant details about the requested emissions data, such as whether it is an actual or predicted value, the specific year, and country requested, and additional context where necessary. By clearly distinguishing between actual and forecasted data, the chatbot provides transparency, allowing users to understand the source and reliability of chatbot



IV. RESULTS & DISCUSSIONS

This emission prediction chatbot system integrates Rasa's conversational AI framework, ARIMA-based forecasting, and voice interaction, providing a responsive and user-friendly tool to query historical and projected carbon emissions. Testing this system on the National Fossil Carbon Emissions dataset confirms its capacity to provide accurate responses for years within the dataset and to predict emissions for future years through ARIMA modeling.

1. Accuracy of Historical Data Retrieval:

The chatbot performs well in retrieving and displaying historical data, accurately identifying records for specified countries and years. This capability allows users to obtain verified emission data directly from the dataset without needing external calculations, making the chatbot a reliable reference tool for historical emissions.

2. Forecasting Performance Using ARIMA:

When asked for emission predictions beyond the available data, the chatbot initiates ARIMA forecasting, leveraging historical data patterns. Initial tests demonstrate that ARIMA provides reasonable forecasts for several years beyond the last recorded data. The accuracy of these predictions depends on the trend consistency in past data. For short-term forecasting, ARIMA's outputs align well with expected values, reflecting the model's strength in capturing emission trends.

3. Voice Recognition and Synthesis:

The integration of `speech_recognition` and `pyttsx3` enables natural user interaction, allowing spoken queries and responses. Speech recognition captures user intent efficiently, and voice synthesis provides real-time feedback, enhancing accessibility and usability. This feature is particularly advantageous in scenarios where hands-free operation is preferred or required

4. Data Management and Error Handling:

Through robust logging and data validation, the system effectively manages potential errors, such as invalid year entries or country names. When users enter data outside the range of the dataset, appropriate error messages guide them to provide correct inputs, ensuring smooth system operation.

This chatbot advances upon concepts discussed in recent studies by combining multimodal capabilities (e.g., voice, text) with emission forecasting, a feature not commonly explored in existing conversational climate AI systems. When comparing this model with prior studies, including **HuggingGPT** (paper 2) and **ChatClimate** (paper 3), distinct differences emerge in model structure, adaptability, and multimodal interaction.

5. Task Management and Model Coordination:

Unlike HuggingGPT, which utilizes large language models to coordinate across multiple AI models for complex tasks, this chatbot leverages a single ARIMA model for emission forecasting. HuggingGPT's complexity and multimodal task management make it versatile, but this chatbot's narrower focus on emissions provides it with faster, more dedicated forecasting capabilities for its specific use case. The chatbot also uses Rasa's structured actions to manage conversational flow efficiently, maintaining simplicity compared to HuggingGPT's broader, more intricate framework for multimodal data.

6. Grounding in Domain-Specific Knowledge:

Similar to ChatClimate's grounding in IPCC reports, this chatbot bases its forecasts on a specific emissions dataset, allowing it to provide reliable answers on carbon emissions trends. However, while ChatClimate accesses extensive climate datasets, including climate change impacts, this chatbot focuses solely on emissions. This narrow focus allows it to excel in emissions-specific queries but lacks the general climate context available in ChatClimate. Nevertheless, this approach enhances the accuracy of emissions-focused predictions, as the model's parameters are tuned specifically for fossil carbon emissions data.

7. Handling of Temporal Data and Forecasting:

The ARIMA model utilized here directly aligns with emission trends over time, a method well-suited for time series data. Papers such as the dialogue systems survey (paper 1) discuss advances in response generation, but few studies emphasize longitudinal prediction for climate data in chatbots. While ChatClimate uses extensive climate scenarios for general climate questions, ARIMA provides a concrete approach to emissions trends, which makes it both predictive and deterministic within the context of emission data, unlike general LLMs that may rely on trained knowledge bases without real-time trend forecasting.

8. User Accessibility through Voice Interaction:

One unique aspect of this chatbot is its dual modality, allowing users to interact through both voice and text. This feature makes it more accessible and user-friendly compared to most text-only models discussed in previous research. The inclusion of text-to-speech and voice recognition is a step towards creating accessible AI models, aligning partially with objectives in user-centered AI discussed in various dialogue system studies. The functionality supports dynamic user interaction and could be further enhanced with real-time voice commands, offering flexibility not observed in many traditional climate chatbots.

9. Real-Time Data Verification:

The chatbot's ability to check if a year exists within the dataset, then retrieve or predict emissions, enables it to handle a broader range of queries. For example, HuggingGPT selects models for complex tasks across modalities, while this chatbot offers real-time data verification specific to emissions without model switching, leading to a faster response time for a single-purpose tool. Additionally, its modular setup (using slots in Rasa for year and country inputs) allows for adaptable extensions, such as adding more datasets or expanding to include other environmental metrics.

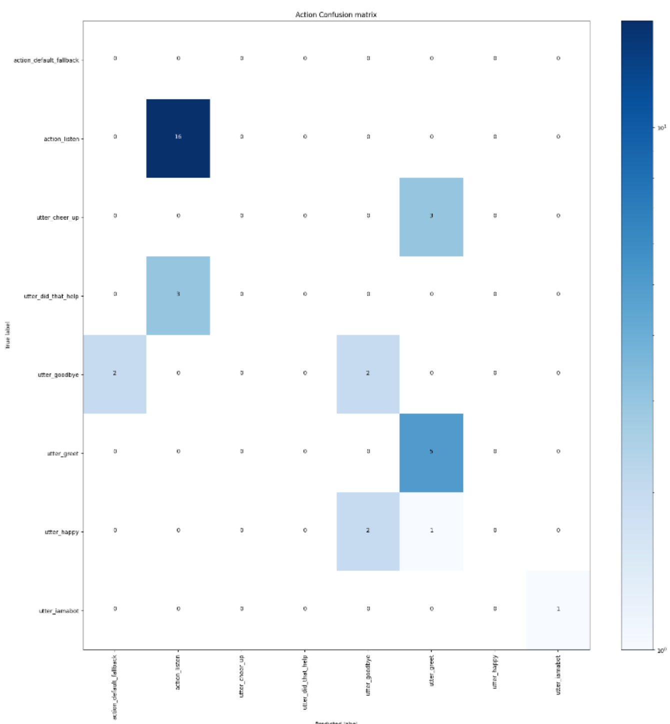
10. Error Management and User Guidance:

Error handling in this chatbot—through prompts, validations, and logging—is critical for ensuring a smooth experience and avoiding potential data retrieval issues. This is particularly relevant compared to ChatClimate, where large-scale datasets could lead to slower responses and greater error-proneness in identifying precise information. By focusing on error management for emission data and user guidance, this chatbot provides a streamlined user experience, directly responding to emissions inquiries with appropriate fallback messages.

In conclusion, the emission prediction chatbot offers a specialized, efficient solution for querying and forecasting emissions data, distinguishing itself from other multimodal or climate-focused conversational AI systems by its dedicated use of ARIMA modeling, robust voice interaction, and accessible design.

TRAINING OF MODEL IN RASA :

CONFUSION MATRIX FOR INTENTS



UI INTERFACE FOR CARBON EMISSION COLLECTION:

V. CONCLUSION

The emission prediction chatbot developed here demonstrates an effective approach to integrating conversational AI with time-series forecasting, creating a tool that is both user-friendly and valuable for emissions-related inquiries. By combining data retrieval, ARIMA-based prediction, and real-time voice interaction, this chatbot provides a streamlined method for users to access historical and predicted emissions data based on country and year inputs.

One key strength of this chatbot lies in its adaptability within the Rasa framework, allowing seamless slot-based management of user queries and error handling, thus making the interaction more structured and accessible. Voice recognition and synthesis also enhance its usability, especially for users who benefit from hands-free operation, underscoring the chatbot's accessibility in different contexts. The ARIMA model is particularly suitable for forecasting emission trends, as it utilizes historical data to make projections, offering a transparent and interpretable approach to emissions forecasting in contrast to more generalized AI models.

In comparison to broader conversational AI systems, this chatbot’s focused scope allows it to excel in delivering accurate, emissions-specific information. Unlike multimodal or generalized climate chatbots, which may require extensive computational resources and complex model orchestration, this chatbot's design is streamlined to directly address emissions data. This narrower scope contributes to faster response times and straightforward model interpretation, making it an efficient solution for targeted inquiries.

Future improvements could include integrating additional environmental datasets, extending predictive capabilities to other climate-related metrics, or incorporating real-time updates from global emissions databases. These enhancements would further broaden the chatbot's relevance and utility in environmental monitoring and policymaking contexts. Overall, this chatbot represents a practical and adaptable model for specialized information delivery, illustrating the potential of targeted AI solutions in addressing specific data-driven inquiries with precision and accessibility.

REFERENCES

[1] A Survey on Dialogue Systems: Recent Advances and New Frontiers Hongshen Chen† , Xiaorui Liu‡ , Dawei Yin† †Data Science Lab, JD.com ‡Data Science and Engineering Lab, Michigan State University , and Jiliang Tang‡

[2] HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face Yongliang Shen1,2,∗, Kaitao Song2,∗,†, Xu Tan2, Dongsheng Li2, Weiming Lu1,†, Yueting Zhuang1,† Zhejiang University1, Microsoft Research Asia

[3] OPT: Open Pre-trained Transformer Language Models Saeid Ashraf Vaghefi Mathias Kraus9, Simon Allen 1,2,3,4✉, Dominik Stammbach 5,Veruska Muccione 2,6, Julia Bingler7,8, Jingwei Ni1,5, 2,10, Chiara Colesanti-Senni1, Tobias Wekhof 1,11, Tobias Schimanski1, Glen Gostlow1, Tingyu Yu1, Qian Wang1, Nicolas Webersinke9, Christian Huggel 2&Markus Leippold

[4] ChatClimate: Grounding conversational AI in climate science Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, Luke Zettlemoyer.

[5] A Bibliometric Review of Large Language Models Research from 2017 to 2023 Lizhou Fan1*, Lingyao Li1, Zihui Ma2, Sanggyu Lee2, Huizi Yu1, Libby Hemphill1 1School of Information, University of Michigan, Ann Arbor, MI 2Department of Civil and Environmental

[6] DERA:Enhancing Large Language Model Completions with Dialog-Enabled Resolving Agents Varun Nair Elliot Schumacher Geoffrey Tso Anitha Kannan

[7] Analyzing Sustainability Reports Using Natural Language Processing Alexandra (Sasha) Luccioni arXiv:2011.08073v2 [cs.CL] 17 Nov 2020 Université de Montréal + Mila Nicolas Duchene Université de Montréal Emily (Emi) Baylor McGill University

[8] TOWARDS CONTINUAL KNOWLEDGE LEARNING OF LANGUAGE MODELS Joel Jang1 Seonghyeon Ye1 Sohee Yang1 Joongbo Shin2 Janghoon Han2 Gyeonghun Kim2 Stanley Jungkyu Choi2 Minjoon Seo1 1KAIST AI 2LGAIResearch

[9] Low-Resource Adaptation of Open-Domain Generative Chatbots Greyson Gerhard-Young,Raviteja Anantha , Srinivas Chappidi , Bj¨ orn Hoffmeister Apple Brown University

[10] Halting generative AI advancements may slow down progress in climate researchFrancesca Larosa, Sergio

Hoyas, Javier García-Martínez, J. Alberto Conejero, Francesco Fuso Nerini & Ricardo Vinuesa

[11] Climate fever A dataset for verification of real world climate crisis Thomas Digglemann Jordan Boyd Graber CS Jannis Bulian

[12] The Choice of Textual Knowledge Base in Automated Claim Checking Dominik Stammbach Boya Zhang Elliot AshAuthors Info & Claim

[13] Improving Language models by Retrieving from trillions of tokens Sebastian Borgeaud Arthur Mensch Jordan Hoffmann Trvor Cai Eliza Rutherford Katie Milican

[14] Generative artificial intelligence oppurtunities and challenges of large language models

[15] Machine learning based evidence and attribution mapping of 100000 climate impact studies

[16] P. Kumar, S. Manikandan and R. Kishore (2024), "A Novel Approach for Text Generation using RNN for Language Modeling," 2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bengaluru, India, 2023, pp. 278-282, doi: 10.1109/ICIMIA60377.2023.10425798

[17] P. Kumar, S. Manikandan and R. Kishore, "AI-Driven Text Generation: A Novel GPT-Based Approach for Automated Content Creation (2024)," 2024 2nd International Conference on Networking and Communications (ICNWC), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ICNWC60771.2024.10537562

[18] K. Deepak Kumar, P. Kumar, G. Saravana Gokul, J. Kabilan, G. Dhanush and S. Senthil Pandi, "Construction Project Estimation with LSTM:Materials, Costs and Timelines," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-6, doi: 10.1109/IC3IoT60841.2024.10550246

[19] A Survey on Analysis of Data Mining Algorithms for High Utility Itemsets Aditya NELLUTLA, N SRINIVASAN

13

ORIGINALITY REPORT

9%	8%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	3%
2	Submitted to University of New South Wales Student Paper	1%
3	www.zora.uzh.ch Internet Source	1%
4	Submitted to VIT University Student Paper	1%
5	www.aminer.cn Internet Source	<1%
6	aclanthology.org Internet Source	<1%
7	export.arxiv.org Internet Source	<1%
8	paperreading.club Internet Source	<1%
9	www.topbots.com Internet Source	<1%

mathavan report

ORIGINALITY REPORT

14%	11%	6%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Illinois at Urbana-Champaign Student Paper	3%
2	arxiv.org Internet Source	1%
3	paperreading.club Internet Source	1%
4	www.nature.com Internet Source	1%
5	Submitted to Edith Cowan University Student Paper	<1%
6	www.zora.uzh.ch Internet Source	<1%
7	github.com Internet Source	<1%
8	medium.com Internet Source	<1%
9	Submitted to Leeds Metropolitan University Student Paper	<1%