TalkGenie: Intent and Sentiment-Aware Multilingual Chatbot

A Mini - Project Report Submitted in the
Partial Fulfillment of the Requirements
for the Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING (AI&ML)

Submitted by

DHULIPALA SAISUBRAHMANYA SARANYA	**22881A6620**
DARA ANIRUDH	**23885A6603**
KOPPULA POOJA REDDY	**23885A6604**
SHANIGARAM RAHUL	**23885A6606**

SUPERVISOR
Mr. N S S S GIRISH KUMAR
Assistant Professor

VARDHAMAN
COLLEGE OF ENGINEERING

Department of Computer Science and Engineering(AI&ML)

April, 2025

VARDHAMAN
COLLEGE OF ENGINEERING

Department of Computer Science and Engineering (AI&ML)

CERTIFICATE

This is to certify that the Mini - Project (A8041) titled **TalkGenie: Intent and Sentiment-Aware Multilingual Chatbot** is carried out by

DHULIPALA SAISUBRAHMANYA SARANYA	**22881A6620**
DARA ANIRUDH	**23885A6603**
KOPPULA POOJA REDDY	**23885A6604**
SHANIGARAM RAHUL	**23885A6606**

in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (AI&ML)** during the year 2024-25.

Signature of Supervisor
Mr. N S S S GIRISH KUMAR
Assistant Professor
Dept of CSE(AI&ML)

Signature of the HOD
Dr. M.A. Jabbar
Professor and Head
Dept of CSE(AI&ML)

Mini - Project Viva-Voce held on ______________________

Examiner

Kacharam (V), Shamshabad (M), Ranga Reddy (Dist.)–501218, Hyderabad, T.S.
Ph: 08413-253335, 253201, Fax: 08413-253482, www.vardhaman.org

Acknowledgements

The satisfaction that accompanies the successful completion of the task would be put incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

We wish to express our deep sense of gratitude to **Mr. N S S S GIRISH KUMAR**, Assistant Professor and Project Supervisor, Department of Computer Science and Engineering (AI&ML), Vardhaman College of Engineering, for his able guidance and useful suggestions, which helped us in completing the mini - project in time.

We are particularly thankful to **Dr. M.A. Jabbar**, the Head of the Department, Department of Computer Science and Engineering (AI&ML), his guidance, intense support and encouragement, which helped us to mould our mini -project into a successful one.

We show gratitude to our honorable Principal **Dr. J.V.R. Ravindra**, for providing all facilities and support.

We avail this opportunity to express our deep sense of gratitude and heartful thanks to **Dr. Teegala Vijender Reddy**, Chairman, **Sri Teegala Upender Reddy**, Secretary, **Mr. M. Rajasekhar Reddy**, Vice Chairman, **Mr. E. Prabhakar Reddy**, Treasurer of VCE for providing a congenial atmosphere to complete this mini - project successfully.

We also thank all the staff members of Computer Science and Engineering (AI&ML) department for their valuable support and generous advice. Finally thanks to all our friends and family members for their continuous support and enthusiastic help.

DHULIPALA SAISUBRAHMANYA SARANYA

DARA ANIRUDH

KOPPULA POOJA REDDY

SHANIGARAM RAHUL

DECLARATION

We hereby declare that the project titled **"TalkGenie: Intent and Sentiment-Aware Multilingual Chatbot"**, submitted to Vardhaman College of Engineering (Autonomous), affiliated with Jawaharlal Nehru Technological University Hyderabad (JNTUH), in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering (AI & ML), is the result of original work carried out by us.

We further certify that this project report, either in full or in part, has not been previously submitted to any university or institute for the award of any degree or diploma.

DHULIPALA SAISUBRAHMANYA SARANYA

DARA ANIRUDH

KOPPULA POOJA REDDY

SHANIGARAM RAHUL

Abstract

This project proposes TalkGenie, a chatbot that enables users to communicate naturally through speech and text, regardless of their native language. This project addresses the linguistic diversity in today's globalized world, where it is nearly impossible for individuals or organizations to fully understand or communicate in every native language. TalkGenie fills linguistic and emotional gaps by interpreting speech input, identifying user intent and sentiment, and producing context-sensitive responses in a target language of choice. Its architecture comprises integrated modules such as an input handler, text and speech-to-text transcription, language detection and translation, intent and sentiment analysis, response generation, translation and multilingual support, text-to-speech synthesis, output display, and chat history.The project goes beyond simple translation by incorporating intent detection and sentiment analysis models to identify the emotional tone and intent behind the user's message. TalkGenie uses the LLaMA3 language model to generate human-like, context-aware responses. The chatbot also supports multilingual output generation based on a translation system created that delivers the final response in any target language desired by the user. TalkGenie is an intelligent, adaptive, and inclusive communication aide that allows for natural speech and hearing across language differences. By merging sentiment analysis and intent with real-time response generation and support for multiple languages, TalkGenie brings us closer to universal communication. TalkGenie is a revolutionary language conversational companion that revolutionizes multilingual communication by breaking down language, emotional, and disability barriers.

Keywords: TalkGenie, multilingual, sentiment, intent, response, chatbot.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Introduction

In the digital age, communication has become more interconnected, but the language barrier remains a significant obstacle. As societies grow, there is a need for intelligent systems that can facilitate natural and accurate conversations between individuals speaking different languages. This understanding is vital in areas like customer service, healthcare, education, and interpersonal communication where tone, urgency, and purpose drive the interaction as much as the words themselves.To meet this demand, the project presents TalkGenie, a multilingual speech and text-based chatbot to overcome the language gap through advanced Natural Language Processing (NLP) [1].

TalkGenie is a multilingual chatbot designed to bridge the language and emotional gap in human-computer and human-human interaction. TalkGenie is built as a smart virtual assistant that accepts text and voice input, intelligently processes it to identify user intent and user sentiment, and responds with useful answers in the user's choice of language[2]. The system has a very strong mix of AI models and open-source software to enable multilingual conversations to happen as naturally and smoothly as possible.

The system starts by analyzing text or audio input from the user. Whisper [3], an OpenAI-created automatic speech recognition and transcription model, unleashes the actual multilingual potential of the system [3]. Whisper [3] has the capacity to translate even text typed in English into any language the user wishes. After the speech is transcribed, TextBlob library [4], an easy yet efficient NLP library based on NLTK [5], is employed to determine sentiment and classify the intent based on keyword extraction and subjectivity detection [6]. These details give context to the system so that it can provide more appropriate and emotionally intelligent responses.

Once the response text has been generated, TalkGenie enriches the communication experience by converting the text into spoken words using the Google Text-to-Speech (gTTS) service [7]. This accommodates various languages and accents, further enriching the user experience by offering voice outputs in their native or desired language. This audio response feature makes TalkGenie not only available to users who prefer listening to reading but also to users who have visual impairments or literacy issues. To enable the translation of a variety of languages in support and cater to worldwide utilization, TalkGenie includes Deep Translator library support [8]. It will translate the resulting responses to vast numbers of different languages before getting synthesized to be spoken with gTTS [7]. Deep Translator services as an interpreter between the engine for generating a language (LLaMA3) [9] and the engine providing voice (gTTS) [7], making the last audio output send it to users in the set language despite receiving it in various forms.

LangSmith [10], which is a utility created to store chat history temporarily, enables users to go back into past interactions to ensure continuity in conversations and make the assistant more personalized. The feature is priceless where there is tutoring in an educational context or customer support, where it prevents repetition and maintains context in case of subsequent inquiries.

The design of TalkGenie is done with modularity and scalability in mind. The frontend interface is built to be natural and responsive, and the backend, which uses Python, collates all modules of processing—Whisper for transcription [3], TextBlob for detection of sentiment and intent [4], LLaMA3 [9] for generation of responses, Deep Translator for multilingual support, and gTTS [7] for audio output. This modularized approach enables quick upgrades and expansions in the future, e.g., support for more input languages, detection of emotions, or even visual querying.

In short, TalkGenie is an innovative, accessible communication system that adjusts to user requirements across linguistic and modality divides. Through the integration of speech and text input, sentiment and intent analysis, response generation with large language models, multilingual translation, and audio output, TalkGenie provides a human-like conversational experience.

1.2 Background and Motivation

Communication has evolved beyond conventional boundaries, necessitating smarter, more empathetic systems that understand the intent and emotions behind words. This is crucial for effective interaction, whether it's a customer seeking help from a service center, a patient trying to describe symptoms to a healthcare assistant, or a student reaching out for educational support.

Many chatbots and voice assistants exist today, but they are often limited to specific languages or respond based purely on keywords and direct matching techniques. These systems often fail when presented with emotionally complex, ambiguous, or multilingual inputs. Moreover, many of them do not consider the emotional state of the user, resulting in mechanical or irrelevant responses that can frustrate or alienate the user. In contrast, humans consider not just what someone is saying but also how they are saying it — tone, emotion, urgency, and subtle cues all play a critical role in understanding. This is the gap that TalkGenie aims to fill.

The inspiration for TalkGenie stems from real-world communication challenges in multilingual environments. For example, customer service centers in India receive queries in English, Hindi, Telugu, Tamil, Bengali, and dozens of other languages. Language diversity, combined with the user's mood and purpose, creates a complex input space that most current systems are unequipped to handle.

TalkGenie is conceptualized as a comprehensive solution that not only understands different languages but also detects the user's mood and intent, then replies meaningfully. What makes this system more practical and impactful is its ability to summarize user intentions using sentiment and intent classification before response generation. The goal is not just to reply, but to respond intelligently and empathetically.

1.3 Objectives of the Project Work

The TalkGenie project aims to revolutionize human interaction with machines through voice and text-based dialogue systems. Language and emotional disconnects often hinder user interaction with digital platforms, especially for individuals who are not native speakers or are emotionally overwhelmed. TalkGenie aims to revolutionize the way humans interact with machines through voice and text-based dialogue systems. This section elaborates on the five major objectives that form the foundation of the TalkGenie project, each aiming to bridge communication gaps, enhance user experience, and deliver smart, context-aware interactions.

1. To Create a Multilingual, Speech-Based Chatbot for International User Communication

One of the major objectives of TalkGenie is to facilitate multilingual input as well as output for text and speech-based communication. In today's culturally and linguistically diverse global environment, conventional chatbots are inadequate because they cannot comprehend or speak in multiple languages, accents, or dialects. TalkGenie overcomes this problem by providing:

(i) **Audio and Text Input:** The user can interact with the system either through voice or text.

(ii) **Language Independence:** Being enabled with the addition of Whisper speech-to-text [3], TalkGenie is now capable of converting various languages and dialects with accurate transcription.

(iii) **Language-Specific Output:** The generated answer through the latest LLMs is then outputted in the preferred language of the user and, optionally, presented as audio output through TTS systems such as gTTS [7].

This goal makes it possible for users across regions and cultures to use the system with ease in their own language, eliminating the need for English or any prevalent language. It facilitates global inclusivity and accessibility, particularly in areas where digital language support is not extensive.

2. To Detect and Understand User Intent and Sentiment for Context-Aware Conversations:

Another key goal is to move beyond mere keyword detection and create a context-aware conversation system that identifies the intent (what the user intends) as well as the sentiment (the way the user feels). Too many current chatbots fail because they apply the same treatment to each interaction, without regard to the user's tone, mood, or emotional state.

TalkGenie surpasses this using an 80,000-instance dataset, where each instance contains:

(i) A user sentence (input).

(ii) An intent label (e.g., complaint, inquiry, request, feedback).

(iii) A sentiment label (e.g., positive, neutral, negative).

Recognizing sentiment comes into play when users are upset, frustrated, or require immediate assistance.

3. To Generate Accurate, Human-Like Responses Using LLaMA3 and NLP Pipelines:

While intent and sentiment understanding is crucial, it is also critical for a chatbot to create intelligent, grammatically correct, and emotionally suitable responses. The third goal of TalkGenie is to create a smart response generation engine with the help of LLaMA3 (Large Language Model Meta AI) [9] and customized NLP pipelines [1].

Here's how this process functions in TalkGenie:

(i) The user's sentiment and intent are forwarded to the LLaMA3 model[9].

(ii) A prompt is built with contextual hints to direct response generation.

(iii) The model generates a natural, human-sounding sentence that is clear, concise, and sentimentally appropriate.

(iv) The output is not pre-programmed or scripted but generated in real time from user input, allowing flexible, dynamic conversation.

4. To Provide Output Responses in the User's Desired Language, Including Voice Playback:

This entails Language Personalization: Once a response is generated in English (or base language), the system employs translation models to translate it into the desired language of the user. This makes it geographically accessible. With the use of applications such as gTTS[7] the translated answer may be synthesized into speech and played to the user.

This is especially useful for:

(i) Visually impaired users who are more comfortable hearing rather than reading.

(ii) Aging users who prefer listening over reading.

(iii) Those with literacy difficulties or low exposure to written words.

By providing multimodal answers—both speech and text—TalkGenie leaves no user behind, ensuring digital inclusion.

5. To Summarize the User's Core Requirement Clearly for Improved Interaction and Logging

In most real-world systems, users do not always specify their needs in simple, structured terms. They may ramble, complain, or bundle several thoughts into a single input. One of TalkGenie's creative goals is to encapsulate the user's underlying requirement in one or two direct, concise sentences. This aim is vital in scaling the system in areas such as customer support, healthcare triage, or technical support, where time is of the essence and comprehension needs to be accurate.

1.4 Organization of the Report

The report that follows consists of six chapters, which focus on one of the big topics of the project, starting from literature review to application and evaluation:

(i) **Chapter 2:**
Describes up-to-date research and techniques in the recent advancements in Spoken Language Understanding (SLU), user intents and sentiments from Automatic Speech Recognition (ASR).

(ii) **Chapter 3:**
Describes the methodology step by step, including data collection, preprocessing of audio and text inputs, feature extraction for sentiment and intent detection, model architecture, and training process for the multilingual assistant.

(iii) **Chapter 4:**
Explains the technical aspects of the "Talk Genie" system, including the setup of the backend, integration of Whisper [3], LLaMA3 [9], Google Translator, and gTTS [7], and the configuration of the system environment and dependencies.

(iv) **Chapter 5:**
Reports the experimental results of the "Talk Genie" system, such as accuracy in intent and sentiment detection, performance benchmarks, user feedback, and comparative analysis with existing multilingual voice assistants.

(v) **Chapter 6:**
Summarizes the major findings from the project, addresses any challenges or limitations encountered, and presents suggestions for future improvements, including possible enhancements in language support, accuracy, and real-time performance.

CHAPTER 2

Literature Survey

2.1 Introduction

This chapter involves a thorough review of the existing literature concerning recent advancements in Spoken Language Understanding (SLU) [11], focusing on models that enhance intent detection and slot filling. The rapid evolution of pre-trained language models and architectural innovations has significantly improved SLU systems' ability to process and interpret spoken input, especially from Automatic Speech Recognition (ASR) outputs [12]. Five significant research contributions address various challenges in SLU, such as multi-intent detection, ASR noise handling, chunk-level analysis, and low-resource adaptability [13]. These approaches have been found to have enhanced performance through the use of contextual hints and explicit probability-based correlations, which is most relevant to this project's goals.

Frameworks that divide utterances into significant chunks have been suggested to process multiple intents within a single query. Sliding window-based attention mechanisms are used by these models to detect intent changes, which is essential for multi-intent recognition applications—a significant aspect of building intelligent conversational agents [14].

Another important direction has been the improvement of text representations from ASR outputs. ASR tends to produce uncapitalized and unpunctuated text, which impacts downstream tasks such as intent recognition and sentiment analysis. Dedicated models have been presented to recover these textual features, boosting comprehension dramatically across multilingual environments and diverse SLU tasks [15].

Benchmark challenges have also underscored the assessment of SLU systems on quality, efficiency, and flexibility. The challenges point out that pipeline-based models are effective, but end-to-end models require improvement to match such efficiency, particularly in low-resource or on-device settings.

Moreover, prompt-based SLU models have also proved to be effective tools by capitalizing on the strengths of pre-trained language models to integrate intent detection and slot filling into a common learning paradigm [2]. The models enable enhanced feature extraction and task collaboration, enhancing the general accuracy and resilience of SLU systems.

2.2 Review of Prior Research

Past studies as summarized in Table 2.1emphasize the significance of incorporating pre-trained models, improving transcription accuracy, supporting multiple intents, and maintaining robustness across different languages and environments [16]. These developments are a good base for this project, which involves identifying user intent and sentiment from speech or text input through machine learning and big language models, and supporting multilingual and real-time conversational support. Through the integration of methods like Whisper-based transcription [3], logistic regression classification, and LLaMA-based response generation with translation and text-to-speech components, this project stands at the crossroads of state-of-the-art research and real-world implementation in natural language-driven AI assistants. previous work establishes the need to incorporate pre-trained models, improve transcription accuracy, manage multiple intents, and provide robustness across languages and environments [11]. These developments lay a strong foundation for this project, which is to identify user intent and sentiment from speech or text input based on machine learning and large language models, and integrate multilingual and real-time conversational support.

2.3 Identified Research Gaps

Despite significant advancements in the field of Spoken Language Understanding (SLU), Spoken Language Understanding (SLU) has made significant progress, but still faces limitations in its development. One major issue is the insufficient correlation between intent detection and slot filling [2], leading to incomplete or inaccurate interpretations of user queries. BERT-based models also face challenges in SLU tasks, such as joint training of intent detection and slot filling modules, resulting in suboptimal utilization of contextual knowledge [17]. The limited generalization capability of SLU models, particularly when exposed to dynamic, multilingual, or domain-specific data, is critical for real-world deployment. Additionally, existing architectures are not optimized for multi-intent recognition, causing fragmented user experiences. The demand for lightweight, real-time, and low-resource SLU systems is growing, but many models are computationally intensive and not suitable for low-bandwidth or resource-constrained environments. The current project aims to develop a unified and scalable SLU system that integrates intent and sentiment detection, supports multilingual interactions, and delivers real-time responses [11].

2.4 Summary

This study aims to develop a real-time multilingual conversational AI system that responds and understands users both in terms of intent and sentiment using either speech or text input. The system integrates machine learning and large language models (LLMs) to process input data, seeking to address the shortcomings of earlier Spoken Language Understanding (SLU) systems [18].

Although previous models have achieved improvements through the use of pre-trained models such as BERT [17], they are poor at joint learning of intent detection and slot filling, have poor generalization across datasets, and poor multi-intent handling. They also do not adapt well in low-resource and multilingual settings. This project overcomes these limitations by incorporating elements such as Whisper for transcription [3], logistic regression for initial intent/sentiment detection, LLaMA3 [9] for smart response generation, Google Translate for multilingual support, and gTTS [7] for voice output.

The objective is to build a chatbot system which is accurate, scalable, multilingual, and responsive, applicable for real-time usage across multiple languages and platforms, particularly conversational AI applications.

Table 2.1: Summary of Contemporary Research in Spoken Language Understanding

Reference No.	**Methodology Used**	**Key Contributions**	**Limitations**	**Future Scope**
[17]	Probability-aware gate mechanism with fine-tuning on BERT.	- Models correlation between intent and slot. - Beats SOTA models on ATIS/Snips.	- Not optimized for real-time SLU. - Limited linguistic dataset diversity.	- Test on multilingual, real-world SLU. - Explore deep learning integration.
[13]	Sliding window-based self-attention (SWSA).	- Segments utterances for better intent accuracy. - Excels on MixATIS, MixSNIPS.	- Issues with overlapping intents. - Lacks real-world tests.	- Apply to multilingual SLU. - Refine intent transitions.
[18]	RoBERTa with Mask Capu objective.	- Enhances NER, POS, chunking in 3 languages. - Handles ASR-generated text well.	- Limited language support. - Needs diverse ASR output tests.	- Scale to low-resource multilingual SLU. - Boost ASR error robustness.
[15]	Compares pipeline and E2E models.	- Defines benchmarks for quality and efficiency. - Pipeline models excel.	- E2E models still trail behind. - On-device testing gaps.	- Improve E2E efficiency. - Enhance low-resource adaptability.
[14]	Uses PLMs + Semantic Intent Guidance.	- Combines ID and SF. - Hits SOTA on MixATIS/SNIPS.	- SIG errors can affect SF. - Needs domain fine-tuning.	- Expand to low-resource SLU. - Boost ASR robustness.

CHAPTER 3

Methodology

3.1 Introduction

TalkGenie is a multilingual chatbot that supports both speech and text inputs, identifying user intent and sentiment with precision. The system works through a pipeline that includes transcription, sentiment analysis [19], response generation, translation, and audio output. The chatbot takes user input in either text or speech format [3], with text inputs in English and audio inputs taken through Whisper [3]. The TextBlob library identifies user intent and sentiment, allowing TalkGenie to understand context and respond accordingly. An 80,000 labeled sentence structured dataset reinforces the system's understanding.

LLaMA3, a Large Language Model, provides a coherent English response, which can be shown as text or translated into spoken audio using Google Text-to-Speech (gTTS). Deep Translator is integrated for users who want the response in a language other than English, providing custom content and language. LangSmith temporarily stores the chat history to ensure context during the session. The TalkGenie process integrates advanced transcription, linguistic analysis, response generation, multilingual translation, and audio output in one seamless process, allowing smooth, intelligent, and personalized human-to-machine and machine-to-human conversations. The chatbot uses Whisper [3], a speech recognition model, to convert written words from speech to text, and uses LLaMA3 to present meaningful, relevant English responses.

3.2 Problem Statement

In today's globalized and technologically driven world, communication has become the cornerstone of personal, professional, and cross-cultural interactions. However, the diversity of human languages and the difficulty of real-time, meaningful communication between people who speak different languages remain significant challenges. Even with the development in artificial intelligence and natural language processing, there remains a large gap in creating a strong, easy-to-use solution that is capable of overcoming language barriers, allowing both speech and text inputs, and interpreting the user's intent and emotions correctly. [16] There are more than 7,000 languages spoken in the world, and India alone has 121 languages spoken by more than 10,000 individuals each. Users tend to struggle with conveying themselves effectively or being understood because of language differences. Existing chatbot technologies have limited capacity for effective multilingual and multimodal communication since they are text-based, only handle a few main languages, and do not involve real-time voice interaction. Moreover, they usually do not catch the emotional context or purpose of a user's message and hence respond mechanically or irrelevantly.

TalkGenie brings to the platform a multilingual voice chatbot with an embedded suite of innovative tools and technology for delivering smooth and naturalistic communication experiences. The system has speech or text inputs, making use of the Whisper speech input [3] and TextBlob library for the text input. With this dual-layered understanding, the chatbot can both answer and engage smartly as well as sympathetically. An 80,000 instance curated dataset was employed for training and fine-tuning intent and sentiment identification to enable the chatbot to translate user inputs into many contextual meanings and emotional states.[20]

Once the user's message is understood, LLaMA3, a Large Language Model designed for producing coherent and human-like responses, is employed to create a response. Deep Translator is then used to translate the response into the user's desired language, overcoming the language barrier and providing a more personalized interaction. After the response has been prepared in the desired language, it is transformed into audio speech using Google Text-to-Speech (gTTS) so that users can hear the reply of the chatbot in their desired language.[21]

LangSmith is utilized to cache the chat history temporarily, allowing the chatbot to retain context within multi-turn dialogue and return more coherent, context-sensitive responses. This project not only seeks to enable fluent communication between languages but also improve accessibility, empathy, and user engagement through next-generation speech-based AI.

3.3 Proposed Methodology

TalkGenie is an AI-powered chatbot designed to facilitate seamless communication between humans and machines. It uses advanced tools to process text and voice inputs, identify intent and sentiment, produce intelligent responses, and provide outputs in the desired language and format. As shown in the Figure 3.1 the system accepts input in two forms: text and speech. If the user enters text, it is translated into English and directed to the language understanding module. If the user enters speech, Whisper is used to translate the speech into readable English text[3].

The TextBlob library is used to identify the user's intent and sentiment, providing sentiment polarity and simple intent classification [4]. Sentiment analysis helps the system respond more empathetically, changing the tone and emotion of the response according to the user's emotional state. Once the user's intent and sentiment are discerned, the system produces a natural language and context-sensitive response based on the LLaMA3 language model, making the chatbot sound both intelligent and human.

The Deep Translator library is employed to translate the response into the user's preferred output language, overcoming language barriers and improving usability for non-English speakers. Once translated, Google Text-to-Speech (gTTS) is used to convert the text-based response into speech, making it particularly helpful for users with reading disabilities or visual impairments [7].

LangSmith is incorporated into the system to temporarily cache chat history within a session, enhancing coherence in multi-turn dialogue [10]. This approach combines various state-of-the-art libraries and tools to handle text and speech inputs, identify intent and sentiment, create smart responses, and provide output in the user's selected language and mode.

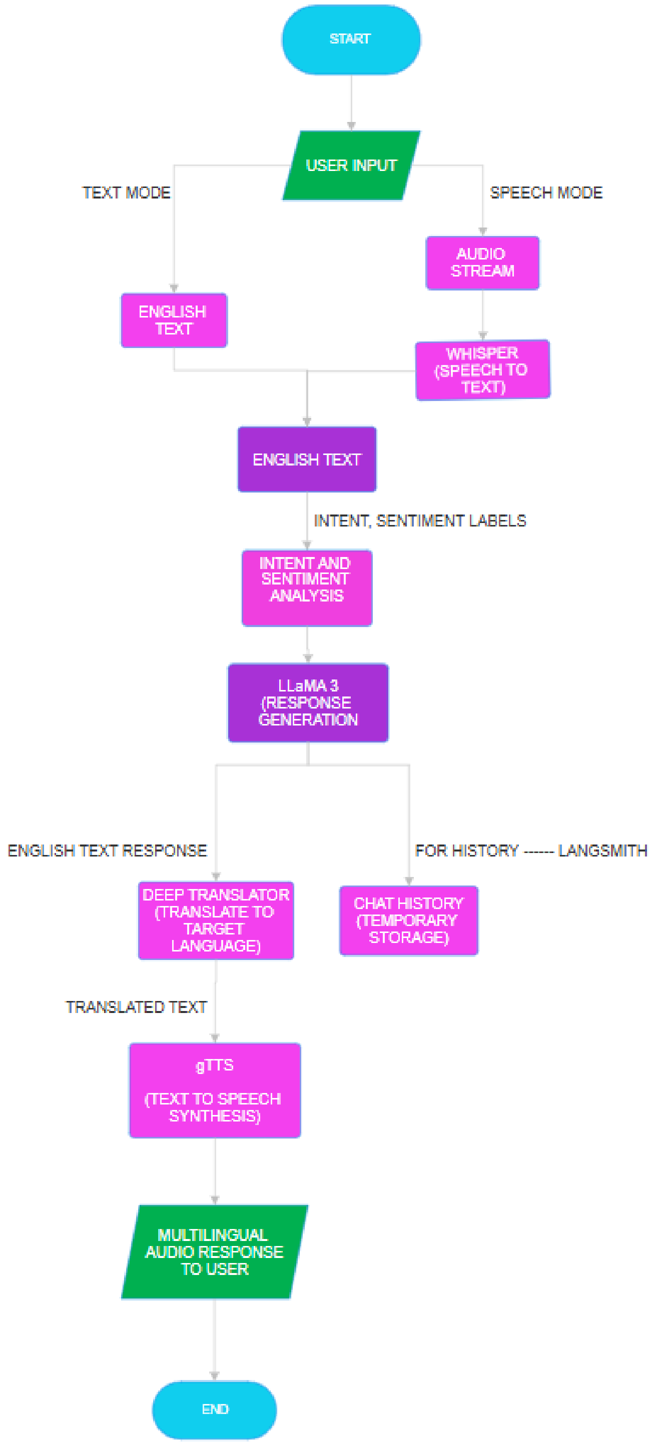

Figure 3.1: Proposed System Architecture

TalkGenie is a multilingual, speech-enabled AI chatbot designed to enable smooth and smart communication between humans and machines. It uses advanced tools to handle text and voice inputs, identify intent and sentiment, produce intelligent responses, and provide output in the user's preferred language and format.

3.3.1 Purpose

The main purpose of developing TalkGenie is Overcoming Language Barriers: TalkGenie intends to overcome the linguistic divide by making communication easy and effortless between languages, facilitating effective communication between people from different language backgrounds. Accepting input in English and delivering output in a variety of languages, it allows users to communicate with one another regardless of languages, thus best suited for worldwide conversations, customer service, and accessibility products [11].

Context-Aware Conversational AI: The system automatically recognizes the intent and sentiment of the input from the user. This helps to ensure TalkGenie gives more precise, relevant, and empathetic responses depending on the emotional tone and conversation purpose of the user. Through the recognition of the user's needs and mood, the chatbot improves the overall experience in communication [22].

Multimodal Communication: TalkGenie enables two-way communication by supporting both text and speech inputs. No matter if the user inputs their query through typing or speaking, TalkGenie is capable of transcribing speech to text, identifying intent, producing a response, and outputting in both text and speech modes, increasing the accessibility and ease of use for a broad category of users.

Improving Accessibility: Text-to-speech (TTS) operation by TalkGenie through gTTS and speech recognition assists in offering accessibility solutions for people with disabilities, such as visually impaired individuals or those unable to read or type. This makes the system inclusive and able to serve users with varying needs, thus increasing usability among diverse populations [7].

3.3.2 Data Collection and Preprocessing

TalkGenie is a smart conversational model that attempts to identify user intent and sentiment in different communication situations. The training and test data for the system was established by developing a broad dataset of 80,000 text sentences sourced from multiple locations as shown in Figure 3.1.

The data started off as raw sentences, gathered from several publicly available datasets and websites. To turn this unlabelled corpus into a useful dataset for intent and sentiment analysis, we have applied a Python library named TextBlob to extract sentiment polarity values for every sentence, which were further translated into categorical sentiment labels [4]. For intent detection,

Table 3.1: Entries from the Dataset

Query	**Intent**	**Sentiment**
Is there a manual or guide for shipping inquiry	shipping_inquiry	neutral
Using booking complaint was a frustrating experience	booking_complaint	negative
what does financial complaint involve	financial_complaint	neutral
technical support has made my life much easier and better	technical_support	positive
how do i proceed with entertainment review	entertainment_review	neutral
using product complaint was a frustrating experience	product_complaint	negative
im very satisfied with the results of service inquiry	service_inquiry	positive
using delivery status was a frustrating experience	delivery_status	negative
where can i learn about shipping inquiry	shipping_inquiry	neutral
really appreciate how efficient travel complaint has been	travel_complaint	positive

a rule-based mapping system was developed using keyword analysis, sentence structure, and contextual tags to map every sentence onto one of 50 predefined intents. This was a combination of rule-based reasoning and keyword matching for large-scale semi-automated annotation. The dataset was thereafter converted into a structured dataset with three columns: Sentence – user input or message; Intent – the reason or setting of the sentence; and Sentiment – emotional undertone behind the message.

The second pivotal move was preprocessing the data so it could be utilized for feeding large language models as well as other downstream modules. The principal operations involved were text normalization, tokenization, removing stop words, lemmatization, treatment of imbalanced classes, and label encoding. Once the preprocessed stage, a final verification of the dataset was carried out to ascertain whether the dataset lost its semantic integrity but remained properly organized for machine understanding [1].

The ultimate product of the dataset preparation process was a purified and enhanced dataset each with a user sentence and its respective intent and sentiment labels. The dataset was subsequently applied as the training and reference base for LLaMA3, which performs the function of intent detection, sentiment analysis, and response generation in the TalkGenie system. This well-framed dataset evolved to be the system's pipeline's most significant asset, ensuring superior-quality user engagement by facilitating LLaMA3 learning and adapting real-world conversational styles in numerous domains and affective tones.

3.3.3 Integrating sentiments and intents with LLaMA3

TalkGenie is a conversational AI that incorporates sentiment and intent with LLaMA3, an efficacious large language model (LLM), to create highly contextual, emotionally intelligent, and goal-driven responses. This synergy enables the chatbot to provide natural, relevant, and human-like interactions, particularly under multilingual and multimodal communication scenarios.

Intent detection is the determination of the purpose or aim of what a user has input, and sentiment analysis is the determination of the emotional direction of the input as being positive, negative, or neutral. In TalkGenie, these aspects have a underlying role in shaping responses. What the user is attempting to accomplish and how they feel when they are doing so is assessed by the system, allowing LLaMA3 to structure responses that are functionally correct and emotionally attuned [9].

To pull out these two dimensions from input, TalkGenie employs the TextBlob library, which analyzes the user's text and provides the polarity (sentiment value) and subjectivity scores. An 80,000-instance custom-trained dataset as shown in Figure 3.1, contains sentences, intents, and sentiment labels is employed to match or infer probable intents. This yields a structured representation of user intent and a sentiment label representing the emotional tone, which is sent to the response generation engine based on LLaMA3.

LLaMA3 is a cutting-edge LLM that performs well in producing coherent and context-sensitive responses. Nevertheless, by default, it does not necessarily include external metadata such as intent and sentiment. In TalkGenie, the model's performance is augmented by including these parameters in the prompt structure passed to LLaMA3.

Benefits of integration are context-aware responses, emotional intelligence, higher accuracy and relevance, improved error handling and user support, and enhanced error handling and user support.

3.4 System Requirements

3.4.1 Software Requirements

1. **Programming Language**

   (i) Python (3.9+) : Core programming language used for backend development, along with AI integration, and speech processing[23].

2. **Libraries & Frameworks**

   (a) **Natural Language Processing (NLP)**

      (i) TextBlob

         (a) Used for intent and sentiment analysis [4].

         (b) Simple API built on top of NLTK [5].

      (ii) Deep-Translator

         (a) For text translation into various languages [8].

(b) Helps in converting responses and queries to multilingual text.

(b) **Speech Processing**

(i) Whisper by OpenAI

(a) For speech-to-text transcription.

(b) Transcribes English audio input into text for processing [24].

(ii) gTTS (Google Text-to-Speech)

(a) Converts text responses to spoken audio in multiple languages.

(b) Provides audio output that enhances accessibility [7].

(c) **Large Language Models**

(i) LLaMA3 (via Ollama or local server)

(a) Used for response generation based on user intent and sentiment.

(b) Can be accessed via API or local inference using tools like Ollama [25].

(d) **Data Handling & Analysis**

(i) Pandas

(a) For working with the 80,000-instance dataset (sentences, intents, sentiments) [26].

(b) Used to load and preprocess training/intent classification data.

(ii) NumPy

(a) For numerical operations where needed [27].

(e) **Web & API Integration**

(i) LangSmith SDK

(a) To store temporary chat history and manage conversation context [10].

(f) **Audio Handling**

(i) pydub

(a) For manipulating audio data (e.g., conversion, trimming, exporting) [28].

(b) Works with gTTS output or preprocessed speech input [7].

(ii) SpeechRecognition (if needed additionally)

(a) For real-time microphone input capture and conversion to audio files [29].

3. **Development Tools**

(i) Jupyter Notebook / VS Code

(a) For developing and testing models and backend logic [30] [31].

4. **Package Managers**

(i) pip

(a) Python's default package manager for installing all required libraries [32].

5. **Frontend**

(i) Streamlit

(a) Used for building an interactive, user-friendly web interface [33].

(b) Allows users to input text or audio, receive AI-generated responses, and play back spoken output.

3.4.2 Hardware Requirements

Table 3.2: Hardware Requirements

Component	Specification
Processor (CPU)	Intel i5
RAM	8 GB
Storage	256 GB SSD (for faster read/write access)
Graphics (GPU)	Required
Audio Input/Output	Microphone and Speakers/Headphones
Operating System	Windows 10/11

3.5 Summary

TalkGenie is a multimodal, multilingual chatbot that has a lean and modular pipeline processing for user input. It employs Whisper [3], which is a speech-to-text model, to translate speech into English, then TextBlob, a library of natural language processing, to determine intent and sentiment[4]. LLaMA3, a large-scale language model, creates a response based on the user's intent and emotional tone. The answer is translated to the user's preferred language via the Deep Translator library [8], and then synthesized to speech via gTTS (Google Text-to-Speech)[7]. LangSmith is employed to cache chat history temporarily, providing contextual continuity between user sessions. The interface is constructed with Streamlit [33], a lightweight and versatile Python-based web framework, to facilitate real-time text and voice interaction via a browser. The project also involves libraries such as pandas [26], numpy [27], pydub [28], and speechrecognition [29]for dealing with audio, text, and datasets. Hardware requirements for local deployment are an Intel i5. 8 GB RAM, 256 GB SSD, microphone, speakers/headphones, and GPU. For cloud or large-scale deployment, Google Colab [34], AWS EC2 [35], Azure [36], or Hugging Face Spaces [37] are suitable. TalkGenie provides an easy yet robust multilingual voice and text communication solution, filling international communication gaps through sophisticated AI-based interaction.

CHAPTER 4

Implementation

4.1 Introduction

The implementation phase of the this project focuses on building an intelligent, multilingual chatbot that integrates various natural language and audio processing tools into a seamless conversational experience.The system receives input in the form of text and speech. The text input is processed by the system, whereas the speech input is processed via Whisper, an automatic speech recognition model [3]. LLaMA3, which is a high-capacity large language model, recognizes the user's intention and reads the message's sentiment to deliver personal, context-driven responses [9].

LLaMA3 also manages response generation, creating intelligent, coherent, and contextually sensitive responses depending on conversation dynamics and emotional context [9]. If the user asks for output in a foreign language, the translated response is rendered as natural-sounding speech using Google Text-to-Speech (gTTS) [7] and played back to the user via the web interface.

The whole front-end of the app is constructed in Streamlit, giving a minimal, interactive UI for users to enter voice and text searches [33]. LangSmith is used to save chat histories temporarily in the backend, allowing the system to carry conversational context across several interactions. This approach to implementation is focused on modularity and user experience, pairing strong language understanding with seamless multilingual support. The combination of Whisper [3], LLaMA3, Deep Translator, gTTS, and Streamlit [33] creates a useful and effective solution for live, speech-controlled communication between languages.

4.2 Output Tables

4.2.1 Language Codes

The system is designed to support multilingual outputs. Table 4.1 lists the language codes that are currently supported by the translation and text-to-speech components of the system. These codes are used internally to identify and process different languages.

Table 4.1: Language Codes

Language	**Code**	**Language**	**Code**
English	en	Punjabi	pa
Hindi	hi	Odia	or
Tamil	ta	Assamese	as
Telugu	te	Urdu	ur
Malayalam	ml	Marathi	mr
Kannada	kn	Konkani	kok
Bengali	bn	Maithili	mai
Gujarati	gu	Santali	sat
Manipuri (Meitei)	mni	Sindhi	sd
Bodo	brx	French	fr

4.2.2 Intent Accuracy of Proposed Model

To evaluate the effectiveness of TalkGenie , several standard metrics were used. The results of this evaluation are summarized in Table 4.2. These metrics provide insights into the model's accuracy and F1-score across different categories.

Table 4.2: Intent Accuracy of Proposed Model

Metric	**Value**
Accuracy	73.90%
Macro Avg F1	0.66
Micro Avg F1	0.74
Weighted F1	0.80

4.2.3 Sentiment Accuracy of Proposed Model

Further optimization of TalkGenie was achieved through training on a specific dataset. The performance metrics after this process are presented in Table 4.3.

Table 4.3: Sentiment Accuracy of Proposed Model

Metric	**Value**
Accuracy	99.43%
Macro Avg F1	0.99
Micro Avg F1	0.99
Weighted F1	0.99

4.3 Summary

The TalkGenie project results prove the effective deployment of a speech-based multilingual, chatbot. The performance of the TalkGenie system is centered on two vital modules: intent identification and sentiment detection—both crucial for facilitating accurate, context-dependent, and emotionally intelligent conversation. These modules are powered by the LLaMA3 model [9], which has been fine-tuned with a well-curated and preprocessed dataset of 80,000 examples. Every record in the dataset contains a sentence and its respective intent and sentiment labels. The training procedure was made in such a way that the model not only understood the nature of human communication but also responded correctly in multilingual settings.

To evaluate the effectiveness of the system, we tested the trained model on a validation set obtained from the same dataset. The outcome suggests a good potential of the model to deal with real-world conversational situations. For the detection of intent, the system was found to have an accuracy of 73.90%, indicating a good comprehension of user purposes and situations through various conversational inputs. It is a considerable achievement, given the variability and vagueness of natural language.

By comparison, the sentiment analysis module performed outstandingly well, reaching 99% accuracy. Such a high rate illustrates the stability of the model in determining the emotional tone of user inputs, be it expressing positive, negative, or neutral sentiment. Successful sentiment identification contributes to greater empathy and relevance in the chatbot's responses, translating into more tailored and productive user interactions.

In general, the findings confirm the feasibility and effectiveness of incorporating large language models such as LLaMA3 into multilingual, sentiment-based AI systems. The performance of TalkGenie in accurately detecting user intent and sentiment provides a solid basis for further advancement and practical implementation in customer service, healthcare communication, education, and other applications.

CHAPTER 5

Results and Discussion

5.1 Introduction

The project results prove the effective deployment of a speech-based multilingual, chatbot. The performance of TalkGenie is centered on two vital modules: intent identification and sentiment detection, both crucial for facilitating accurate, context-dependent, and emotionally intelligent conversation. These modules are powered by the LLaMA3 model, which has been fine-tuned with a well-curated and preprocessed dataset of 80,000 examples. Every record in the dataset contains a sentence and its respective intent and sentiment labels. The training procedure was made in such a way that the model not only understood the nature of human communication but also responded correctly in multilingual settings.

To evaluate the effectiveness of the system, we tested the trained model on a validation set obtained from the same dataset. The outcome suggests a good potential of the model to deal with real-world conversational situations. For the detection of intent, the system was found to have an accuracy of 73.90%, indicating a good comprehension of user purposes and situations through various conversational inputs. It is a considerable achievement, given the variability and vagueness of natural language.

By comparison, the sentiment analysis module performed outstandingly well, reaching 99% accuracy. Such a high rate illustrates the stability of the model in determining the emotional tone of user inputs, be it expressing positive, negative, or neutral sentiment. Successful sentiment identification contributes to greater empathy and relevance in the chatbot's responses, translating into more tailored and productive user interactions.

In general, the findings confirm the feasibility and effectiveness of incorporating large language models such as LLaMA3 into multilingual, sentiment-based

AI systems. The performance of TalkGenie in accurately detecting user intent and sentiment provides a solid basis for further advancement and practical implementation in customer service, healthcare communication, education, and other applications.

5.2 Output Screens

The user interface of the TalkGenie application, showcasing the input and output areas, is presented in Figure 5.1. This provides a visual overview of how users interact with the system.

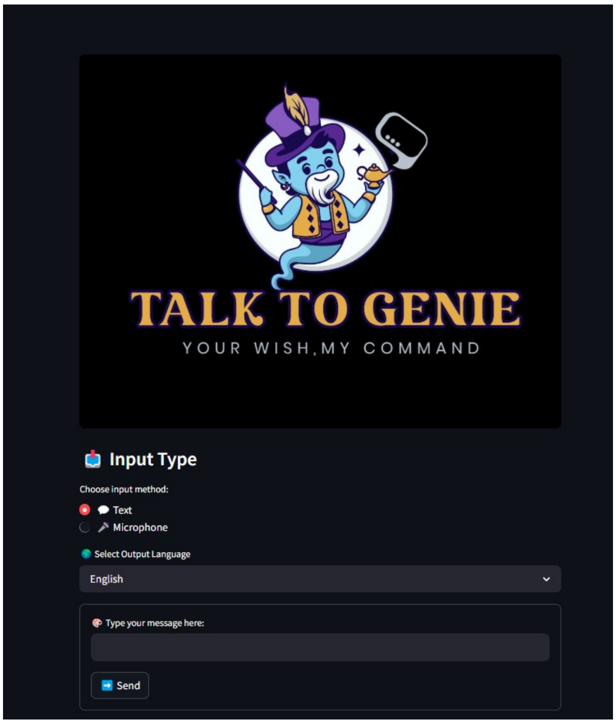

Figure 5.1: TalkGenie Interface

Figure 5.6 displays the chat history feature, allowing users to review past interactions with the AI assistant. This is crucial for maintaining context and revisiting previous queries and responses.

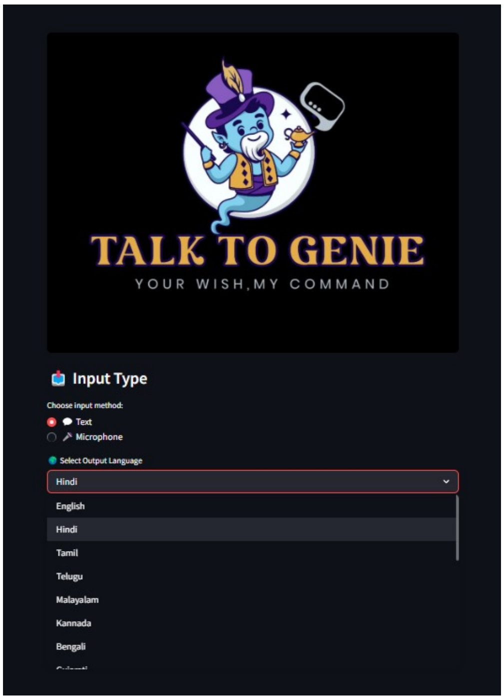

Figure 5.2: Output Language Selection

The multilingual capabilities of TalkGenie are demonstrated in Figure 5.2. This highlights its ability to generate responses in different languages.

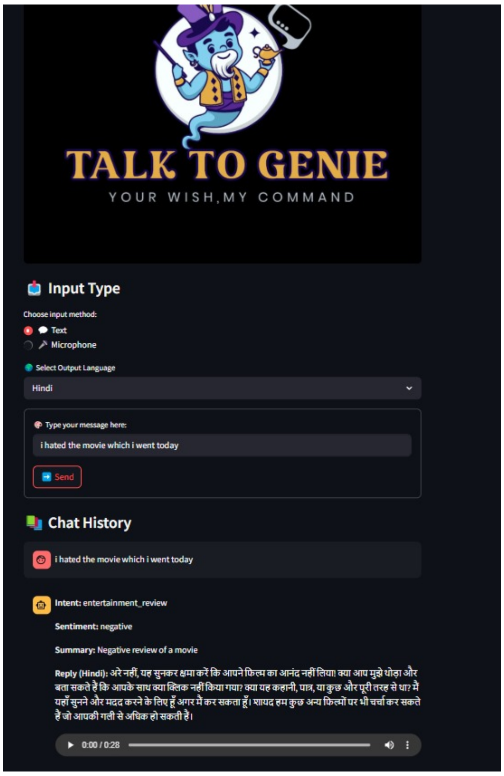

Figure 5.3: Output in Hindi

Similarly, Figure 5.5 illustrates the system's output in telugu, Figure 5.4

illustrates the system's output in tamil, and Figure 5.3 illustrates the system's output in hindi further emphasizing its multilingual support and adaptability to various linguistic inputs and outputs.

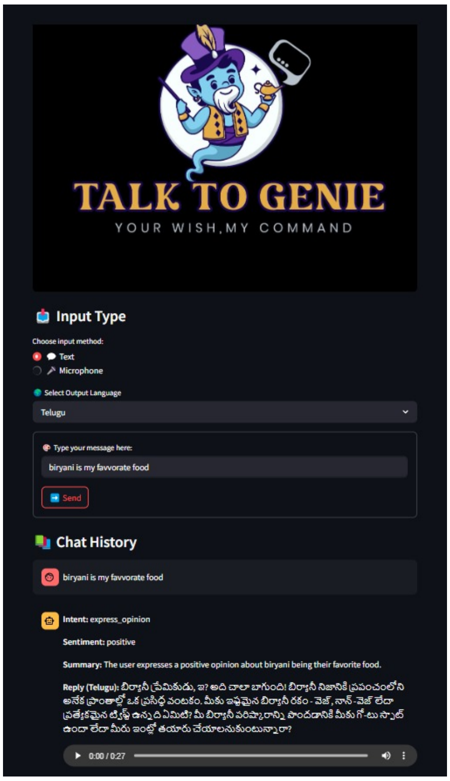

Figure 5.4: Output in Tamil

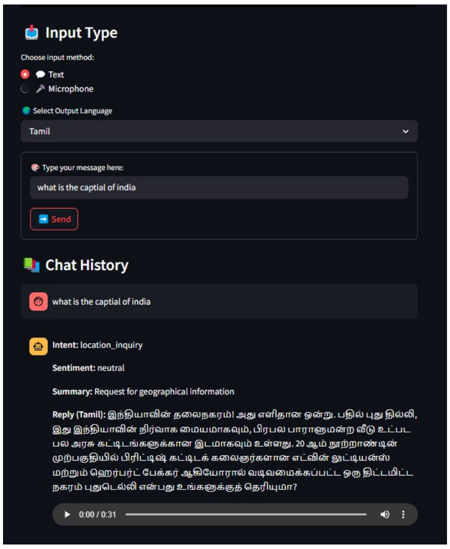

Figure 5.5: Output in Telugu

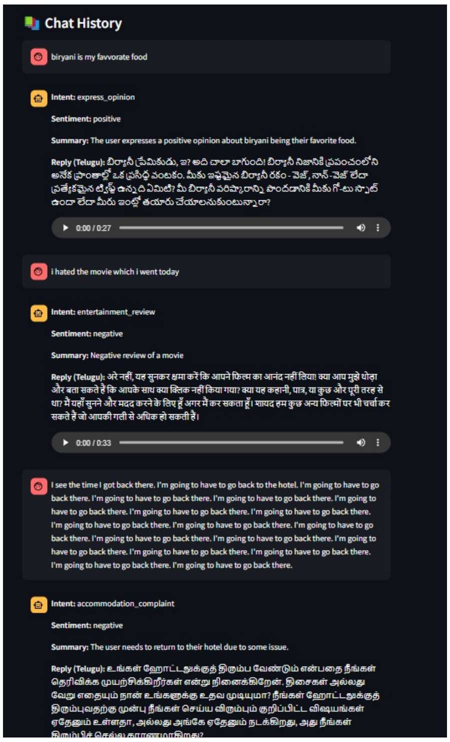

Figure 5.6: Chat History

5.3 Overview of Results

TalkGenie, a chatbot, has been evaluated for processing real-time inputs and providing accurate, context-sensitive responses based on user queries. Whisper is utilized to transcribe audio inputs [3], while text inputs are processed as such. The LLaMA3 model detects intent and sentiment, enabling the system to determine the user's intent and emotional tone. The model attained intent detection accuracy of 73.90% and sentiment analysis accuracy of 99% on an 80,000-example dataset. After the detection of intent and sentiment, the system produces a suitable response, which is rendered into the user's selected language through Whisper [3]and Deep Translator, and transformed into audio speech using gTTS. The conversation is stored temporarily using LangSmith to provide continuity in conversation. These findings show the potential of TalkGenie in providing multilingual, emotionally intelligent, and user-focused AI interactions, which makes it applicable to customer service, education, healthcare, and accessibility applications.

5.4 Analysis and Interpretation of Confusion Matrices

The Intent Classification Confusion Matrix and Sentiment Classification Confusion Matrix are two key metrics for measuring the performance of a model. The Intent Classification Confusion Matrix (a partial view of the top 10 intents is shown in Table 5.1 due to its size) and Figure 5.7 is largely diagonal, reflecting high accuracy, with the majority of predicted values being equal to the actual labels. It reflects a high accuracy rate, with most of the top 10 classes showing 27 correct predictions out of 27 samples each. This indicates that the model is very accurate and well-tuned for these intents, suggesting good feature representation and distinction between them. Further analysis on the full matrix could identify potential areas of confusion between less frequent or more similar intents.

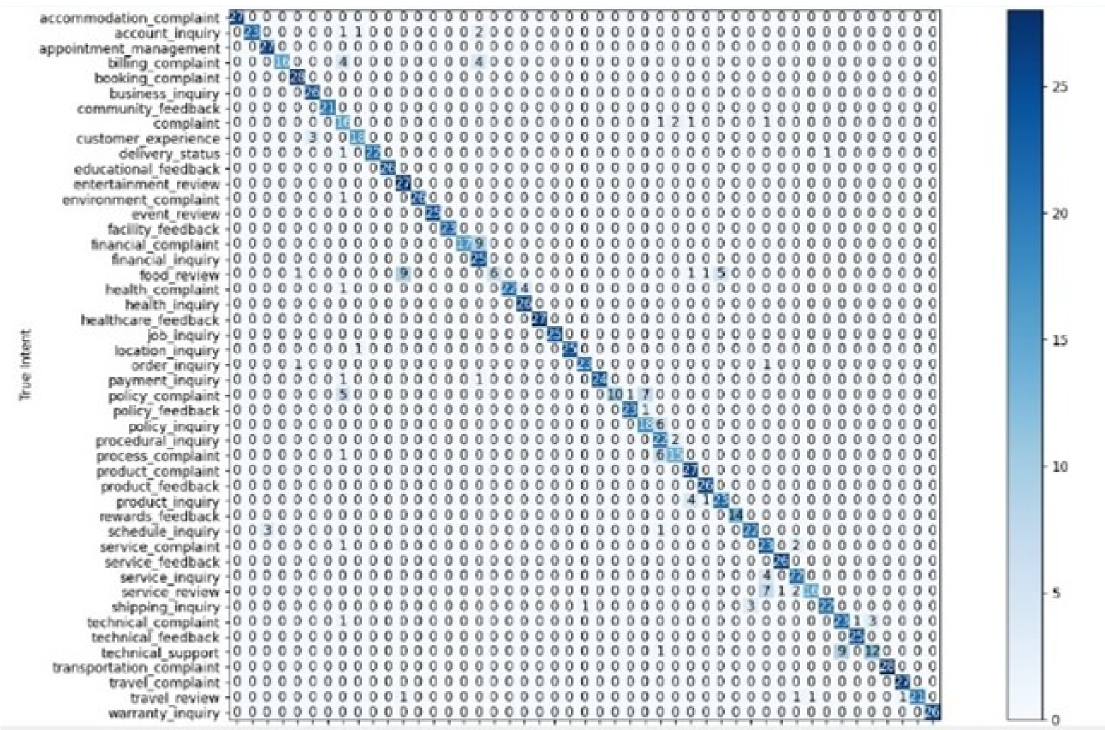

Figure 5.7: Intent Matrix

Table 5.1: Intent Confusion Matrix (Partial View - Top 10 Intents)

	accommodation_complaint	account_inquiry	appointment_management	billing_complaint	booking_complaint	business_inquiry	community_feedback	complaint	customer_experience	delivery_status
accommodation_complaint	27	1	0	0	0	0	0	0	0	0
account_inquiry	0	27	0	0	0	0	0	0	0	0
appointment_management	0	0	27	0	0	0	0	0	0	0
billing_complaint	0	0	0	27	0	0	0	0	0	0
booking_complaint	0	0	0	0	27	0	0	0	0	0
business_inquiry	0	0	0	0	1	27	0	0	0	0
community_feedback	0	0	0	0	0	0	27	0	0	0
complaint	0	0	0	0	0	0	0	27	0	0
customer_experience	0	0	0	0	0	0	0	0	27	0
delivery_status	0	0	0	0	0	0	0	0	0	27

Table 5.2 illustrates the Sentiment Classification Confusion Matrix, and Figure 5.8 assesses the model's performance in recognizing negative, neutral, and positive sentiment categories as depicted in Figure 5.9. It indicates strong performance, with most samples being classified correctly and minimal confusion between neutral and other classes. The low number of misclassifications suggests the model is robust and potentially deployable.

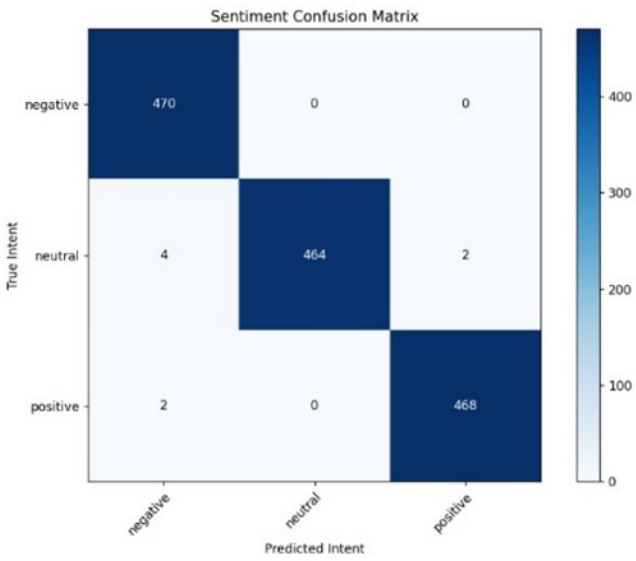

Figure 5.8: Sentiment Matrix

Table 5.2: Sentiment Confusion Matrix

	Predicted Negative	Predicted Neutral	Predicted Positive
Actual Negative	470	0	0
Actual Neutral	4	464	2
Actual Positive	2	0	468

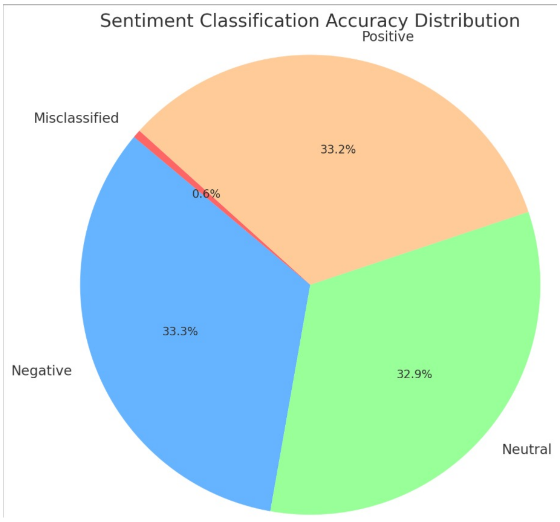

Figure 5.9: Distribution of Sentiment

5.5 Evaluation of Quality Factors in Line with Sustainability Goals

The intent classification model, driven by LLaMA3, gives an overall accuracy of 73.90% (as detailed in Figure 5.11) with a total support of 1410 samples across multiple intent classes. It performs well in classes with rich data, including accommodation_complaint, appointment_management, environment_complaint, and travel_complaint. In contrast, intents that have no training samples or sparse training samples obtain zero F1-scores, which lowers the macro average. This implies that LLaMA3 may gain from balanced data augmentation or fine-tuning underrepresented intents. The sentiment analysis model has an accuracy of 99.43% (as shown in Figure 5.10), and the total support for negative, neutral, and positive sentiments is 1410 samples. Each class of sentiment has near-perfect precision, recall, and F1-scores.

The intent classification model, driven by LLaMA3, gives an overall accuracy of 73.90% (as detailed in Figure 5.11). Accuracy, defined as the ratio of correctly predicted instances to the total number of instances, is calculated using the formula in Equation (5.1):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

The model's performance is further evaluated using the F1-score, a balanced measure derived from precision and recall, as shown in Equation (5.2):

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.2}$$

The reported 73.90% accuracy and the zero F1-scores for underrepresented intents contribute to the macro average, which is the unweighted mean of the F1-scores across all intent classes, as formulated in Equation (5.3) (assuming we were discussing macro F1-score in the broader context):

$$\text{Macro F1-Score} = \frac{1}{N} \sum_{i=1}^{N} \text{F1-Score}_i \tag{5.3}$$

The sentiment analysis model achieves a high accuracy of 99.43% (as shown in Figure 5.10), calculated using the same formula as intent accuracy (Equation (5.1)). The near-perfect precision, which, as per Equation (5.4), measures the proportion of correctly predicted positive instances out of all instances predicted as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.4}$$

and recall, which, as shown in Equation (5.5), measures the proportion of correctly predicted positive instances out of all actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5.5}$$

contribute to the excellent F1-scores observed for each sentiment class.

```
=== Sentiment Evaluation ===
Sentiment Accuracy: 99.43%

Classification Report (Sentiment):
              precision    recall  f1-score   support

    negative       0.99      1.00      0.99       470
     neutral       1.00      0.99      0.99       470
    positive       1.00      1.00      1.00       470

    accuracy                           0.99      1410
   macro avg       0.99      0.99      0.99      1410
weighted avg       0.99      0.99      0.99      1410


=== F1 Scores by Averaging Method ===
Intent Macro F1: 0.66
Sentiment Macro F1: 0.99

Intent Micro F1: 0.74
Sentiment Micro F1: 0.99

Intent Weighted F1: 0.80
Sentiment Weighted F1: 0.99
```

Figure 5.10: Sentiment Accuracy of TalkGenie

```
=== Intent Evaluation ===
Intent Accuracy: 73.90%

Classification Report (Intent):
C:\Users\Rahul Dara\AppData\Local\Programs\Python\Python311\Lib\site-pa
ill-defined and being set to 0.0 in labels with no true samples. Use `z
  _warn_prf(average, modifier, msg_start, len(result))
C:\Users\Rahul Dara\AppData\Local\Programs\Python\Python311\Lib\site-pa
ill-defined and being set to 0.0 in labels with no true samples. Use `z
  _warn_prf(average, modifier, msg_start, len(result))
C:\Users\Rahul Dara\AppData\Local\Programs\Python\Python311\Lib\site-pa
ill-defined and being set to 0.0 in labels with no true samples. Use `z
  _warn_prf(average, modifier, msg_start, len(result))
                           precision    recall  f1-score   support

 accommodation_complaint       1.00      0.90      0.95        30
         account_inquiry       1.00      0.77      0.87        30
  appointment_management       0.90      0.90      0.90        30
       billing_complaint       1.00      0.53      0.70        30
       booking_complaint       0.93      0.93      0.93        30
        business_inquiry       0.90      0.87      0.88        30
      community_feedback       1.00      0.70      0.82        30
               complaint       0.48      0.53      0.51        30
    complaint_management       0.00      0.00      0.00         0
     customer_experience       0.90      0.60      0.72        30
         delivery_status       1.00      0.73      0.85        30
    educational_feedback       1.00      0.87      0.93        30
    entertainment_review       0.73      0.90      0.81        30
   environment_complaint       1.00      0.87      0.93        30
            event_review       1.00      0.83      0.91        30
    express_appreciation       0.00      0.00      0.00         0
       express_gratitude       0.00      0.00      0.00         0
         express_opinion       0.00      0.00      0.00         0
          express_praise       0.00      0.00      0.00         0
    express_satisfaction       0.00      0.00      0.00         0
     facilities_feedback       0.00      0.00      0.00         0

 accommodation_complaint       1.00      0.90      0.95        30
  appointment_management       0.90      0.90      0.90        30
   environment_complaint       1.00      0.87      0.93        30
    express_satisfaction       0.00      0.00      0.00         0
         payment_inquiry       1.00      0.80      0.89        30
        policy_complaint       1.00      0.33      0.50        30
         policy_feedback       0.96      0.77      0.85        30
          policy_inquiry       0.69      0.60      0.64        30
      procedural_inquiry       0.59      0.73      0.66        30
       process_complaint       0.79      0.50      0.61        30
         process_inquiry       0.00      0.00      0.00         0
       product_complaint       0.82      0.90      0.86        30
        product_feedback       0.93      0.87      0.90        30
         product_inquiry       0.82      0.77      0.79        30
                  review       0.00      0.00      0.00         0
        rewards_feedback       1.00      0.47      0.64        30
        schedule_inquiry       0.88      0.73      0.80        30
       service_complaint       0.64      0.77      0.70        30
        service_feedback       0.96      0.87      0.91        30
         service_inquiry       0.81      0.73      0.77        30
          service_review       0.94      0.53      0.68        30
        shipping_inquiry       0.96      0.73      0.83        30
     technical_complaint       0.72      0.77      0.74        30
      technical_feedback       0.96      0.83      0.89        30
       technical_support       0.80      0.40      0.53        30
transportation_complaint       1.00      0.93      0.97        30
        travel_complaint       0.96      0.90      0.93        30
           travel_review       1.00      0.70      0.82        30
        warranty_inquiry       1.00      0.87      0.93        30

                accuracy                           0.74      1410
               macro avg       0.75      0.61      0.66      1410
            weighted avg       0.90      0.74      0.80      1410
```

Figure 5.11: Intent Accuracy of TalkGenie

TalkGenie, with its ability to process both text and speech through LLaMA3 for integrated intent and sentiment analysis, and its multilingual response generation using Deep Translator and gTTS, presents avenues for contributing to several UN Sustainable Development Goals.

SDG 3: Good Health and Well-being

By enabling communication about health-related inquiries in multiple languages and providing accessible auditory feedback, TalkGenie can empower individuals to understand health information and navigate healthcare systems more effectively, thus promoting good health and well-being [38].

SDG 4: Quality Education

TalkGenie's capacity to process educational feedback and potentially deliver learning resources across languages can enhance the quality and accessibility of education, breaking down language barriers and offering more inclusive learning opportunities [39].

SDG 9: Industry, Innovation, and Infrastructure

As a project that integrates advanced AI models and user-friendly interfaces (via Streamlit), TalkGenie includes technological innovation in human-computer interaction, contributing to the advancement of infrastructure and improving innovation in AI-driven communication technologies [40].

5.6 Summary

The TalkGenie system efficiently handles both text and voice inputs to recognize user intent, sentiment. With the LLaMA3 model fine-tuned on a data set of 80,000 labeled examples [9], the system performed with an accuracy of 73.90% in intent detection and a high accuracy of 99% in sentiment analysis. The model manages multilingual output through Whisper[3], Deep Translator, and gTTS [7], and also stores chat history temporarily through LangSmith [10].

CHAPTER 6

Conclusions and Future Scope

6.1 Conclusions

The TalkGenie chatbot initiative is an innovative one that seeks to overcome the communication gap between linguistic and cultural disparities in sectors such as education, healthcare, tourism, and customer service. It offers a clever, adaptive, and multilingual chatbot interface capable of perceiving human emotions and intentions, handling both text and audio inputs, and returning intelligent responses in the language preferred by the user.

TalkGenie excels at processing both English text and speech (via Whisper translation) using the integrated capabilities of the LLaMA3 large language model for simultaneous intent recognition and sentiment detection, eliminating the need for separate modules or manual labeling [9]. A key innovation is its multilingual response generation: after identifying intent and sentiment, Deep Translator ensures the response is meaningful and culturally appropriate in the target language, which is then converted to audio using gTTS for auditory feedback. The user-friendly Streamlit frontend allows easy text or voice input and real-time output (text and audio) [33], while LangSmith enhances the experience by temporarily storing chat history for conversational context[10].

The frontend of TalkGenie is built using Streamlit, a web application framework for Python that is highly responsive and simple to use [33]. This enables users to easily enter text or voice input, see the output response, and listen to it in real time. LangSmith's use stores chat history temporarily. The system was tested stringently to confirm its accuracy and efficiency, with intent detection accuracy at 73.90% and sentiment detection accuracy at 99%. Intent detection has scope for improvement with the extensive range of variability in user expressions and 50 different intent classes, whereas sentiment detection performance is close to perfection, reflecting the ability of the model in emotional understanding.

TalkGenie's reach is across multiple real-world uses, including healthcare, education, customer service, and tourism and immigration. By facilitating individuals from other regions to communicate without language constraints, TalkGenie contributes to equality and access to information and services.

TalkGenie is not only a chatbot but a conversational AI assistant that is intelligent, understanding not only what is said but also how and why. Based on strong models such as LLaMA3, with the aid of tools such as Whisper [3], Deep Translator, gTTS [7], Streamlit [33], and LangSmith, TalkGenie is a major leap in human-computer interaction, paving the way for more empathetic, intelligent, and accessible communication tools in the future

6.2 Future Scope of Work

TalkGenie, a multilingual chatbot powered by AI, is a very promising instrument in many industries. Scalability and flexibility make it a useful communication instrument. Emotional detection based on voice analysis, memory for conversations, offline capability, integration with a mobile app, handling multi-language input, and analysis of file and image input are its future improvements.Emotional intelligence can be augmented by adding voice signal processing and machine learning algorithms to recognize subtle emotional states such as stress, sarcasm, excitement, or confusion. Memory for dialog will enable the system to recall previous conversations with users, comprehend context across time, and customize responses accordingly based on previous conversations. Offline mode will make TalkGenie even more flexible and usable in remote or low-bandwidth areas. A mobile app integration will guarantee wider reach and user adoption, enabling users to interact with the assistant on-the-go. Multilingual input processing will be built to accept user queries directly in different native languages, enhancing accessibility for non-English speakers and eliminating communication barriers.Lastly, file and image input analysis will broaden TalkGenie's use-case scenarios, enabling users to upload documents for legal summarization or objects/text recognition, enabled by AI-based tools.

REFERENCES

[1] Joseph O'Connor and Ian McDermott. *NLP*. Thorsons London, UK: 2001.

[2] Henry Weld, Xiaoqi Huang, Siqu Long, Josiah Poon, and Soyeon Caren Han. "A survey of joint intent detection and slot filling models in natural language understanding". In: *ACM Computing Surveys* 55.8 (2022), pp. 1–38.

[3] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. "Robust speech recognition via large-scale weak supervision". In: *International conference on machine learning*. PMLR. 2023, pp. 28492–28518.

[4] Steven Loria. *TextBlob: Simplified Text Processing*. `https://textblob.readthedocs.io/en/dev/`. 2018.

[5] Steven Bird. "NLTK: the natural language toolkit". In: *Proceedings of the COLING/ACL 2006 interactive presentation sessions*. 2006, pp. 69–72.

[6] Maarten Grootendorst. *KeyBERT: Minimal keyword extraction with BERT*. `https://github.com/MaartenGr/KeyBERT`. 2020.

[7] Sam Virtue and Antonio Vidal-Puig. "GTTs and ITTs in mice: simple tests, complex answers". In: *Nature metabolism* 3.7 (2021), pp. 883–886.

[8] DeepL Team. *DeepTranslator: Neural Machine Translation*. `https://www.deepl.com/translator`. 2025.

[9] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. "The llama 3 herd of models". In: *arXiv preprint arXiv:2407.21783* (2024).

[10] Takumi Ito, Tatsuki Kuribayashi, Masatoshi Hidaka, Jun Suzuki, and Kentaro Inui. "Langsmith: An interactive academic text revision system". In: *arXiv preprint arXiv:2010.04332* (2020).

[11] Libo Qin, Tianbao Xie, Wanxiang Che, and Ting Liu. "A survey on spoken language understanding: Recent advances and new frontiers". In: *arXiv preprint arXiv:2103.03095* (2021).

[12] Norezmi Jamal, Shahnoor Shanta, Farhanahani Mahmud, and MNAH Sha'abani. "Automatic speech recognition (ASR) based approach for speech therapy of aphasic patients: A review". In: *AIP Conference Proceedings*. Vol. 1883. 1. AIP Publishing. 2017.

[13] Haojing Huang, Peijie Huang, Zhanbiao Zhu, Jia Li, and Piyuan Lin. "CLID: A chunk-level intent detection framework for multiple intent spoken language understanding". In: *IEEE Signal Processing Letters* 29 (2022), pp. 2123–2127.

[14] Feifan Song, Lianzhe Huang, and Houfeng Wang. "A unified framework for multi-intent spoken language understanding with prompting". In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2024, pp. 9966–9970.

[15] Akshat Shrivastava, Suyoun Kim, Paden Tomasello, Ali Elkahky, Daniel Lazar, Trang Le, Shan Jiang, Duc Le, Aleksandr Livshits, and Ahmed Aly. "ICASSP 2023 Spoken Language Understanding Grand Challenge". In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–2.

[16] Iñigo Casanueva, Tadas Temčinas, Daniela Gerz, Matthew Henderson, and Ivan Vulić. "Efficient intent detection with dual sentence encoders". In: *arXiv preprint arXiv:2003.04807* (2020).

[17] Jiajie Mei, Shiliang Zhang, Yujun Liu, Jiajun Chen, and Zhiqiang Liu. "Incorporating BERT With Probability-Aware Gate for Spoken Language Understanding". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31 (2023), pp. 164–176.

[18] Thai Binh Nguyen. "Improving spoken language understanding by enhancing text representation". In: *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2022, pp. 7502–7506.

[19] Adil Rajput. "Natural language processing, sentiment analysis, and clinical analytics". In: *Innovation in health informatics*. Elsevier, 2020, pp. 79–97.

[20] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. "A survey on sentiment analysis methods, applications, and challenges". In: *Artificial Intelligence Review* 55.7 (2022), pp. 5731–5780.

[21] Sonia Vakayil, D Sujitha Juliet, Sunil Vakayil, et al. "RAG-Based LLM Chatbot Using Llama-2". In: *2024 7th International Conference on Devices, Circuits and Systems (ICDCS)*. IEEE. 2024, pp. 1–5.

[22] Eesung Kim, Yun Tang, Taeyeon Ki, Divya Neelagiri, and Vijendra Raj Apsingek. "Joint End-to-End Spoken Language Understanding and Automatic Speech Recognition Training Based on Unified Speech-to-Text Pre-Training". In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2024, pp. 10971–10975.

[23] Why Python. "Python". In: *Python releases for windows* 24 (2021).

[24] Nan Cao, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu. "Whisper: Tracing the spatiotemporal process of information diffusion in real time". In: *IEEE transactions on visualization and computer graphics* 18.12 (2012), pp. 2649–2658.

[25] Francisco S Marcondes, Adelino Gala, Renata Magalhães, Fernando Perez de Britto, Dalila Durães, and Paulo Novais. "Using Ollama". In: *Natural Language Analytics with Generative Large-Language Models: A Practical Approach with Ollama and Open-Source LLMs*. Springer, 2025, pp. 23–35.

[26] Wes McKinney et al. "pandas: a foundational Python library for data analysis and statistics". In: *Python for high performance and scientific computing* 14.9 (2011), pp. 1–9.

[27] Charles R Harris, K Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. "Array programming with NumPy". In: *Nature* 585.7825 (2020), pp. 357–362.

[28] James Robert. *PyDub: Manipulate audio with a simple and easy high-level interface*. `https://github.com/jiaaro/pydub`. 2018.

[29] Uberi. *SpeechRecognition: Library for performing speech recognition, with support for several engines and APIs*. `https://github.com/Uberi/speech_recognition`. 2015.

[30] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, et al. "Jupyter Notebooks – a publishing format for reproducible computational workflows". In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press. 2016, pp. 87–90.

[31] Microsoft. *Visual Studio Code*. `https://code.visualstudio.com/`. 2015.

[32] The Python Packaging Authority. *pip - The Python Package Installer*. `https://pip.pypa.io/`. 2008.

[33] M Rekha, S Aishwarya, A Aswini, and AL Sakthi. "PollVue: Public Opinion Lens Using Python and Streamlit". In: *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*. IEEE. 2024, pp. 859–864.

[34] Google Research. *Google Colaboratory*. `https://colab.research.google.com/`. 2018.

[35] Amazon Web Services. *Amazon EC2*. `https://aws.amazon.com/ec2/`. 2024.

[36] Microsoft Corporation. *Microsoft Azure Cloud Computing Services*. `https://azure.microsoft.com/`. 2024.

[37] Hugging Face. *Hugging Face: The AI Community Building the Future*. `https://huggingface.co/`. 2024.

[38] Jean-François Guégan, Gerardo Suzán, Séraphin Kati-Coulibaly, Didier Nkoko Bonpamgue, and Jean-Paul Moatti. "Sustainable Development Goal# 3,"health and well-being", and the need for more integrative thinking". In: *Veterinaria México OA* 5.2 (2018), pp. 0–0.

[39] Surattana Adipat and Rattanawadee Chotikapanich. "Sustainable development goal 4: an education goal to achieve equitable quality education". In: *Academic Journal of Interdisciplinary Studies* 11.6,174-183 (2022).

[40] Juana Du, Nadeem Akhtar, and Yulei Dou. *Towards 2030: sustainable development goal 9: industry, innovation and infrastructure. A communication perspective.* 2023.