IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Empowering Tamil-Speaking Farmers: A Tanglish Supported Audio-Based Agriculture Chatbot for Real- Time Query

[1]Shafrin Farzana U, [2]Mohamed Irzath S, [3]Muhsina Sultana M Y, [4]Soundarya D, [5]Nafeesa Sanad S

[1,2,4,5] Student, [3] Assistant Professor

[1]Department of Computer Science and Engineering,

[1]Aalim Muhammed Salegh College of Engineering, Chennai, India

Abstract: The lack of accessible, localized agricultural support poses a major challenge for Tamil-speaking farmers, especially in rural areas with limited literacy. This project focuses on the development and evaluation of an AI-powered audio-based chatbot that supports Tanglish (Tamil written in English), Tamil, and English. Leveraging advanced speech recognition (Whisper) and natural language processing (GPT-4), the system allows farmers to interact through voice and receive real-time responses to queries related to crop management, pest control, weather updates, and market prices. Our research evaluates the chatbot's accuracy, response time, and user satisfaction, aiming to improve accessibility, inclusivity, and decisionmaking in agriculture. The proposed solution bridges the digital divide and empowers farmers through multilingual, voice- enabled agricultural assistance.

Index Terms **-Tanglish Chatbot, AI in Agriculture, Voice-Based Agricultural Assistance, Tamil-Speaking Farmers**

I. INTRODUCTION

The digital transformation in agriculture is crucial to addressing challenges faced by farmers in accessing timely and accurate information. In India, especially among Tamil-speaking communities, a major barrier to adopting digital tools is the lack of systems that support regional languages and voice-based interaction. Traditional agricultural advisory platforms often rely on text-based interfaces and English language support, which limits accessibility for farmers with low literacy levels or limited technical skills. Moreover, many farmers communicate in Tanglish—Tamil language written in English script— creating a need for systems that can understand and respond in this hybrid language format.

To bridge this gap, this project proposes the development of an AI powered chatbot that facilitates agricultural assistance through audio based queries in Tanglish, Tamil, and English. The chatbot uses advanced speech recognition technology (Whisper by OpenAI) to transcribe spoken input and employs Natural Language Processing (NLP) with GPT-4 to understand user intent and generate relevant responses. This approach enables real-time support on key topics such as crop management, pest control, weather forecasts, and market prices, empowering farmers to make more informed decisions While the integration of AI in agriculture is not new, the use of multilingual and hybrid language voice input systems remains underexplored, particularly for rural and regional user bases. Voice-enabled systems not only improve accessibility but also reduce the digital literacy barrier, offering a more natural way for farmers to interact with technology. However, challenges such as speech recognition in noisy environments, understanding regional accents, and maintaining response accuracy must be addressed to ensure effectiveness.

This study aims to develop, implement, and evaluate a chatbot system that supports audio input in Tanglish, processes it using state-of-theart AI tools, and delivers real-time agricultural advice. The performance of the system will be assessed based on metrics such as speech-to-text accuracy, response time, and user satisfaction. The outcomes of this research will provide valuable insights into how AI can be adapted to support regional language users, and contribute to the broader goal of inclusive digital agriculture

In the broader context of smart agriculture and digital empowerment, this chatbot represents a step toward democratizing access to expert knowledge for all farmers, regardless of education level or language proficiency. Future expansions of the system may include features such as image-based disease detection, integration with IoT-based soil sensors, and offline functionality for regions with limited internet connectivity. By focusing on real-world usability and regional language support, the Tamil Agriculture Chatbot not only enhances productivity but also contributes to building a more inclusive and sustainable agricultural ecosystem

II. BACKGROUND

India's agricultural sector is home to millions of small-scale farmers who often face challenges in accessing timely and relevant agricultural information. Traditional sources of knowledge such as extension officers, printed manuals, and TV programs are often not personalized or immediately available when farmers face critical issues in the field. With the rise of smartphones and internet connectivity in rural areas, there is a growing opportunity to bridge this gap through digital tools. However, language barriers, limited digital literacy, and lack of localized content have hindered widespread adoption of technology among rural farmers—particularly those who speak regional languages such as Tamil. While there are several AI-based chatbot systems available globally, most are developed in English or major languages and require text-based interaction, making them less effective for non English speakers or users who are not comfortable with typing. Furthermore, many farmers prefer to speak in Tanglish (Tamil spoken but written using English alphabets), a hybrid linguistic style commonly used in informal digital communication. This presents a unique challenge for Natural Language Processing (NLP) systems, which must understand and interpret speech input in a non-standardized, codemixed language format.

Recent advancements in speech recognition and AI language models have enabled more natural and conversational human-computer interactions. Models like OpenAI's Whisper and GPT-4 have demonstrated strong capabilities in transcribing and understanding multilingual audio input. By integrating these technologies, it is possible to create a chatbot that supports voice-based queries in Tanglish, Tamil, and English—helping farmers receive instant answers about crop diseases, pest control, weather, and market trends. Such a system can dramatically improve decision-making and reduce reliance on third-party sources or local agents who may not always be available Despite the benefits, developing a multilingual, speech-driven chatbot system comes with its own set of challenges. Accents, background noise, and inconsistent pronunciation can impact speech-to-text accuracy. Tanglish, lacking formal grammar or structure, requires advanced NLP models that can interpret context and intent accurately. This study addresses these gaps by implementing a user-friendly chatbot that combines speech recognition, code-mixed language support, and real-time AI- generated responses. The project not only contributes to AI in agriculture but also offers a scalable framework for inclusive and accessible technology solutions in other regional and low resource language communities.

In addition to addressing language and accessibility challenges, this chatbot also aligns with national and global efforts toward smart agriculture and digital empowerment. Government initiatives such as Digital India and eNAM (National Agriculture Market) emphasize the use of technology to support farmers, but often fall short in regional language integration and user-friendliness. By offering a system tailored to the linguistic and cultural context of Tamil-speaking farmers, this project fills a critical gap in current agricultural tech solutions. Moreover, its adaptable architecture allows for future integration with other data-driven tools like IoT- based soil sensors or satellite imagery for precision farming, paving the way for a comprehensive digital support ecosystem for Indian agriculture.

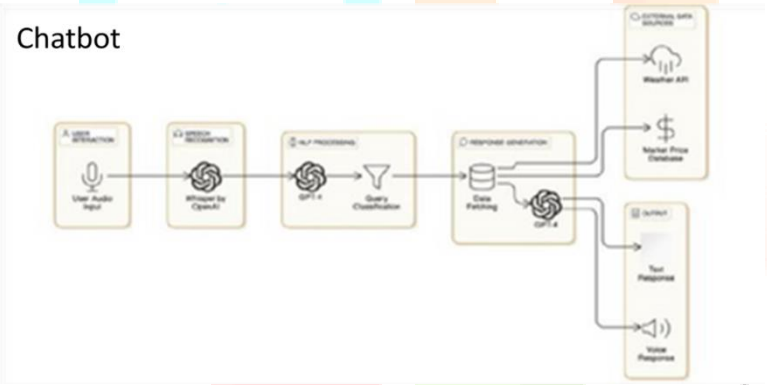
III. ARCHITECTURE

The Tamil Agriculture Chatbot is built to enable seamless, audio- based interaction between Tamil-speaking farmers and an AI-powered agricultural knowledge system. The architecture is centered around voice-first communication, allowing users to speak in Tanglish, Tamil, or English, which is then processed through a multi-layered pipeline of AI components. The system begins with a voice input module that records or captures audio queries from the user. These audio files are passed through the Whisper speech recognition engine, which transcribes the spoken content into text, even accounting for informal or hybrid language patterns like Tanglish

Once the audio is converted to text, the transcribed query is sent to the Natural Language Processing (NLP) layer, powered by OpenAI's GPT-4, which understands the user's intent, context, and tone. This NLP model is fine-tuned to handle agricultural vocabulary and informal phrasing common among rural users. After interpreting the query, the system generates a natural, conversational response in the same language format used by the user.

6 to 4 or Teredo, NAT64 is based on algorithmic mapping translation rather than encapsulation, allowing for direct communication over an IPv6-only transport network. This mechanism is particularly useful in datacenters, cloud environments, ISPs, and enterprise networks that are transitioning to IPv6 while maintaining access to legacy IPv4-based services.

Figure 1:System Architecture of the Tamil Agriculture



Key Components of the Proposed System

A.User Audio Input

The system begins with user interaction through voice commands .Farmers or users speak their queries in Tanglish—a blend of Tamil spoken using the English alphabet. This mode of communication makes it easier for users who are familiar with the Tamil language but not native English speakers. The input is captured using a microphone enabled mobile or web-based interface.

B.Speech Recognition(Whisper by OpenAI)

The captured audio is processed using Whisper, OpenAI's multilingual speech recognition model, which converts the Tanglish speech into corresponding text. Whisper is capable of handling accent variability and informal speech patterns, making it particularly effective for rural and regional dialects. This stage ensures accurate transcription, which is critical for downstream language understanding

C.NLP Processing and Query Classification

The transcribed text is passed through GPT-4, which serves as the Natural Language Processing (NLP) engine of the system. GPT-4 performs semantic analysis to understand the context and intent of the query. The system employs a classification mechanism to categorize the query into predefined types,

such as weather inquiries, crop suggestions,fertilizer usage, or current market prices. This classification aids in directing the query to the appropriate data retrieval module

D. Data Fetching from External Sources

Depending on the classified query type, the system fetches relevant data from external sources. Weather information is retrieved from a live weather API, while market price data is pulled from a regularly updated agricultural price database. These integrations allow the system to provide real-time, location-specific insights to farmers, enhancing the decision-making process in farming operations.

E.*Response Generation(GPT-4)*

After retrieving the required data, GPT-4 is again leveraged to format and contextualize the response. Instead of returning raw data, GPT-4 generates human-like answers that are conversational, clear, and customized to the user's query. This step ensures that users receive actionable responses tailored to their needs and literacy levels.

F.Output Generation

The final response is delivered through two output modes: a textual output displayed on the screen, and anaudioresponse synthesized using a Text-to-Speech (TTS) engine. The voice output is especially useful for users who may have reading difficulties or prefer listening. This multimodal output design increases the accessibility and inclusiveness of the system, making it suitable for use in low-literacy and rural environments.

.

IV. CONCLUSION

The development of AI-powered audio-based agricultural assistance systems represents a transformative step in bridging the digital divide in rural areas, particularly for non-native English speakers. By combining speech recognition, natural language processing, and realtime data retrieval, the proposed chatbot empowers farmers to access critical agricultural information in a user-friendly and intuitive manner. Our system, built on OpenAI's Whisper and GPT-4 models, demonstrates that Tanglish—a blend of Tamil and English—can effectively serve as a medium for voice-based interaction in regional contexts.

Despite its advantages, the proposed system also presents certain challenges. Accurate speech recognition in noisy rural environments remains a concern, as does handling code-mixed language input that varies by region and speaker. Furthermore, the reliance on real-time APIs means the system is dependent on stable internet connectivity and the availability of reliable external databases, which may not always be feasible in remote locations.

To ensure broad and sustainable adoption, strategic factors such as offline fallback mechanisms, dialect-specific model fine-tuning, and regular updates to agricultural databases must be prioritized. Additionally, while the system currently supports two output modes— text and voice—expanding to support multimedia responses (e.g., images or videos) could further enhance user engagement and comprehension

Ultimately, this chatbot architecture serves as a scalable, low-barrier technological solution for precision farming and agricultural decision making. It holds the potential to become a key enabler in digital agriculture, promoting informed practices and improving livelihood outcomes among farming communities in linguistically diverse regions.

V. REFERENCE

[1] A. Radford, J. Witte, R. Tezak, and M. Kim, "Whisper: OpenAI'sautomatic speech recognition system," OpenAI, 2022. [Online]. Available: https://openai.com/research/whisper

[2] OpenAI, "GPT-4 Technical Report," OpenAI, 2023. [Online].
Available: https://openai.com/research/gpt-4

[3] S. Kaur and R. Singh, "AI-Based Voice Assistants for Farmers: AReview," International Journal of Computer Applications, vol. 183, no.
45, pp. 10–15, 2022.

[4] M. Sharma, A. Patel, and K. Verma, "Development of a voice enabled chatbot for agricultural queries using NLP," in 2021
International Conference on Smart Innovations in Design, Environment,
Management, Planning and Computing (ICSIDEMPC), pp. 72–78, IEEE, 2021

[5] Government of India, "AgriMarket: Real-Time Market Price
Information," Ministry of Agriculture & Farmers Welfare, [Online].
Available: https://agmarknet.gov.in/

[6] A. S. Nair and B. Pillai, "Speech-to-text systems in Indian languages: Challenges and recent developments," in 2019 11th
International Conference on Communication Systems & Networks
(COMSNETS), pp. 531–535, IEEE, 2019