## AI-POWERED PLANT DISEASE DETECTION

MINOR PROJECT REPORT

By

**PVSK ROHIT (RA2311027010191)**

**RAHUL GOVARDHAN (RA2311027010201)**

Under the guidance of

#### Dr. M. Anand

*In partial fulfilment for the Course*

of

**21CSC206T – ARTIFICIAL INTELLIGENCE**

In

Dept. of Data Science and Business Systems



**FACULTY OF ENGINEERING AND TECHNOLOGY SCHOOL OF COMPUTING**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR**

**MAY 2025**

# SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

**(Under Section 3 of UGC Act, 1956)**

## BONAFIDE CERTIFICATE

Certified that this minor project report for the course **21CSC206T ARTIFICIAL INTELLIGENCE** entitled in "AI VOICE ASSISTANT" is the bonafide work of

**PVSK ROHIT (RA2311027010191), RAHUL GOVARDHAN (RA2311027010201)** who carried out the work under my supervision.

### Signature Signature

Dr. M. Anand Assistant Professor

Data Science and Business Systems SRMIST

Kattankulathur

Dr. V. Kavitha HOD & Professor

Data Science and Business Systems SRMIST

Kattankulathur

## TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO** | **CONTENTS** | **PAGE NO** |
| **1** | **ABSTRACT** | **2** |
| **2** | **INTRODUCTION** | **3-4** |
| **3** | **TECH STACK & WORK FLOW** | **5** |
| **4** | **DEMO** | **6** |
| **5** | **APPLICATIONS, SCALABILITY & COST** | **7** |
| **6** | **BENEFITS & CONCLUSION** | **8** |
| **7** | **REFERENCES** | **9** |

**CHAPTER 1 ABSTRACT**

In the modern era of digital transformation, conversational AI is reshaping the way humans interact with technology, especially in fields that demand real-time access to knowledge and support. Users across domains often face challenges in obtaining immediate, reliable assistance due to limited technical literacy, lack of hands-free solutions, or unavailability of domain experts. This leads to inefficiencies, delayed decision-making, and increased dependency on inconsistent online content. While textual AI tools and search engines are widely available, they often fall short in accessibility, ease of use, and natural interaction. This project aims to bridge this gap by developing an intelligent, voice-enabled assistant chatbot that offers real-time, context-aware responses using state-of-the-art speech recognition and generative AI. The system captures spoken input from the user, processes it via a large language model (Gemini), and delivers spoken replies through a text-to-speech engine, ensuring hands-free, inclusive, and user-friendly interaction. This solution emphasizes accessibility, natural communication, and intelligent guidance ultimately supporting more efficient, intuitive, andpersonalizedhuman-computerinteraction.

## CHAPTER 2 INTRODUCTION

Voice interaction is emerging as a transformative interface in the digital era, particularly as societies strive for more accessible, intuitive, and efficient ways to engage with intelligent systems. Despite the growing presence of chatbots and AI-based services, many users—especially those with limited digital literacy or physical accessibility constraints—face persistent challenges in retrieving timely and accurate information. A key issue is the lack of natural communication channels that allow users to interact without typing, navigating menus, or interpreting complex textual responses.

While the internet offers a vast amount of information, most existing systems are text-centric, lacking conversational fluidity, contextual understanding, and auditory feedback. As a result, users often feel disconnected, overwhelmed, or unable to use AI systems effectively. Moreover, current tools do not integrate voice interaction with powerful generative models or support personalized, human-like conversations in real-time.

With rising demand for inclusive and intelligent communication systems, there is a pressing need for a voice-enabled assistant that leverages speech recognition, text-to-speech synthesis, and large language models to deliver seamless, hands-free interaction. Such a system would empower users to access information, get personalized assistance, and interact with AI more naturally and efficiently.

#### Objectives

The main objective of this project is to design and develop an intelligent, voice-enabled chatbot system that:

* Understands and processes natural language queries through speech recognition,
* Provides accurate, real-time responses using generative AI (Gemini),
* Delivers spoken replies using a text-to-speech engine for hands-free interaction,
* Enhances accessibility for users with limited literacy or technical skills,
* Enables fluid, conversational interaction across various topics,
* Lays the foundation for integration into assistive technologies, smart devices, and customer support tools.

#### Functional Needs

**Voice-Based Interaction**

* Allow users to speak queries directly into a microphone.
* Use speech recognition (Google STT) to convert spoken input into accurate text.

**Natural Language Understanding & AI Response**

* Process transcribed queries using a generative AI model (Gemini) to generate intelligent, relevant responses.
* Understand diverse topics and respond conversationally with context-awareness.

**Spoken Response Delivery**

* Convert AI-generated responses into audible speech using a text-to-speech engine (pyttsx3).
* Support asynchronous playback for non-blocking, smooth voice interaction.

**User Accessibility and Control**

* Enable features like “stop speaking,” “repeat response,” or “restart session” for better usability.
* Make the system accessible for users with limited literacy or technical skills.

**Context-Aware Dialogue Support**

* Optionally retain conversation context for multi-turn interactions.
* Adapt responses based on prior input or user-specific settings (e.g., region, language).

**Integration and Extensibility**

* Support modular design for integration with mobile apps, web platforms, or smart devices.
* Prepared for multilingual support, offline mode, or task-specific extensions in future versions.

**CHAPTER 3**

**TECH STACK & WORK FLOW**

Voice Interface (Frontend Equivalent)  
The system operates through a simple, voice-driven interface that can be extended to a web or mobile application. Users interact with the assistant using a microphone, without the need for visual input or text typing. This design supports accessibility and ease of use across user groups.

Backend Development  
The backend is developed in Python and orchestrates the interaction between various modules:

* Speech Recognition Module: Captures and processes user voice input using the speech\_recognition library and Google Speech-to-Text API.
* AI Interaction Module: Sends the processed text to the Google Generative AI (Gemini) model via the google.generativeai library to obtain intelligent responses.
* Text-to-Speech Module: Delivers AI responses using pyttsx3, allowing asynchronous audio output for smooth voice communication.

AI Integration  
The Gemini AI model is used to generate relevant and context-aware replies. It supports a wide range of conversational topics, from general knowledge to specific tasks. The interaction flow includes:

* Accurate query interpretation and language understanding
* Response generation based on large-scale language modeling
* Adaptability for future customization and domain-specific dialogue

Deployment  
The system can be run locally or hosted on a server (e.g., using Flask or FastAPI) to provide a web-based or API-driven voice assistant. It is designed to be:

* Lightweight for edge device integration (e.g., Raspberry Pi)
* Scalable for integration into mobile apps, customer service platforms, or smart home environments
* Modular, allowing future enhancements such as multilingual support, offline capabilities, or integration with physical sensors

## CHAPTER 4 DEMO

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**CHAPTER 5 APPLICATIONS, SCALABILITY & COSTS**

**Real Time Applications**

**Applications:**

* General Users: Hands-free access to AI-powered knowledge, productivity tools, and personal assistance in everyday tasks.
* Visually Impaired Users: Enhanced accessibility through fully voice-operated interaction, reducing reliance on screen-based systems.
* Customer Support Systems: Automated voice responders for common queries in domains like banking, e-commerce, and healthcare.
* Educational Platforms: Interactive learning assistants that provide spoken answers, helping students learn via conversational engagement.
* Smart Devices and IoT: Integration into smart homes, appliances, or kiosks for intuitive control via voice commands.

**Scalability:**

* Language Expansion: Incorporate support for regional languages and dialects to reach broader user demographics.
* Platform Portability: Deploy on mobile apps, web browsers, desktop systems, or embedded in smart devices.
* User Base Growth: Scalable backend that supports high volumes of concurrent users via cloud infrastructure.
* Modular Integration: Ready for integration with other AI tools (e.g., vision, automation) or APIs in enterprise ecosystems.

**Costs:**

* Development: Initial investment in integrating Google Generative AI, STT, and TTS modules.
* Operational: Usage-based costs for Gemini API, cloud hosting for speech processing, and background services.
* Scalability: Infrastructure upgrades, multi-language model support, and offline deployment modules.
* Revenue Models: Freemium model, paid API access, licensing for business solutions, or embedded OEM partnerships.

## CHAPTER 6 BENEFITS & CONCLUSION

Benefits

Improved Accessibility: Enables seamless interaction through voice, making AI tools more accessible for users with visual or physical limitations.

Hands-Free Efficiency: Facilitates multitasking and reduces reliance on keyboards or screens, enhancing user productivity and convenience.

Cost-Effective: Eliminates the need for expensive hardware or software by using lightweight, open-source speech libraries and cloud APIs.

Natural User Experience: Delivers human-like, conversational responses using generative AI, creating an intuitive and personalized interaction.

Scalability: Modular architecture allows for easy expansion to support multiple languages, platforms, and advanced use cases such as IoT or smart assistants.

Conclusion

The proposed voice assistant chatbot system represents a significant step forward in natural and inclusive human-computer interaction. By combining speech recognition, generative AI, and text-to-speech technologies, the system delivers real-time, intelligent, and hands-free communication. This approach not only enhances accessibility but also empowers users to engage with complex AI capabilities through simple spoken language.

The system's flexibility allows it to be adapted for various applications, from education and healthcare to customer support and smart environments. Its scalable architecture ensures that it can evolve alongside user needs and technological advancements.

Ultimately, this voice-based chatbot platform sets the foundation for more human-centered digital experiences—where users interact with technology as naturally as they would with another person.

**CHAPTER 7 REFERENCES**

* Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.). Stanford University.  
  This foundational text provides in-depth coverage of natural language processing, speech recognition, and voice synthesis—all central to the chatbot’s design.
* Tanwar, S., Patel, S., & Tyagi, S. (2020). *Voice Assistants: Architecture, Opportunities, and Challenges.* IEEE Access, 8, 14120-14134.  
  Discusses voice assistant system architectures and real-world applications, aligning with this project’s focus on hands-free interaction.
* Google Cloud. (2023). *Generative AI with Gemini API – Developer Documentation.*  
  Official documentation for using Google Generative AI (Gemini) for intelligent, context-aware response generation in AI applications.
* Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). *A survey of the usage of deep learning in agriculture.* Computers and Electronics in Agriculture, 147, 70-90.  
  While agriculture-focused, this reference supports the project's potential applications in smart assistants for rural and agricultural domains.
* Povey, D., et al. (2011). *The Kaldi Speech Recognition Toolkit.* IEEE ASRU.  
  While not used directly, this toolkit and its principles relate to the speech recognition layer of the system and highlight advancements in open-source speech tools.
* Python Libraries:
  + speech\_recognition – Python Library for Speech Recognition.
  + pyttsx3 – Text-to-Speech conversion library in Python.
  + google.generativeai – Google’s SDK for accessing Generative AI (Gemini) AP