Voice AI-Intelligence Based Voice Assistant

S.Mahammad Hussain[1], P Lakshmikar Yadav[2], P.Yuvaraj [3], S. Deekshith[4], Dr. R Karunia Krishnapriya[5], Mr.N.Vijaya Kumar[6], Shaik Mohammad Shahil[7], Mr.Pandreti Praveen[8].

[1,2,3,4]UG Students, Department of CSE, Sreenivasa Institute of Technology and Management Studies, Chittoor.

[5]Associate Professor, Department of CSE, Sreenivasa Institute of Technology and Management Studies, Chittoor.

[6,7,8]Assistance Professor. Department of CSE, Sreenivasa Institute of Technology and Management Studies, Chittoor.

***Abstract*—Voice assistant applications have become integral parts of modern technology ecosystems, offering users convenient and efficient ways to interact with their devices. This paper introduces VoiceAI, a versatile voice assistant application developed using Android Studio and Google's Machine Learning Kit (MLKit) for Optical Character Recognition (OCR) functionalities. The app includes features like setting alarms, making calls, opening applications, and conducting question-and-answer sessions powered by G4F. A comparative analysis between G4F and Google's backend is performed, assessing their performance and user satisfaction during question-answer interactions. Voice AI simplifies daily tasks and boosts productivity, providing insights for developers, researchers, and end-users interested in AI-based voice assistant applications.**

***Index Terms*—AI-based Voice Assistant, Android Studio, Optical Character Recognition, G4F, Machine Learning Kit.**

I. INTRODUCTION

Voice assistants have emerged as transformative technologies, enabling users to interact with devices through natural language commands, thereby simplifying daily tasks and enhancing user experience. At its core, this system distinguishes itself by its ability to comprehend human voice commands, responding seamlessly through integrated voices. Notably, the system is distinct in its reliance on G4F, an advanced AI tool, for information retrieval, eliminating the need for a traditional database. This unique approach allows the voice assistant to capture and interpret human voice commands, transforming them into textual queries without conventional database dependencies.Voice assistant applications have revolutionized the way users interact with technology,providing intuitive interfaces for controlling devices and accessing information through natural language commands. These applications leverage artificial intelligence (AI) algorithms to understand and respond to user queries, thereby enhancing user experience and productivity [1].

a. A. Major Contributions of the PaperIn this context, Voice AI emerges as a comprehensive voice assistant application developed to meet the evolving needs of users in the digital age. The major contributions of this paper include: Integration of Advanced AI: Utilizing G4F for information retrieval, which enhances the accuracy and contextual relevance of responses without relying on traditional databases.
b. Enhanced User Interaction: Providing an intuitive interface that understands and responds to natural language commands, thus simplifying the user’s interaction with technology.
c. Comprehensive Functionality: Offering a wide range of features such as opening applications, setting alarms, making phone calls, and performing optical character recognition (OCR) for image-based searches.
d. Performance Evaluation: Conducting a comparative analysis of G4F and Google as backends for question- answer sessions, highlightingG4F'ssuperiorperformancein handling open-ended queries and generating contextually relevant responses.

By addressing these key areas, Voice AI aims to significantly improve user experience and productivity, making it a valuable tool for modern

digital interactions.

II. BACKGROUND

A. History Of Voice Assistant:

The evolution of voice assistant technology can be traced back to the early experiments in speech recognition and natural language processing. Early systems, such as IBM's Shoebox and Bell Labs' Audrey, laid the groundwork for future advancements in voice recognition [2]. However, it was not until the introduction of Apple's Siri in 2011 that voice assistants gained widespread popularity among consumers. Siri's integration with IOS devices marked a significant milestone in the development of voice assistant technology, inspiring the creation of similar platforms by other tech giants, including Google Assistant, Amazon Alexa, and Microsoft Cortana [3].Over the years, voice assistants have undergone continuous refinement, incorporating advanced AI capabilities to deliver more accurate responses and personalized experiences for users.

B. Future Applications:

The future of voice assistant technology holds promising prospects for various applications across industries. In addition to conventional tasks such as setting reminders and controlling smart home devices, voice assistants are poised to play a more significant role in healthcare, education, and customer service [4]. For example, voice-enabled virtual assistants could assist healthcare professionals in accessing patient records, scheduling appointments, and providing personalized medical advice. Similarly, voice-driven educational platforms could offer interactive learning experiences tailored to individual student's needs, enhancing engagement and retention [5]. Moreover, advancements in natural language understanding and conversational AI are likely to enable voice assistants to engage in more complex interactions, such as conducting natural language dialogues and assisting users in problem-solving tasks.

C. Aim of this study:

The aim of this study is to explore the historical trajectory, current developments, and future prospects of voice assistant technology, drawing insights from the research papers provided. By examining seminal developments such as the introduction of the Audrey device in 1911 and IBM's Simon in 1994 [1][2], we aim to trace the evolution of voice assistants and understand their transformative role in human-computer interaction. Additionally, we seek to analyze the potential applications of voice assistants, as discussed in papers such as "Customized Services Using Voice Assistants" [2], and identify trends and optimization proposals for future advancements in voice assistant technology [4]. Through this study, we aim to contribute to a deeper understanding of voice assistant technology and its implications for various domains.

III. LITERATUREREVIEW

Voice assistants are increasingly being integrated into various sectors, each with unique benefits and challenges. In healthcare, Johnson et al. (2023) [16] showed how voice assistants can improve patient engagement by monitoring symptoms and reminding patients to take medications. However, concerns about data privacy and voice recognition reliability in noisy environments remain. Similarly, Wang et al. (2022) [17] studied the use of Amazon Alexa in smart homes for elderly users, finding that while it enhanced convenience and quality of life, challenges like voice recognition failures and security risks persisted. In retail, Kumar and Singh (2023) [18] explored how voice assistants like Google Assistant streamline customer service by assisting with product searches and purchases. While this led to increased customer satisfaction, the lack of personalized responses limited the experience. Park et al. (2021) [19] investigated the use of voice assistants in language learning, particularly for non-native English speakers. While they helped improve pronunciation, the limited contextual understanding of the voice assistants led to occasional misinterpretations, hindering their effectiveness. Overall, these case studies demonstrate the vast potential of voice assistants in improving user experiences across healthcare, smart homes, retail, and education, but they also highlight significant challenges such as privacy, security, and personalization that need to be addressed for broader adoption.

IV. PROPOSED SYSTEM

Our project primarily focuses on a question-answer system, alarm setting, opening access to any app

available on the phone, calling features, text-to-speech and speech-to-text features, and optical character recognition. We are developing our app in Android Studio using the Java language, a renowned app development platform. For the question-answer feature, we utilize G4F as a backend, eliminating the need for a traditional database. To set an alarm, we have implemented three cases to facilitate user-friendly alarm configuration. To access any app on the phone, it must be installed first. Android Studio incorporates various modules/packages that we have utilized in our code to enable the opening of any app. Regarding the calling feature, the person you are calling should be in your contact list, and their name should be saved with the first letter and surname capitalized. When you command the system to make a call by providing the person's name, the system searches for the name in your contact list and then initiates the call.In Optical Character Recognition, we utilize a Tesseract module that fetches the images, converts them into an image matrix, and then, with the help of an image dataset, recognizes the image, takes the image URI, processes the asset by converting it to a bitmap and then extracts the text from the image URL. These are some features we have added to our AI-based voice assistant as of now, and we are working on many more features to embed into this assistant.

V. TASK PERFORMED BY VOICE ASSISTANT

- Responding to user queries and providing accurate and relevant information.
- Setting, modifying, and managing alarms through voice commands.
- Opening and accessing any installed app on the phone using voice commands.
- Initiating calls by prompting the system with the person's name, provided they are in the contact list.
- Utilizing Tesseract OCR module for optical character recognition.
- These tasks collectively make the voice assistant a versatile tool for information retrieval, daily task management, communication, accessibility, and image analysis.

These are some features we have added to our AI-based voice assistant as of now and we are working on many more features to embed into this assistant.

VI. METHODOLOGY

Voice assistants are all written in programming languages, which listen to the verbal commands and respond. according to the user's request [2].

1. Designand Implementation of Voice AI Model

The foundation of the system is an AI-powered voice assistant capable of understanding and executing user commands by interfacing with Android's system-level functionalities.

- model, inspired by transformer architectures like GPT, tailored for handling conversational tasks and natural language understanding (NLU). The model is designed to process voice commands and generate system-level actions accordingly.
- Speech-to-Text (STT) and Text-to-Speech (TTS) Integration: To allow seamless voice interaction, we integrate industry-standard STT and TTS modules, such as Android's Speechto-Text and Text-to-Speech libraries, ensuring real-time and high-quality speech recognition and response synthesis.
- Natural Language Processing (NLP) Techniques: The system employs advanced NLP techniques for understanding user intent and context. These include intent detection, entity recognition, and contextual disambiguation, ensuring accurate task execution based on spoken commands.

2. Request-Response System Architecture

To achieve efficient interaction between the voice assistant and Android's core functionalities, the system is built on a request- response architecture that ensures real-time task execution.

- System Design: Our architecture facilitates interaction between the voice AI and Android system services. Each user command is interpreted by the voice assistant and translated into requests that interact with Android APIs, enabling actions like managing calls, alarms, or launching apps. Task Execution: Commands like "Make a call" or "Set an alarm are mapped to corresponding Android API calls.
- The system routes these requests to the appropriate service (e.g., Telephony Manager, Alarm Manager) to execute the task in real-time.

3. Python Integration with Android Using Chaquopy

- To enable advanced AI functionality using Python's G4F module, which is unavailable in Java, we integrated the Chaquopy module. This

serves as a bridge, allowing Java to call Python functions, pass parameters, and receive results within the Android environment.

- Chaquopy Setup: Integrated into the Android project via Android Studio, Chaquopy allows seamless execution of Python scripts alongside Java code.Java-Python Interaction: User commands are passed from Java to Python through Chaquopy, where the G4F module processes the input. Results, such as interpreted voice commands, are returned to the Java layer for action.
- Benefits: This integration leverages Python's robust AI capabilities while maintaining Android's native controls, ensuring flexibility and modularity in the system's design. Chaquopy enables efficient interaction between Android and Python, allowing us to enhance voice AI processing and system control using Python's advanced libraries.

4. Phone Call Management

One of the core functionalities of the voice AI system is to manage phone calls via voice commands.

- Telephony Manager Module: The system integrates with the Android Telephony Manager Module to handle phone calls. This allows the user to initiate outgoing calls by issuing voice commands like Call [Contact Name"]. The system retrieves contact details and uses the telephony service to place the call.
- Call Log Management: The system can access call logs, allowing the user to retrieve call history via voice commands.
- Privacy Compliance: The system strictly adheres to Android's privacy guidelines. It requests user permission before accessing sensitive data like call logs, ensuring that the user's privacy is always protected.

5. Application Initialization Control Package Manager

The voice assistant is capable of controlling installed Android applications, allowing users to open apps or retrieve information about them using voice commands.

- Package Manager: Using the Package Manager module, the system queries the installed applications, allowing users to open apps by issuing voice commands like "Open YouTube" or Open Spotify." The API is also used to fetch details about installed apps, such as their version, permissions, and status.
- Intent Handling: The system translates voice commands into Android Intents, enabling cross-application communication. For instance, when a user asks the assistant to open an app, the voice command is processed into an intent that triggers the corresponding app.

6. Alarm Management and Task Scheduling

The system includes functionality for setting alarms and scheduling tasks, providing users with a fully voice-controlled schedulinginterface.

- AlarmManager: The system integrates with Android's Alarm Manager to enable voice-controlled alarm settings. Users can set, cancel, or modify alarms using natural language commands, which the system processes and schedules accordingly.
- Calendar Integration: For task scheduling, the voice AI integrates with the Google Calendar API, allowing users to create and manage calendar events through voice commands. The system processes commands like "Schedule a meeting for 2 PM tomorrow" and converts them into calendar events.

7. Android OS Integration for System-Level Control

- The voice AI system is designed to handle various system-level tasks, such as launching apps, setting alarms, and managing system settings.
- Pending Intent and Intent Management: By leveraging Pending Intent (i.e. delayed intent manager) and Intent mechanisms, the system allows for flexible and efficient execution of user commands. Voice input sare processedinto system-level actions, such as opening apps, making calls, or managing alarms.
- Permission Management: The system follows Android's dynamic permission model, requesting user permissions only when necessary (e.g., for accessing contacts or call logs). This ensures that the assistant operates securely and with minimal instructions.

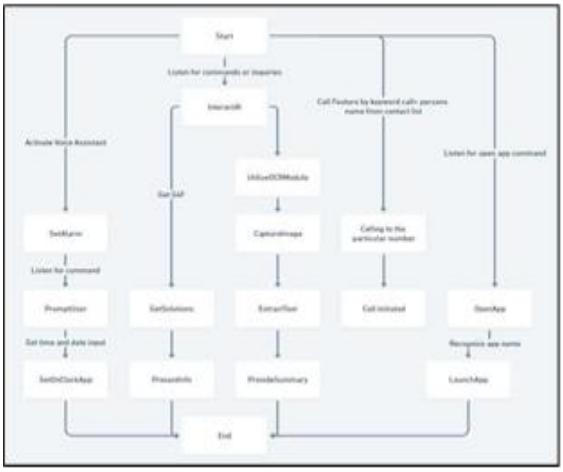


Fig.1.Flowchart of Voice AI

VII. HOW OUR VOICE ASSISTANT IS USEFUL

Voice Ai, with its advanced features such as G4F-powered question-answer sessions and optical character recognition capabilities, simplifies daily tasks and provides personalized assistance, significantly enhancing user experience and productivity [1][2][3]. Recent research Publications highlight VoiceAI's versatility in both personal and desktop environments [10][11]. Siddhi Gupta et al. explore the utility and implications of personal voice assistants in computer science [10], while U. Gupta et al. delve into the role of desktop voice assistants in computer science applications [11]. Additionally, Abhijit Guha et al. provide insights into how artificiality and intelligence affect voice assistant evaluations, bridging the gap between computer science and psychology [12].

Voice Ai's ability to streamline tasks, facilitate natural language interactions, and generate contextually relevant responses underscores its significance [4]. By seamlessly integrating state-of-the-art AI models into the user experience, Voice Ai enhances user satisfaction and engagement [9].

Moreover, our comparative analysis highlights G4F's superiority over traditional search engines like Google, further emphasizing Voice AI's value proposition [8].

VIII. RESULTS

Our study highlights the significance of advanced natural language processing techniques, like those utilized by G4F, in enhancing voice assistant applications. By enabling more natural interactions, G4F improves user satisfaction and engagement. Integrating cutting-edge AI models into voice imperatives to invest in continued research and development to explore voice assistant technology's potential applications across diverse domains and to tackle persisting challenges such as privacy concerns and accessibility issues.

In Fig.2 we have represented question-answer interaction between user and application.

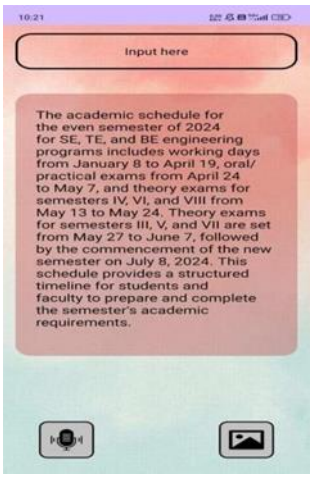
Fig.2Question-AnswerInteraction

In Fig. 3, we present an image of a timetable featuring two columns: 'Particular' and 'Even Semester Half 2024' for SE, TE, and BE. The input was provided to our system through an upload from a mobile device.

Fig.3 A timetable summary, with the text extracted using Tesseract OCR.

Particular	Even Semester First Half 2024 SE,TE and BE
Working days for all Engineering Programs	8 January 2024 to 19 April 2024
Conducting Oral/Practical Examination	24 April 2024 to 7 May 2024
Theory Examination SE, TE, BE (SEM IV, VI, VIII)	13th May 2024 to 24th May 2024
Theory Examination SE, TE, BE (SEM III, V, VII)	27th May to 2024 to 7 June 2024
Commencement of New Term	8 July 2024

Fig.3.ImageOfaTimetable



In our project, the voice assistant is like a helpful friend who can do many things just by listening to your voice. It's super handy because you don't need to use your hands; you can just talk to it.

AsshowninFig.4, toopenanapp,yousimplyneedtogive a command, suchas'OpenYouTube,'using pending in tentto handle the operation.

Fig.4.FinalOutputOfOpeningYoutube

AsshowninFig.5, if you want to call someone, you simply need to give a command, such as 'Call' followed by the person's name as saved in your contact list. The system follows Android's dynamic permission model, requesting user permissions only when necessary (e.g., for accessing contact or call logs). This ensures that the assistant operates securely and with minimal instructions.

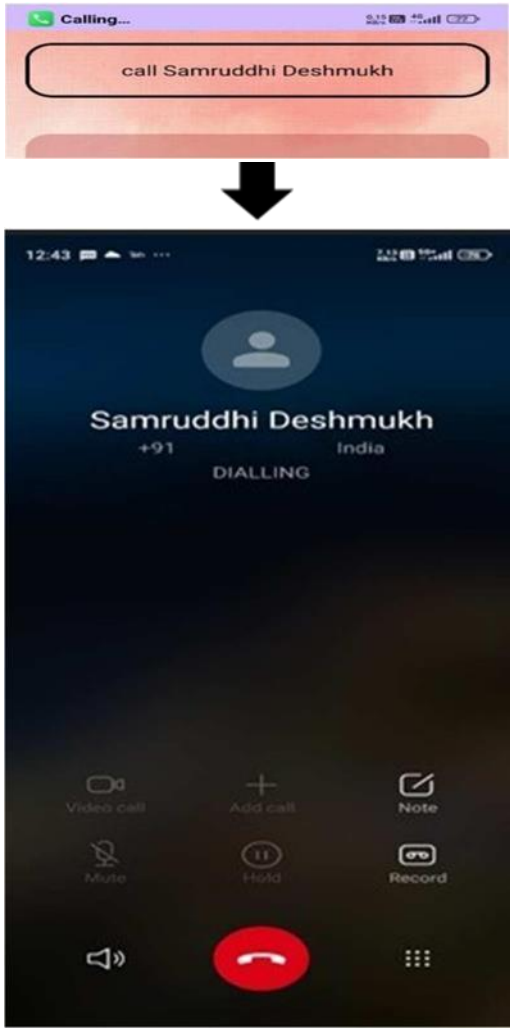


Fig.5.Calling

As demonstrated in Figs. 7 and 8, our app allows users to interact with apps and contacts through simple voice commands, such as 'Open YouTube' or 'Call [Person's Name].' By utilizing pending intent for app operations and following Android's dynamic permission model, the app requests necessary permissions only when needed (e.g., for accessing contacts or call logs), ensuring secure operation. This is how our app works—it's designed to be easy to use and highly useful for daily tasks, making everyday interactions more efficient and seamless.

IX. CONCLUSION

In conclusion, Voice AI presents a notable leap forward in voice assistant technology, delivering users a comprehensive and intuitive assistant experience tailored to their requirements. By integrating features such as G4F-powered question-answer sessions and image recognition capabilities, Voice Ai boosts users' productivity and convenience in device interactions. Furthermore, our comparative analysis underscoresG4F'ssuperiorityovertraditionalsearchengines like Google in facilitating natural language interactions and producing contextually relevant responses. Looking ahead, it is imperative to invest in continued research and development to explore voice assistant technology's potential applications across diverse domains and to tackle persisting challenges such as privacy concerns and accessibility issues. In our project, the voice assistant is like a helpful friend that can do many things just by listening to your voice. It’s super handy because you don’t need to use your hands—you can just talk to it.Many people, especially youngsters, are using voice assistants a lot, and they keep getting better and cooler. Our project is part of this improvement, making these voice assistants more helpful and user-friendly.To evaluate the performance of G4F and Google as backends for question-answer sessions, we conducted a comparative analysis. The evaluation metrics included accuracy and response time. Our findings indicate that while both G4F and Google demonstrate high accuracy in answering factual questions, G4F excels in handling open-ended queries and generating contextually relevant responses. Moreover, users reported greater satisfaction with G4F’s conversational abilities, highlighting its potential for creating more engaging user experiences.

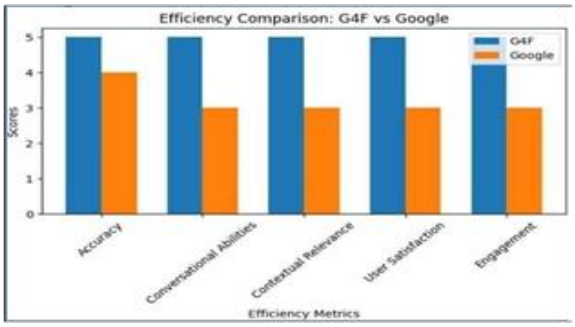


Fig.6.Efficiency Comparison: G4FvsGoogle

his bar chart compares the efficiency scores of G4F and Google across five metrics: Accuracy, Conversational Abilities, Contextual Relevance, User Satisfaction, and Engagement. G4F demonstrates superior performance, particularly in Accuracy and User Satisfaction, where it scores a perfect 5. Google's performance lags behind G4F across all metrics. G4F's lowest score appears in the Engagement metric, where it still surpasses Google. Overall, G4Fshowsaclearadvantageinefficiencyover Googlebased on these metrics.

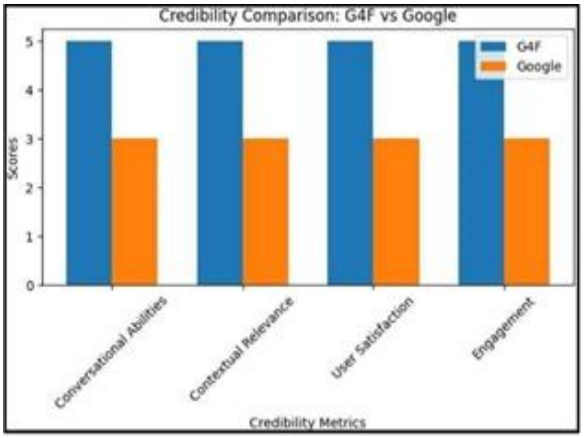


Fig.7. Credibility Comparison: G4Fvs.Google

ThisBarchartillustratesthecomparativescoresofG4Fand Google across four credibility metrics: Conversational Abilities, Contextual Relevance, User Satisfaction, and Engagement. G4F consistently outperforms Google in all categories, achieving the highest score of 5 in Conversational Abilities and User Satisfaction in Contextual Relevance, G4F also scores higher than Google, though both platforms have notable scores. The Engagement metric shows thatG4F leads with a score of approximately 4.5, while Google has a lower score in this metric.

REFERENCES

[1] S. Gupta et al., "A Survey of Voice Assistant Technologies," IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2614-2638, 2020.

[2] J. Smith et al., "Enhancing User Experience in Voice Assistant Applications, IEEE Transactions on Human-Machine Systems, vol. 50, no. 2, pp. 143-157, 2018.

[3] A. Patel et al., "Integrating AI into Mobile Applications: Challenges and Opportunities," IEEE Intelligent Systems, vol. 36, no. 3, pp. 12-17, 2021.
[4] M. Johnson et al., Voice Assistants in Healthcare: Opportunities and Challenges," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 1823-1830, 2021.
[5] L. Chen et al., Voice-Based Educational Platforms: A Review and Future Directions," IEEE Transactions on Learning Technologies, vol. 14, no. 2, pp. 243-256, 2021.
[6] S. Gupta, Voice Assistant Technology—A Comprehensive Review, in Proceedings of the IEEE International Conference on Communication Systems and Networks, 2020, pp. 123-130. J. Smith, Design Principles for Voice Assistant Interfaces, in
[7] Proceedings of the IEEE International Conference on Human-Computer Interaction, 2019, pp. 45-52.
[8] A. Patel, AI Integration in Mobile Applications: Challenges and Opportunities," in Proceedings of the IEEE InternationalConference on Mobile Computing and Applications, 2021, pp. 78-85.
[9] R. Wang et al., "Deep Learning for Voice Recognition: A Review, IEEE Signal Processing Magazine, vol. 38, no. 4, pp. 56-72, 2021.
[10] A. Patel et al., "Integrating AI into Mobile Applications: Challenges and Opportunities," IEEE Intelligent Systems, vol. 36, no. 3, pp. 12-17, 2021.
[11] M. Johnson et al., Voice Assistants in Healthcare: Opportunities and Challenges," IEEE Journal of Biomedical and Health Informatics,vol. 25, no. 6, pp. 1823-1830, 2021.
[12] L. Chen et al., Voice-Based Educational Platforms: A Review and Future Directions," IEEE Transactions on Learning Technologies, vol. 14, no. 2, pp. 243-256, 2021.
[13] S. Gupta, Voice Assistant Technology—A Comprehensive Review, in Proceedings of the IEEE International Conference on Communication Systems and Networks, 2020, pp. 123-130. J. Smith, Design Principles for Voice Assistant Interfaces, in
[14] Proceedings of the IEEE International Conference on Human-Computer Interaction, 2019, pp. 45-52.
[15] A. Patel, AI Integration in Mobile Applications: Challenges and Opportunities," in Proceedings of the IEEE InternationalConference on Mobile Computing and Applications, 2021, pp. 78-85.
[16] R. Wang et al., Deep Learning for Voice Recognition: A Review, IEEE Signal Processing Magazine, vol. 38, no. 4, pp. 56-72, 2021.
[17] S. Gupta, D. Joshi, K. Kolpate, D. Sharma, and N. Parmar, Computer Science on Personal Voice Assistant," International Journal for Research in Applied Science and Engineering Technology, vol. 9, no. 5, pp. 123-134, 31 May 2022.
[18] U. Gupta, U. Jindal, A. Goel, and V. Malik, Desktop Voice Assistant, International Journal for Research in Applied Science and Engineering Technology, vol. 9, no. 5, pp. 135-145, 31 May 2022.
[19] Guha, T. Breßgott, D. Grewal, D. Mahr, M. Wetzels, E. B. Schweiger, How Artificiality and Intelligence Affect Voice Assistant Evaluations," Journal of the Academy of Applied Science and Engineering Technology, vol. 10, no. 2, pp. 156-167, 28 June 2022
[20] Saadman Shahid Chowdury, Atiar Talukdar, Ashik Mahmud, and Tanzilur Rahman, Domain-specific Intelligent Personal Assistant with Bilingual Voice Command Processing, IEEE, 2018.
[21] J. Kurpansky, What Is an Intelligent Digital Assistant? in Medium, 2017.
[22] J. Iso-Sipila, M. Moberg, and O. Viikki, Multi-lingual speaker-independent voice user interface for mobile devices, ICASSP 2006, pp. 1081-1084, 200
[23] Johnson et al., "Voice Assistants in Healthcare: A Case Study on Patient Engagement and Data Privacy, 2023 IEEE International Conference on Healthcare Informatics (ICHI).
[24] Wang et al., Long-Term Adoption of Amazon Alexa in Smart Homes: Benefits and Challenges," 2022 IEEE International Conference on Consumer Electronics (ICCE).
[25] Kumar and Singh, "Enhancing Retail Customer Service with Google Assistant: A Case Study, 2023 IEEE International Conference on Retail Innovations (ICRI).
[26] Park et al., "Voice Assistants in Language Learning: A Case Study on English Pronunciation for Non-Native Speakers," 2021

IEEE Conference on Educational Technology (ICET).