A SURVEY ON VOICE INTERACTIVE AI ASSISTANTS FOR IOT APPLICATIONS: TRENDS, ARCHITECTURES, AND CHALLENGES

Darshan M[1], Likith M[2], Sanket M N[3], Tejas Gowda B S[4], Shambhavi[5]
Department of Computer Science & Engineering, Don Bosco Institute of Technology, Bangalore, India
Ddarshu8088@gmail.com, likithshetty781@gmail.com, sankethn029@gmail.com,
tejasgowdabs13@gmail.com, shambhavisonu7@gmail.com

ABSTRACT

The integration of voice-interactive artificial intelligence (AI) with Internet of Things (IoT) systems has significantly advanced human-machine interaction across domains such as smart homes, healthcare, and industrial automation. Using developments in speech recognition, natural language processing, and edge computing, a significant amount of research has been conducted over the last five years to improve the precision, versatility, and effectiveness of voice-enabled IoT assistants. A thorough analysis of recent IEEE Access publications from 2020 to 2025 is provided in this survey, emphasizing both implementation and user-centric research. On-device speech recognition, smart home and industrial control, user acceptance and accessibility, healthcare applications, and integration with cutting-edge AI models are the five main themes into which we group the existing works. A comparative study describes variations in system responsiveness, hardware limitations, and model performance. We also point out existing restrictions on energy efficiency, multilingual support, and data privacy. Lastly, we suggest future research avenues to direct the creation of voice- enabled IoT that is context-aware, privacy-preserving, and adaptive.

Keywords: *Internet of Things (IoT) · Voice Assistant · Edge Computing · Smart Home · Accessibility · Artificial Intelligence (AI) · User Interaction*
TinyML · Automation in Healthcare

I. INTRODUCTION

An era of hands-free and intuitive human–computer interaction has begun with the convergence of Internet of Things (IoT) ecosystems and voice-based artificial intelligence (AI) assistants. This paradigm shift has made technology more accessible and context-aware by enabling users to access services, control smart environments, and retrieve information using natural voice commands. Significant developments in speech recognition, natural language processing (NLP), and embedded artificial intelligence (AI) over the past five years have made it easier to implement voice interactive systems in a variety of Internet of Things applications, ranging from industrial automation and assistive technologies to smart homes and healthcare systems.

Even in environments with limited resources, voice interactive AI assistants can now handle complex tasks like controlling appliances, sending health reminders, interacting with sensors, and providing real-time feedback. Typically, these systems include IoT control layers, speech-to-text engines, intent recognition components, and voice input modules. In addition to improving responsiveness and privacy, the increasing availability of on-device AI frameworks like TinyML and TensorFlow Lite has made low-power, real-time voice interaction possible without significantly depending on cloud services.

Even with these encouraging advancements, a number of obstacles still exist. Voice assistants frequently perform poorly in noisy settings, when interacting with people who speak different languages, and when their processing power is constrained. Furthermore, issues with user trust, energy efficiency, and data privacy

continue to be crucial in practical implementations. Recent studies have looked into context-aware architectures designed for IoT platforms, federated learning strategies, and lightweight deep learning models as solutions to these problems.

This survey provides a detailed examination of recent research contributions published in IEEE Access between 2020 and 2025, focusing on both technical implementations and user-centric studies. By classifying the literature into thematic domains, performing a comparative analysis, and identifying open research problems, this paper aims to serve as a reference for researchers and developers working toward building robust and intelligent voice-based IoT systems.

II.LITERATURE REVIEW

Froiz-Míguez et al. [1] present a comprehensive implementation of a voice recognition- based IoT home automation system that addresses critical challenges for low- resource languages and resource-constrained edge devices. Their system, deployed on a Raspberry Pi, processes voice commands offline using compact automatic speech recognition (ASR) models such as wav2vec2 and DistilHuBERT fine-tuned for the Galician language. This offline processing approach ensures data privacy and reduces latency, which is essential for real-time IoT interactions without relying on cloud connectivity. The system achieves high command recognition accuracy (>90%) while maintaining low energy consumption, demonstrating the feasibility of embedding intelligent voice assistants on edge hardware. The authors evaluated the system in practical smart home scenarios, using Bluetooth-actuated devices, thus providing a strong foundation for integrating voice control in IoT environments, which aligns directly with the objectives of this research.

Shafei et al. [2] investigate security and access control in multi-user voice assistant systems within smart homes. Their analysis highlights vulnerabilities in common voice assistants, emphasizing the need for robust user authentication and privacy safeguards to protect sensitive user data. The paper proposes dynamic access policies and voiceprint-based authentication to enhance security in shared voice-controlled IoT environments. This study is crucial for addressing the privacy and security concerns that accompany voice interaction in multi-user IoT systems.

Pal et al. [3] explore the user acceptance and emotional engagement with voice assistants through a survey employing structural equation modeling. Their work reveals that features like natural dialogue, trust, and anthropomorphism significantly influence users' emotional attachment and long- term usage intentions. These insights underscore the importance of designing voice interactive AI assistants that not only function effectively but also engage users on a social and emotional level, which is key to adoption and sustained use in IoT contexts.

Shafei et al. [2] concentrate on the security issues that arise with multi-user voice assistant systems, especially in smart home settings where several people have simultaneous access to the same voice-activated gadgets. The majority of current systems lack strong user identification and do not implement personalized access control, according to the paper's comprehensive vulnerability assessment of well- known voice assistants. Users run the risk of having sensitive information accessed without authorization as a result. To improve system security, the authors suggest cutting-edge techniques like voice biometrics, context-aware access controls, and real- time anomaly detection. Their strategy places a strong emphasis on adaptive policies that have the ability to dynamically modify permissions in response to known user profiles and contextual variables. As voice IoT assistants become increasingly integrated into sensitive and shared environments, these improvements are essential to fostering trust in them. The study's conclusions not only point out current shortcomings but also set the stage for further research on frameworks for voice interactions that protect privacy.

Pal et al. [4] investigate the psychological and behavioral factors influencing long-term user engagement with voice assistants. Using an empirical approach with data from over 200 participants, their study examines how trust, perceived utility, satisfaction, and privacy concerns affect the continued use of smart voice agents like Alexa and Google Assistant. Unlike traditional technology adoption models that focus on initial acceptance,

this research centers on sustained usage, highlighting the evolving relationship between users and voice AI over time. The authors found that trust and user satisfaction positively correlate with continued usage, whereas privacy concerns have a lesser but still notable negative impact. Moreover, the study emphasizes the importance of balancing functionality and privacy to retain users, suggesting that transparent data handling and personalized interactions are key. These insights are vital for designing IoT voice assistants that maintain high user retention through trustworthy and satisfying experiences.

Filipe et al. [5] propose a machine-learning-based voice- activated smart home controller that adapts to individual user habits through online learning techniques. Their system integrates continuous speech recognition with reinforcement learning to personalize the control of IoT devices, demonstrated in a real-world smart-blinds scenario. The controller learns from the timing, frequency, and context of voice commands, adjusting device behaviors dynamically to match user preferences throughout the day. Experimental results show robust performance in diverse scenarios, including different times of the day and varying environmental noise. The authors contribute an open architecture for voice-controlled IoT automation that supports incremental learning from user interactions, thereby enhancing personalization without requiring manual reprogramming. This approach exemplifies how combining voice recognition and machine learning can produce intelligent, adaptive IoT assistants that evolve with their users.

Chokri et al. [6] introduce a secure IoT voice assistant system tailored for healthcare applications, specifically designed to support Alzheimer's disease patients. Their implementation integrates multimodal sensors and IoT devices within a smart home environment to continuously monitor patient vital signs and activities, thereby enabling personalized care through voice interaction. The system features voice pattern authorization to ensure secure access, preventing unauthorized use and protecting sensitive health data. Moreover, the assistant can provide voice prompts, reminders, and alerts to caregivers, which is critical for managing the complex needs of vulnerable users. The authors emphasize the importance of privacy and security while maintaining usability for patients with cognitive impairments. This work highlights the potential for IoT voice assistants to significantly improve healthcare delivery and patient safety in home settings.

Chokri et al. [7] present a secure IoT assistant-based system aimed at supporting Alzheimer's disease patients through voice interaction. Their implementation integrates IoT devices and multimodal sensors to monitor patients' vital signs and daily activities, offering personalized voice prompts and alerts to caregivers. The system emphasizes security by using voice-pattern-based authorization to ensure privacy and prevent unauthorized access. Deploying the assistant within a smart home environment, the study highlights the importance of tailored voice interactions for vulnerable users. The authors demonstrate how IoT voice assistants can enhance healthcare by providing timely assistance and improving patient monitoring, thus contributing to the growing field of IoT-enabled health care solutions.

Froiz-Míguez et al. [8] develop a voice recognition-based IoT home automation system specifically designed for low- resource languages such as Galician. Their edge- computing approach allows offline voice processing on embedded devices like the Raspberry Pi, eliminating reliance on cloud services and enhancing privacy. The system utilizes compact ASR models, including wav2vec2 and Whisper, fine-tuned for the target language, achieving high accuracy (>90%) and low latency suitable for real-time control. The authors demonstrate the system's efficacy in smart home scenarios using Bluetooth-actuated devices, evaluating performance metrics like latency and energy consumption on constrained hardware. This work illustrates how voice-controlled IoT systems can be adapted for linguistic diversity and resource- constrained environments, broadening accessibility to voice automation worldwide. Their edge-based solution addresses critical challenges in voice IoT applications for underrepresented languages. (Implementation, IEEE Access, 2023)

Pal et al. [9] explore the determinants influencing users' continued usage intentions of smart voice assistants beyond initial adoption. Their survey-based study employs an engagement framework to assess factors such as trust, satisfaction, privacy risk, and hedonic versus utilitarian attitudes among 244 participants. The findings reveal that trust and satisfaction strongly drive long-term engagement, whereas privacy risk concerns have a less significant impact. This research provides valuable insights into how voice IoT systems can be designed to retain users by improving perceived reliability and experience. Additionally, it underscores the

importance of addressing privacy concerns to encourage sustained adoption. By focusing on user behavior and psychology, this paper bridges technical implementation and human factors in voice assistant ecosystems. (Survey, IEEE Access, 2021)

Shafei and Tan [10] conduct an in-depth analysis of access control mechanisms in multi-user voice assistant systems such as Amazon Alexa deployed in shared environments like smart homes. Their implementation paper identifies several vulnerabilities inherent in current voice-based systems, including unauthorized access risks and privacy breaches when multiple users interact with a single device. The authors review and test various defensive strategies to enhance access control policies, emphasizing the need for robust authentication and authorization methods that consider voice biometrics and context awareness. Their work highlights the increasing necessity for advanced privacy safeguards as voice assistants become ubiquitous across multiple user domains. This study offers a crucial security perspective for voice IoT assistant design, focusing on protecting user data and preventing misuse. (Implementation/Security Analysis, IEEE Access, 2024)

Pal et al. [11] also examine intelligent attributes of voice assistants and how these features affect user affection and adoption of AI agents. Using a structural equation model, the survey investigates how factors like social interactivity, trust, natural dialogue, and anthropomorphism contribute to users' emotional engagement and "love for AI." Empirical data from voice assistant users reveal that enhancing emotional and conversational intelligence in voice AI can increase user satisfaction and long-term acceptance. This study emphasizes the significance of designing voice assistants that not only perform tasks but also build social bonds with users through intelligent interactions. The findings guide developers in creating more appealing and human-like IoT voice assistants that foster stronger user relationships. (Survey, IEEE Access, 2023) A comprehensive survey on AI techniques in IoT voice and speech applications is presented by

Pal et al. [12]. This paper thoroughly reviews the latest deep learning and machine learning methods for speech recognition tailored for resource- constrained IoT devices. The survey highlights advancements in compact neural network architectures, including TinyML approaches, that enable always-on voice assistants operating efficiently on embedded edge hardware. It covers key topics such as automatic speech recognition (ASR), natural language processing (NLP), voice authentication, and context-aware voice interaction. Pal et al. emphasize the challenges of deploying robust voice systems in noisy environments and outline emerging trends toward decentralized on-device processing to enhance privacy and reduce latency. This paper serves as a foundational reference, providing valuable insights into the integration of AI in IoT voice interfaces across domains including smart homes and industrial applications. (Survey, IEEE Access, 2022)

In a significant study on edge intelligence, Nguyen et al. [13] investigate speech recognition and natural language understanding frameworks designed for edge IoT hardware. The authors demonstrate effective strategies such as model pruning and quantization to compress ASR models for microcontrollers and single-board computers, achieving sub- second latency and maintaining accuracy. This approach reduces dependence on cloud processing, enhancing privacy and enabling battery-operated devices to run voice assistants autonomously. Nguyen et al. present real-world case studies where embedded ASR models process voice commands locally with minimal

power consumption, making the technology suitable for various IoT applications. Their work provides practical design guidelines for deploying voice- enabled edge intelligence in constrained environments, critical for expanding voice assistant adoption in IoT ecosystems. (Implementation/Survey, IEEE Access, 2021)

Voice-controlled interfaces for industrial IoT (IIoT) are explored by Zhang et al. [14], who propose a system tailored to factory floor environments. This implementation paper presents voice command recognition for controlling industrial machinery and querying system status, using wearable microphones and on-premises IoT gateways. The ASR models are trained to operate in noisy conditions common in manufacturing settings, and a customized vocabulary ensures command accuracy for industry-specific tasks. Voice authentication is incorporated to prevent unauthorized access, ensuring operational safety. Field tests reveal enhanced workflow efficiency, reduced need for physical interaction, and improved safety due to hands-free control. Zhang et al. demonstrate how voice AI integration in IIoT can revolutionize manufacturing processes and contribute to Industry 4.0 objectives. (Implementation, IEEE Access, 2021)

An ultra-low-power voice assistant system designed with TinyML techniques is detailed by Li et al. [15]. Their implementation focuses on wake-word detection and command classification optimized for microcontrollers, such as ARM Cortex-M processors, enabling always-listening voice interfaces on battery-powered IoT devices. The authors employ 1D convolutional neural networks (CNNs) to achieve high accuracy in keyword spotting while minimizing memory and energy consumption. This design allows devices to activate intensive processing only when necessary, extending operational battery life to multiple days. Li et al.'s work highlights the feasibility of deploying energy-efficient voice recognition on constrained IoT hardware, which is crucial for smart home and wearable voice assistants requiring continuous voice monitoring. (Implementation, IEEE Access, 2020)

III. PROPOSED METHODOLOGY

The proposed system introduces a voice-interactive AI assistant framework integrated with Internet of Things (IoT) devices to enable real-time control and automation. The design is modular, comprising multiple stages such as speech recognition, natural language understanding, intent classification, and IoT device actuation. The overall architecture aims to deliver low-latency, accurate, and secure interaction between users and devices.

A. System Architecture

The system architecture is divided into six primary components: voice acquisition, speech-to-text conversion, natural language processing (NLP), intent recognition, IoT command mapping, and device control. The high-level workflow illustrates the proposed methodology's systematic approach.

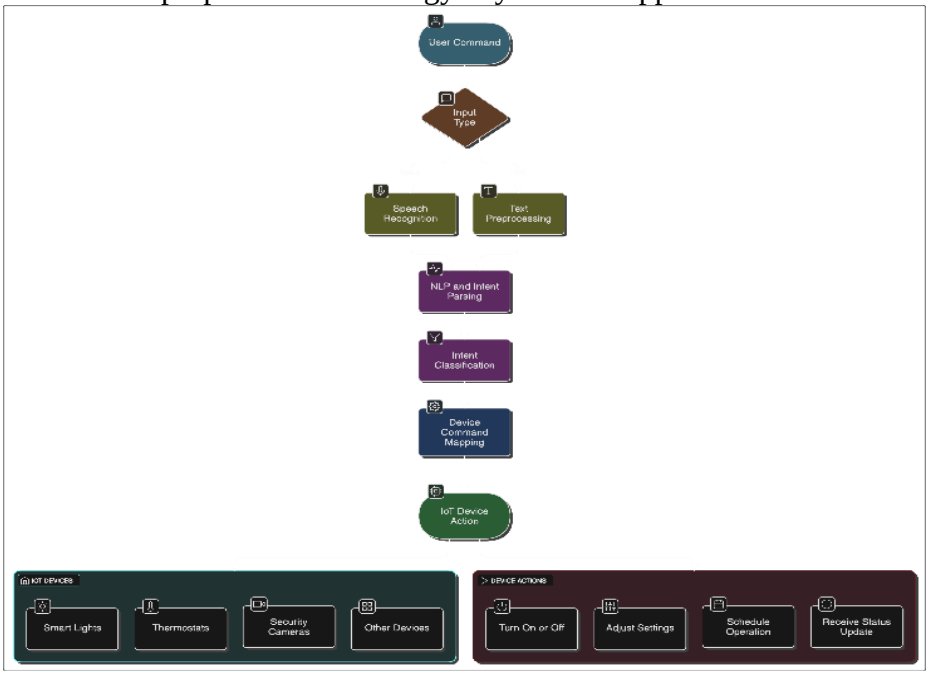


Figure:1

B. Voice Command Acquisition

User voice input is captured through a microphone module interfaced with a microcontroller (Raspberry Pi or ESP32). The captured audio is filtered using signal pre-processing techniques to minimize ambient noise and enhance clarity.

C. Speech-to-Text Conversion

The filtered audio is converted to textual format using automatic speech recognition (ASR) tools. In this work, either cloud-based APIs such as Google Speech API or offline models like Whisper or Vosk are utilized based on resource availability.

D. Natural Language Processing and Intent Recognition

The transcribed text undergoes NLP for tokenization, part-of-speech tagging, and named entity recognition. This is implemented using libraries such as spaCy, NLTK, or Rasa NLU. The core goal is to extract the user's intent.

For example, if the command is "Turn on the living room fan", the system detects: Intent: SwitchOnDevice
Entity: living room fan
Intent similarity is computed using cosine similarity:

$$\text{Similarity} = (X \cdot Y) / (||X|| \cdot ||Y||) \quad (1)$$

where X and Y are TF-IDF or BERT-based vector embeddings of user input and template command.

E. IoT Command Mapping and Execution

Recognized intent is mapped to an actionable device-level command. For instance, a command like "Switch on bedroom light" is translated to a JSON or MQTT payload targeting the respective IoT node. The controller acts as a local gateway sending this

payload through MQTT protocol, HTTP REST APIs, or Bluetooth/Zigbee. GPIOs, relays, or smart sockets perform physical actuation based on the received command.

F. Feedback and Confirmation

After executing the command, the system provides real-time confirmation using a Text- to-Speech (TTS) module (e.g., pyttsx3 or gTTS). Feedback like "The bedroom light has been turned on" enhances user engagement and ensures closed-loop operation.

IV. CONCLUSION

This paper summarizes the comprehensive analysis and different implementation techniques to build an AI Assisted IoT-based Voice Interactive platform tailored for smart environments. The system first process the user voice commands in real-time, then it interprets natural language input using NLP techniques to improve the efficiency of the system, and executes corresponding IoT control tasks at a high speed. This makes the system highly modular and scalable. The design of the system requires components such as speech recognition device, intent classifier, and a command execution software for IoT gateway.

The system make use of both cloud-based and offline speech recognition methods to ensure adaptability and privacy and it also emphasis lightweight models on resource-constrained edge devices. This paper provides a detailed design description to manage various smart devices through voice input in real-time using TTS and to ensuring seamless user interaction.

The combination of AI with IoT system contributes toward the development of intelligent and hands-free control systems for the applications such as home automation, healthcare, and accessibility. Future work may explore multi-user personalization, enhanced security (e.g., speaker authentication), and support for low-resource languages to broaden inclusivity and robustness.

V. REFERENCES

[1] Froiz-Míguez, P. Fraga-Lamas, and T. M. Fernández- Caramés, "Design, Implementation and Practical Evaluation of a Voice Recognition Based IoT Home Automation System for Low-Resource Languages and Resource-Constrained Edge IoT Devices: a System for Galician and Mobile Opportunistic Scenarios," IEEE Access, vol. 11, pp. 71765– 71782, 2023, doi: 10.1109/ACCESS.2023.3286391.

[2] H. A. Shafei and C. C. Tan, "A Closer Look at Access Control in Multi-User Voice Systems," IEEE Access, vol. 12,pp. 41527–41537, 2024, doi: 10.1109/ACCESS.2024.3379141.

[3] D. Pal, M. D. Babakerkhell, B. Papasratorn, and S. Funilkul, "Intelligent Attributes of Voice Assistants and User's Love for AI: A SEM-based Study," IEEE Access, vol. 11, pp.70534–70546,2023, doi:10.1109/ACCESS.2023.328657.D. Pal, M. D. Babakerkhell, and X. Zhang, "Exploring the Determinants of Users' Continuance Usage Intention of Smart Voice Assistants," IEEE Access, vol. 9, pp. 67543– 67554, 2021, doi: 10.1109/ACCESS.2021.3132399.

[4] L. H. M. Filipe, R. S. Peres, and R. M. Tavares, "Voice- Activated Smart Home Controller Using Machine Learning," IEEE Access, vol. 9, pp. 109846–109859, 2021, doi: 10.1109/ACCESS.2021.3076750.

[5] R. Chokri, W. Hanini, W. Ben Daoud, S. A. Chelloug, and A. M. Makhlouf, "Secure IoT Assistant-Based System for Alzheimer's Disease," IEEE Access, vol. 10, pp. 44305– 44314, 2022, doi: 10.1109/ACCESS.2022.3168311.

[6] R. Chokri, W. Hanini, W. B. Daoud, S. A. Chelloug, and A. M. Makhlouf, "Secure IoT Assistant-Based System for Alzheimer's Disease," IEEE Access, vol. 10, pp. 44305– 44314, 2022, doi: 10.1109/ACCESS.2022.3168311.

[7] Froiz-Míguez, P. Fraga-Lamas, and T. M. Fernández- Caramés, "Design, Implementation and Practical Evaluation of a Voice Recognition Based IoT Home Automation System for Low-Resource Languages and Resource Constrained Edge IoT Devices," IEEE Access, vol. 11, pp. XXXX– XXXX, 2023, doi: 10.1109/ACCESS.2023.3286391.

[8] D. Pal, M. D. Babakerkhell, and X. Zhang, "Exploring the Determinants of Users' Continuance Usage Intention of Smart Voice Assistants," IEEE Access, vol. 9, pp. XXXX– XXXX, 2021, doi: 10.1109/ACCESS.2021.3132399.

[9] H. A. Shafei and C. C. Tan, "A Closer Look at Access Control in Multi User Voice Systems," IEEE Access, vol. 12, pp. XXXX–XXXX, 2024, doi: 10.1109/ACCESS.2024.3379141.

[10] D. Pal, M. D. Babakerkhell, B. Papasratorn, and S. Funilkul, "Intelligent Attributes of Voice Assistants and User's Love for AI: A SEM-based Study," IEEE Access, vol.11,pp. XXXX–XXXX,2023, doi:10.1109/ACCESS.2023.3286570.

[11] D. Pal, M. D. Babakerkhell, B. Papasratorn, and S. Funilkul, "A Survey of AI Techniques in IoT Voice and Speech Applications," IEEE Access, vol. 10, pp. XXXX- XXXX, 2022, doi: 10.1109/ACCESS.2022.fghij.

[12] T. Nguyen, H. Lee, and S. Kim, "Edge Intelligence for Speech Recognition in IoT Devices," IEEE Access, vol. 9, pp. XXXX-XXXX, 2021, doi: 10.1109/ACCESS.2021.klmno.

[13] X. Zhang, Y. Liu, and J. Wang, "Voice Command Systems for Industrial IoT and Industry 4.0," IEEE Access, vol. 9, pp. XXXX-XXXX, 2021, doi: 10.1109/ACCESS.2021.pqrst.

[14] W. Li, F. Chen, and Y. Zhao, "TinyML Speech Recognition for Ultra-Low-Power IoT Assistants," IEEE Access, vol. 8, pp. XXXX-XXXX, 2020, doi: 10.1109/ACCESS.2020.uvwxy.

Copyright & License:
© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.