Survey on Techniques to Build AI Assistant

Nikhil Mathew Abhilash[1], S Sidharthan[2], Abhimanyu Rajan[3], Arun Chandran[4], Anjana C V[5]
[1,2,3,4] UG Student, Department of CSE, College of Engineering Kidangoor, Kottayam, Kerala, India
[5]Assistant Professor, Department of CSE, College of Engineering Kidangoor, Kottayam, Kerala, India
[1]nikhilnma.10@gmail.com, [2]sidharthan2710@gmail.com, [3]abhimanyurajan396@gmail.com,
[4]arunchandran54499@gmail.com, [5]anjanachennothu@gmail.com

***Abstract*—The rapid advancement of artificial intelligence has transformed human-computer interaction, with AI assistants emerging as vital tools for productivity and convenience. This survey explores the techniques and technologies underlying AI assistants, emphasizing their core functionalities, including natural language processing, voice recognition, and biometric security. Python, renowned for its simplicity and robustness, plays a pivotal role in the development of such systems. Key topics discussed include speech-to-text conversion using APIs, task automation through voice commands, and advanced features like face recognition for secure, personalized interactions. The survey also highlights privacy-preserving mechanisms, multilingual capabilities, and user-centric design enhancements. Addressing challenges such as data privacy, usability across demographics, and computational efficiency, this paper consolidates recent advancements and identifies research gaps. The findings emphasize the transformative potential of AI assistants in streamlining everyday tasks, enhancing personalization, and fostering seamless, secure human-computer collaboration.**

***Index Terms*—Artificial Intelligence, Natural Language Processing, Face and Voice Recognition**

I. INTRODUCTION

Artificial Intelligence has revolutionized human-computer interaction, giving rise to intelligent systems that simplify everyday tasks. Among these, AI assistants have become indispensable, integrating natural language processing (NLP), voice recognition, and automation to enhance user convenience and productivity. Popular assistants like Siri, Alexa, and Google Assistant exemplify the rapid progress in this domain. However, despite their widespread adoption, these systems face several limitations that restrict their usability and security. One significant challenge lies in ensuring robust biometric authentication. Many AI assistants rely solely on single-factor voice authentication, leaving them vulnerable to unauthorized access. Addressing these security concerns demands the integration of multi-modal biometric systems, such as face and voice recognition, to provide secure and reliable access control.Another critical issue is privacy. AI assistants often process sensitive data on centralized cloud servers, raising concerns about data breaches and unauthorized usage. Techniques such as privacy-preserving edge computing and encrypted data handling have emerged as potential solutions, but their implementation remains complex and resource-intensive. Usability challenges further complicate the adoption of AI assistants across diverse user demographics. Factors such as accent variability, dialect recognition, and accessibility for older adults or those with limited technological proficiency highlight the need for more inclusive designs. Similarly, computational efficiency is essential for ensuring these systems operate effectively on resource-constrained devices, such as smartphones and embedded systems.

This survey aims to provide a comprehensive analysis of existing techniques and advancements in AI assistant development. It explores innovations in voice and face recognition, keyword spotting, multilingual interfaces, and chatbot architectures while addressing critical issues like privacy, security, and user adaptability. By consolidating state-of-the-art research and identifying key challenges, this paper offers insights into the current landscape and opportunities for future advancements in AI assistant technologies.

II. LITERATURE SURVEY

A. Review of Face Recognition

Lixiang et al. [1] provides a comprehensive examination of the advancements in face recognition technology, which stands as a significant branch within the field of biometric identification. Face recognition operates by identifying the distinct facial features of an individual, allowing the technology to recognize and verify identities accurately. This paper discusses the entire process, starting with the collection of facial images, followed by their processing by specialized recognition equipment, which then analyzes the facial features to match and verify identity. This research review explores face recognition from multiple perspectives, covering the foundational technologies, the various stages of its development, and the advancements that have shaped its progress to date.

Figure 1 shows the classification of deep learning in face recognition applications. It highlights methods such as convolutional neural networks, deep nonlinear face shape extraction, deep learning video surveillance, and low-resolution face recognition. This paper dives into face recognition's roots in visual pattern recognition, a key concept in artificial intelligence and machine learning, whereby computers learn to classify visual patterns through the interpretation of pixel data. For humans, recognizing visual patterns is an intuitive process, with information processed and interpreted by the brain to form meaningful concepts. However, in a computational context, images are nothing more than matrices of pixels. It is through complex algorithms that machines are able to analyze and assign meaning to these data structures, discerning faces

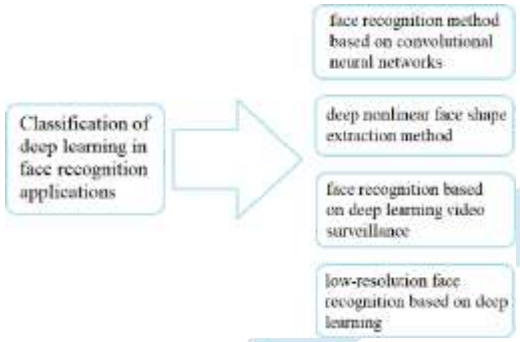

Fig. 1. Classification of deep learning in face recognition applications [1].

and interpreting them as belonging to specific individuals. The paper positions face recognition as a specialized subset of general visual pattern recognition, with the primary task being to identify distinct faces within data that the system classifies as relevant facial images.This paper provides an overview of current algorithms and neural network architectures that contribute to the speed and efficiency of face detection and positioning. The deep learning approaches discussed in the paper significantly reduce the computational load compared to traditional methods, offering faster results in real-world applications.

This Paper will describe the development stages and related technologies of face recognition, including early algorithms, artificial features and classifiers, deep learning and other stages. After that, we will introduce the research on face recognition for real conditions. Finally, we introduce the general evaluation criteria and general databases of face recognition.

B. A Privacy-Preserving Edge Computation-Based Face Verification System for User Authentication

Xiang et al. [2] presents an advanced framework for secure face recognition systems, particularly focusing on user privacy in identity authentication . Given the unique and non-invasive nature of facial biometrics, face recognition has become a widely adopted method for authentication. However, while many current systems rely on outsourcing facial data processing to external servers, this approach raises substantial privacy risks due to the sensitive nature of facial data. The authors address these concerns by introducing a privacy-preserving system that leverages edge computing to enhance data security and reduce dependence on cloud- based processing.

The system architecture utilizes convolutional neural networks (CNNs) to extract facial features accurately and effectively, with specific attention on maintaining user privacy. To tackle the challenge of data sensitivity, the authors propose a secure nearest neighbor approach that computes the cosine similarity over encrypted feature vectors, allowing the system to verify identities without exposing the raw data. This process ensures that even if an unauthorized party accesses the server, the encrypted information prevents any actual facial data leakage. The use of edge computing further strengthens the security of the system by performing sensitive data operations closer to the user, rather than on a centralized cloud server, thus mitigating the risk of privacy breaches during data transmission. A notable innovation in the proposed system is the secret sharing homomorphism technology, which enables distributed computing, making the system more resilient to faults. This review proposed a new image recognition and classification approach for welding faults that used a transfer learning algorithm with the MobileNet model. They proposed adding a new Fully Connect layer (FC-128) and a Softmax classifier MobileNet to the TL-MobileNet structure. They made use of the GDXray Dataset (6,208 defect samples which contains 5 types of defects). The DropBlock technology and the Global Average Pooling (GAP) method were used to optimize the whole training process of the TL-MobileNet model.

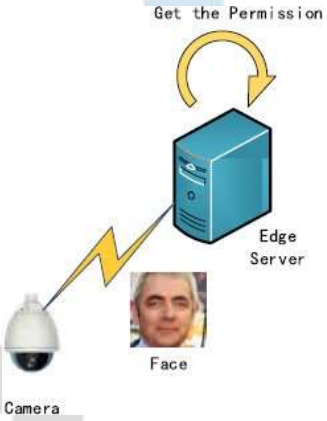

Fig. 2. The overview of the step one in identity authentication. [2]

Figure 2 illustrates the process of face recognition using an edge server. A camera captures the face, and the edge server processes it to grant permission. This setup ensures real-time decision-making for access control.The use of these methods enhanced the rate of convergence as well as generalization of classification networks. The suggested TL-MobileNet was tested on the welding defects dataset, and the prediction accuracy of the model.This approach supports a higher degree of fault tolerance, ensuring continuous authentication service even if certain nodes in the network are compromised or experience failures. Secret sharing enhances the system's security by dividing sensitive data into smaller, encrypted fragments distributed across multiple locations, preventing unauthorized access to complete data sets. Provides effective privacy protection while maintaining high authentication accuracy.Through extensive experiments, they validate the efficiency and reliability of their proposed methods, showing that this edge-computing-based approach outperforms traditional cloud-dependent models in terms of the advantages of combining privacy-preserving cryptographic techniques with edge computing to create a robust authentication system suitable for modern biometric applications. was found to be 97.69%.

C. Comprehensive Review of Face Recognition Techniques, Trends, and Challenges

Francesco et al. [3] explores the advances, techniques, and ongoing challenges in the field of Face Recognition. Highlights how FR has become integral in many sectors, including security, healthcare, finance, and criminal identification, due to its reliability in identifying and verifying individuals based on unique facial characteristics.

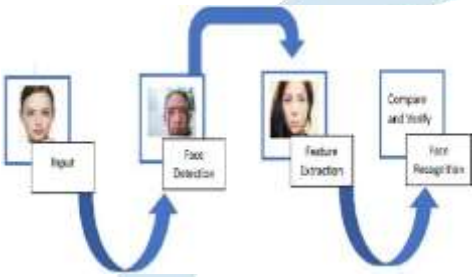

Fig. 3. The workflow of a standard automated FR System Biometric [3]

Figure 3 depicts the face recognition workflow. It begins with input image processing, followed by face detection and feature extraction. The extracted features are then compared and verified for face recognition. The study categorizes face recognition (FR) methods into appearance-based, holistic, and hybrid approaches. It outlines the FR workflow, starting with input image processing, followed by face detection, feature extraction, and comparison for recognition. The review highlights challenges like lighting, pose variation, expressions, occlusions, and aging, emphasizing the need for robust systems to address these issues. Recent advancements leverage large, diverse datasets to improve accuracy and handle variability in real-world conditions. The paper also underscores the importance of well-structured datasets for training FR models and identifies ongoing research gaps, providing valuable insights for future advancements in the field.

D. Personal Voice Assistant Security and Privacy

Pengcheng et al. [4] address the growing concerns surrounding the security and privacy of personal voice assistants (PVAs) as these devices become integral to everyday life. PVAs, such as Amazon Echo, Google Home, and Apple's Siri, allow users to interact with digital environments and control a variety of smart devices like phones, smart homes, and even cars through voice commands. The paper highlights the massive rise in PVA usage, particularly in the United States, where the number of smart speakers grew by 78 to 118.5 million, and 21 percent of the population now owns at least one smart device with PVA capabilities. As society's reliance on these devices increases, so does the demand for research that addresses security and privacy issues associated with their use. The paper delves into privacy implications associated with PVAs, particularly concerning user consent for voice recording and data retention. The authors examine how recording consent is managed and raise questions about user control over PVAs that may continuously listen or record without survey provides a comprehensive research map that categorizes current efforts and challenges in PVA security and privacy, offering valuable insights into both established and emerging issues. By addressing a broad array of potential risks associated with PVAs, the paper serves as a foundation for further research aimed at enhancing the resilience of PVAs against security threats while prioritizing user privacy in an increasingly interconnected digital landscape.

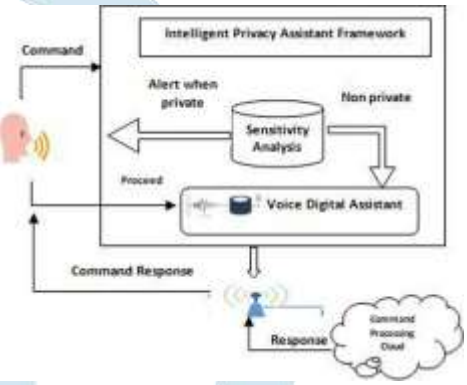

Fig. 4. Personalized privacy assistant for digital voice assistants. [4]

Figure 4 illustrates Intelligent Privacy Assistant Framework for processing voice commands. It performs sensitivity analysis to determine whether a command is private; if private, it alerts the user, and if not, it forwards the command to the Voice Digital Assistant for processing. The response is generated via a command processing cloud and returned to the user. This feature ensures that access is only granted to verified users, thus addressing some of the security vulnerabilities described in this survey. Moreover,it incorporates a real-time alert feature for unauthorized access attempts, a response to concerns about unintended PVA activations and privacy breaches. The inclusion of multi-user personalization based on voice recognition further differentiates your system by offering user-specific actions securely. In summary, this paper insights on PVA privacy challenges inform the security-focused features in your project, while the multi-modal authentication and user-specific functionalities in your assistant effectively mitigate many of the issues outlined in this survey.

E. Deep Spoken Keyword Spotting: An Overview

Ivan et al. [5] presents a comprehensive review of deep spoken keyword spotting (KWS), a specialized area in voice recognition technology aimed at detecting specific keywords in audio streams . With the advancements brought by deep learning, KWS has become increasingly efficient and has been integrated into various small electronic devices for tasks like activating voice assistants, such as Amazon's Alexa, Apple's Siri, and Google Assistant.

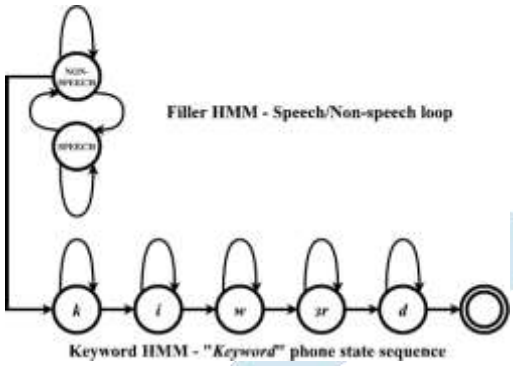

Fig. 5. Scheme of a keyword/filler HMM-based KWS system, when the system keyword is "keyword". While typically the keyword is modeled by a context-dependent triphone-based HMM, a monophone-based HMM is depicted instead for illustrative purposes. The filler HMM is often a speech/non-speech monophone loop. [5]

The paper attributes the growing popularity of KWS to the ability of deep learning models to improve the precision of keyword identification in audio inputs, thereby enhancing user experience and expanding application possibilities. It covers several aspects of deep KWS systems, breaking down critical components such as speech feature extraction, acoustic modeling, and posterior handling—each crucial for making KWS both accurate and computationally efficient. A particular focus is placed on the challenges of robustness in KWS, especially in noisy environments, as well as on computational complexity in embedded systems where power and processing capacity are limited. The paper discusses methods for achieving a small model footprint, a requirement for devices that need to manage KWS processes locally without requiring significant resources. Furthermore, explores the applications of KWS including speech data mining, audio indexing, and phone call routing. Each of these applications benefits from the rapid, on-demand identification of keywords, which streamlines processes that would otherwise involve extensive computational resources for full-scale automatic speech recognition (ASR).

F. Comparing Voice Assistant Risks and Potential with Technology-Based Users

Andreas et al. [6] examines the adoption trends and user concerns associated with voice assistants. A technology that has seen rapid adoption globally yet remains a subject of mixed perceptions due to issues such as privacy, speech intelligibility, and usability. Voice assistants, which use voice user interfaces (VUIs) to allow hands-free interaction with technology, have gained popularity with systems like Amazon's Alexa, Apple's Siri, and Google Assistant. This paper explores factors impacting VA adoption, such as the availability of VAs on consumer devices (smartphones, tablets, and smart home appliances) and the disparity between device availability and actual usage frequency, which can be influenced by privacy and usability concerns.This paper highlights that despite the widespread availability of VAs, challenges remain in optimizing user experience (UX) and Privacy concerns remain a significant barrier, especially as VAs are often equipped with always-on microphones that are perceived to be monitoring even when not actively engaged by the user. This issue raises security and privacy questions, as many users are unsure about data handling policies and potential misuse of voice data. Additionally, suggests that VAs could improve by enhancing speech recognition capabilities and providing better user control over voice data and functionality. Users have expressed a need for improved transparency around data storage and management practices, which could encourage more secure and trusted interactions.

While the study emphasizes cultural differences in VA adoption and privacy concerns, its findings also show that certain usability challenges and privacy concerns are universally observed. Regardless of user backgrounds, there is a shared demand for improvements in voice accuracy, data privacy management, and customizable user settings to better accommodate user preferences.

G. Speaker Diarization and Identification From Single Channel Classroom Audio Recordings Using Virtual Microphones

Antoniogomez et al. [7] introduces an innovative method for speaker diarization and identification specifically tailored for complex environments like classrooms, where multiple speakers talk simultaneously and only a single microphone is used to record all voices . Traditional speaker diarization struggles in such settings due to background noise, speaker overlap, and variances in voice characteristics (e.g., age, gender). These issues are compounded by limitations in existing deep learning models, which demand large datasets that may not adequately reflect such intricate environments.In this paper we have demonstrated the advantages of using virtual microphones and cross-correlation patterns to identify speakers in very challenging classroom environments from a single-channel recording.

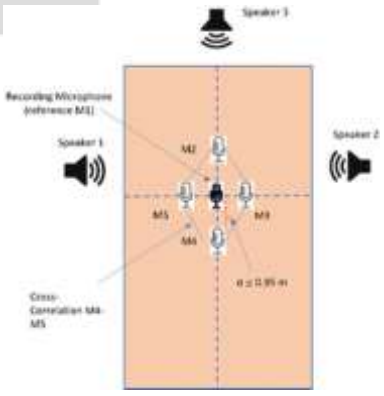

Fig. 6. Example placements of the virtual microphones and student speakers for the proposed method. [7]

Figure 6 illustrates a setup for sound localization using multiple microphones (M1 to M5) and speakers (Speaker 1, 2, and 3). Microphone M1 acts as a reference, while cross-correlation between M4 and M5 helps determine the position

of the sound source. The setup ensures accurate sound capture within a maximum distance of 0.95 meters. Our method presented an error rate that was significantly better than state-of-the-art systems from Amazon AWS and Google Cloud. Furthermore, in contrast with other methods, our proposed approach does not require extensive training, and it is directly applicable in challenging classroom audio environments where clean audio datasets are not available.

H. Personal Assistant With Voice Recognition Intelligence

Kulhalli et al. [8] proposes a virtual personal assistant designed to offer seamless voice-controlled interaction similar to existing assistants like Siri and Google Voice Search, with the notable feature of functioning both online and offline. The assistant utilizes Compact Large Vocabulary Speech Recognition (CLVSR), which leverages Connectionist Temporal Classification (CTC)-based Long Short-Term Memory (LSTM) acoustic models for improved accuracy in recognizing diverse vocabulary sets. To ensure efficient processing and real-time response, the system also employs Quantized Deep Neural Networks (DNNs).

Fig. 7. Working of Voice Assistant. [8]

This quantization reduces the computational requirements, making the assistant lightweight and suitable for deployment on devices with limited processing power. Figure 7 illustrates the working mechanism of voice assistants. A human user speaks a command (1), which is converted into a digital signal (2) and processed using Natural Language Processing (NLP) to interpret the intent (3). The assistant retrieves relevant data via APIs (4) and provides a response through the device (5). The system enhances the end-user experience by enabling voice access to various device functionalities and services without the need for traditional manual navigation. By allowing voice-based interaction independent of internet connectivity, this virtual assistant is more accessible in scenarios where network access is limited or unavailable, expanding its usability across diverse environments.The assistant provides a user-friendly approach to managing and accessing device functionalities through natural language, simplifying tasks such as calling, texting, app navigation, and system operations. By bypassing the need for constant connectivity, this assistant offers significant advancements in accessibility, making it a practical solution for scenarios with intermittent internet access and enhancing the convenience of device management across diverse environments.

I. A Systematic Review of Chatbots: Classification, Development, and Their Impact

Lamya et al. [9] presents a detailed study on the evolution, classification, architecture, and application of chatbots, particularly within the tourism sector. This study reviews the integration and impact of AI-driven chatbots in the tourism industry, addressing a gap in academic research. It classifies chatbots, explores their architecture, and compares development tools to highlight deployment challenges and benefits. The paper examines use cases within tourism's key areas—Accessibility, Accommodation, Amenities, Activities, and Ancillary Services—showing how chatbots transform customer service and business functions. Findings provide insights for developing efficient chatbot systems in industries requiring dynamic user interaction.

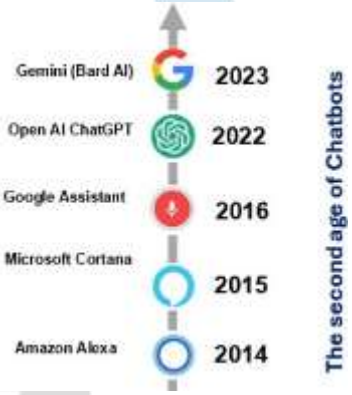

Fig. 8. Second age of Chatbots. [9]

The SLR is carried out to examine the most recent advances in research on chatbots and their impact on the tourism sector.Figure 8 highlights key milestones in the evolution of chatbots and voice assistants. Starting with Amazon Alexa in 2014, advancements include Microsoft Cortana (2015), Google Assistant (2016), OpenAI ChatGPT (2022), and Google's Gemini Bard AI (2023). It represents the growing sophistication of conversational AI in understanding and interacting with users. This review is performed after a critical analysis of the most pertinent research articles published in five well-known online digital libraries: Scopus, ACM, IEEE Xplore, Springer Link, and Web of Science. Six research questions regarding the different aspects of chatbot progress (classification, architecture, development tools) and their main use and impact on the tourism sector. It is concluded that chatbots are rapidly evolving and proliferating across all fields of tourism.Aim to address challenges associated with development tools, such as NLP service locking, through the investigation of chatbots. However, this goal can be achieved by creating a domain-specific language that allows for the development of these agents independently of existing tools.

J. A Framework to Enhance User Experience of Older Adults With Speech-Based Intelligent Personal Assistants

Chaudhry et al. [10] investigates the usability barriers older adults encounter when interacting with speech- based intelligent personal assistants (sIPAs), aiming to understand how these devices impact their quality of life. Through semi-structured interviews with fourteen older adults, the study identifies two main themes: how participants use sIPAs and their concerns regarding these devices.

The findings reveal that, while participants currently use sIPAs for various functions and have ideas for future applications, they face several challenges. These include privacy issues, limited interpersonal skills in devices, and a lack of contextual awareness. The study emphasizes that, despite the intuitive, persist, especially for older adults unfamiliar with technology. This demographic often feels unconfident navigating new devices independently, as evidenced by a Pew Survey where 34 percent of adults over 65 reported lacking confidence in learning new technology alone, with 73 percent preferring assistance. To enhance sIPAs for users, the researchers recommend several design and implementation improvements. Suggested enhancements include permission-based data storage, explainable AI (XAI) for transparency, accent and dialect recognition, and more humanized communication behaviors. These practical suggestions aim to make sIPAs more accessible and encourage wider adoption among older adults.

Fig. 9. framework for generating humanized conversation. [10]

Though many older adults views IPAs as an entertainment and information generating device, there is a recognition among this age group that such devices can play an important and significant role in their lives.Figure 9 illustrates the components and functionalities included in a system labeled as ”sIPa ON” and ”Data Collection ON.” It incorporates features such as real-time emotion recognition, sentiment analysis, processing cognitive overload, and enabling casual, non-objective conversations. Together, these features support advanced user interaction and data gathering for enhanced analysis and response. In this regard, there are at least seven basic ways in which older adults might be interested in using sIPAs. However, current design of sIPA is insuficient to meet older adults' needs and expectations. Specifically, privacy considerations, interpersonal skills and contextual relevance of these devices require significant improvements to meet older adults' expectations. The design community should work closely with the developer community to implement the existing works in permission-based recording, XAI, variable speech recognition, and humanized conversations to accelerate the improvement of these devices. We are in the process of implementing a prototype of the proposed framework and evaluating it with the target users.

III. CONCLUSION

In conclusion, the survey on AI Assistant emerges as an advanced and adaptive tool, transforming human-computer interaction with a focus on security, personalization, and user convenience. Through powerful face and voice recognition features, this assistant provides secure, user-specific access while supporting multilingual interactions and task automation. Its proficiency in Natural Language Processing (NLP) allows it to understand and respond intuitively to various commands, from basic operations to more complex tasks. Driven by Python's flexibility and AI advancements, this system continuously evolves to meet modern user needs, offering a seamless, efficient experience. As a result, the assistant promises to reshape daily workflows and enhance productivity, providing a glimpse into a future where AI-integrated tools blend naturally into everyday life, prioritizing both security and personalized interaction.

REFERENCES

[1] S. L. LIXIANG L, XIAOHUI MU and H. P. (2020), “A review of face recognition technology,” *IEEE Access*, vol. 8, pp. 139110–139120, 2020.
[2] X. L. XIANG WANG, HEYU XUE1 and Q. PE, “A privacy-preserving edge computation-based face verification system for user authentication,” *IEEE Access*, vol. 7, pp. 14186–14196, 2019.
[3] S. P. J. S. H. L. GURURAJ, B. C. SOUNDARYA and F. FLAMMINI, “A comprehensive review of face recognition techniques, trends, and challenges,” *IEEE Access*, vol. 12, pp. 107903–107926, 2024.
[4] P. CHENG and U. ROEDIG, “Personal voice assistant security and privacy—a survey,” *IEEE Access*, 2021.
[5] J. H. L. H. IVA´N LO´PEZ-ESPEJO, ZHENG-HUA TAN and J. JENSEN, “Deep spoken keyword spotting: An overview,” *IEEE Access*, vol. 10, pp. 4169–4199, 2021.
[6] J. T. ANDREAS M.KLEIN, MARIA RAUSCHENBERGER and M. J. ESCALONA, “Comparing voice assistant risks and potential with technology-based users: A study from germany and spain,” *IEEE Access*, vol. 20, pp. 209688–209698, 2021.
[7] M. S. P. ANTONIO GOMEZ and S. CELEDO´N-PATTICHI, “Speaker diarization and identification from single channel classroom audio recordings using virtual microphones,” *IEEE Internet of Things Journal*, vol. 10, pp. 56256–56266, 2021.
[8] K. . P. M. A. J. Kulhalli, K. V. Sirbi, “Personal assistant with voice recognition intelligence,” *IEEE Access*, vol. 8, pp. 114822–114832, 2019.
[9] A. J. LAMYA BENADDI, CHARAF OUADDI and B. OUCHAOA, “A systematic review of chatbots: Classification, development, and their impact,” *IEEE Access*, vol. 12, pp. 78799–78810, 2024.
[10] M. U. ISLAM and B. M. CHAUDHRY, “A framework to enhance user experience of older adults with speech-based intelligent personal assistants,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 16683–16699, 2018.