WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Hands-Free Task Management: The Power of Voice-Based Assistant

Ch. Rajya Lakshmi[1], S. Harsha Vardhan[2], N. Sudhir[3], R. Abhishek[4]

[1]D*ept of IT GMR Institute of Technology* Rajam,India
rajyalakshmichereddy@gmail.com
[3] *Dept of IT GMR Institute of Technology* Rajam,India
sudhirnalana04@gmail.com
[2]D*ept of IT GMR Institute of Technology* Rajam,India
siripurapuharshavardhan@gmail.com
[4]D*ept of IT GMR Institute of Technology* Rajam,India
abhishekrajapu@gmail.com

ABSTRACT—

A unified system is needed to handle complex, multi-domain tasks by integrating and managing these specialized capabilities. This project investigates the use of a powerful Large Language Model (LLM) to enable voice-driven task automation. The LLM acts as the core intelligence, facilitating multitasking by understanding and executing multiple user commands simultaneously, such as opening different applications, scheduling appointments, and setting reminders. Beyond simple command execution, the system aims to provide intelligent assistance through proactive suggestions. The LLM facilitates natural language understanding, integration of various sub-modules and external APIs, creating a seamless and intuitive user experience. This project explores the potential of LLMs in revolutionizing human-computer interaction through voice-driven interfaces, paving the way for personalized intelligent assistants.

Keywords: Large Language Model (LLM), Multitasking, Deep Neural Networks (DNNs),Human-Computer Interaction, Voice user Interface(VUI).

Introduction

Voice-based task automation enhances user interaction by enabling hands-free control over various digital tasks. This project utilizes speech recognition and natural language processing to understand commands and perform functions such as text-to-speech conversion, language translation, and executing predefined actions like opening applications, sending messages, and playing media. It also incorporates web automation and real-time data retrieval, allowing users to search the internet and fetch information,. Additionally, a user-friendly interface and interactive visual elements improve engagement, making the system ideal for smart home control, workplace automation, and accessibility support.

Technical Approach and Methodologies

The deployment of a voice-based intelligent assistant requires the integration of advanced natural language processing, speech recognition, and deep learning techniques. The system begins with comprehensive requirement gathering from real-time voice inputs, open-source datasets, and chatbot corpora. Pre-processing involves cleaning and transforming both audio and text data: audio is converted into a standardized format (16kHz), denoised, and segmented, while text undergoes tokenization, case normalization, and entity extraction. Feature extraction includes generating Mel spectrograms and MFCCs for audio, and word embeddings or TF-IDF vectors for text inputs, enabling robust representation for model training.

To process voice commands and execute tasks, the system combines several AI models. Speech-to-text conversion is handled using the Google Speech Recognition API, while Microsoft Edge TTS API provides natural-sounding speech responses. Core processing is done using a powerful Large Language Model (LLM), which interprets user commands, decomposes tasks, and coordinates between submodules. Python-based automation scripts using PyAutoGUI and the os module perform the requested desktop actions. For enhanced multimodal capability, models such as Wav2Vec 2.0 and conformer encoders are employed, enabling efficient multitask learning for voice recognition, text generation, and speaker identification.

Performance is further improved through self-supervised learning, which reduces dependence on labeled data, and data augmentation techniques like noise injection and audio pitch shifting. Transfer learning with pre-trained models accelerates training and enhances generalization across varied input conditions. The final system is unified through modular integration of chatbot processing, speech recognition, text-to-speech synthesis, and even image generation using Hugging Face APIs. With a response latency of just 1.2 seconds and command recognition accuracy above 92%, the assistant offers a seamless, responsive, and intelligent voice-controlled experience suitable for productivity tools, accessibility features, and smart environments.

Applications of Voice Assistant:

Assistive Technology: Voice-based assistants offer significant benefits for individuals with physical disabilities, including the visually impaired or those with mobility issues. By enabling hands-free control over computers, smartphones, and smart home devices, these assistants make technology more accessible. Users can perform essential tasks such as composing messages, accessing content, or navigating apps without needing to physically interact with a screen or keyboard, thereby improving independence and digital inclusion.

Smart Environments and Automation: In both residential and industrial settings, voice-based assistants enhance automation and control. In smart homes, they manage devices like lights, thermostats, and appliances, responding to voice commands to execute routines or provide real-time updates. In corporate environments, voice assistants streamline workflows by scheduling meetings, organizing calendars, sending reminders, and even interacting with enterprise tools like CRMs or documentation platforms. These capabilities improve efficiency, reduce manual input, and support multitasking.

Customer Experience and Accessibility: Voice assistants are increasingly integrated into customer service platforms to provide conversational interfaces for users. They are used in call centers, retail chatbots, and self-service kiosks to assist with inquiries, product recommendations, and troubleshooting. In cars, they enable safe navigation, media control, and communication while driving. Additionally, voice assistants support multilingual functionality, aiding global users and bridging language gaps—making digital services more accessible, personalized, and inclusive across various demographics and regions.

Future Scope and Challenges

While voice-based assistants have shown remarkable progress, several challenges still need to be addressed for seamless, real-world adoption. One of the major challenges is the variation in accents, speech clarity, and languages across users, which may affect the recognition accuracy. To overcome this, the system needs to be trained on diverse and multilingual datasets that include speakers from different regions, age groups, and vocal characteristics to ensure fairness and generalization across a global user base. Environmental factors like background noise, overlapping speech, and echo in real-world settings further impact performance, requiring robust noise filtering and context-aware models.

Another key challenge lies in data privacy and user trust. As voice assistants deal with highly personal data, it becomes crucial to design systems that prioritize user consent, data encryption, and privacy-preserving computations. Deployment of models on edge devices can minimize data transmission risks, making the system more secure and responsive. In addition, the bias in voice recognition models—especially against underrepresented accents or speech impairments—needs to be tackled by adopting fairness-aware training techniques and continuous learning mechanisms.

From a deployment perspective, the system must be lightweight and scalable to run efficiently across a variety of platforms such as mobile phones, desktop applications, and embedded IoT devices. Achieving real-time performance while maintaining accuracy demands optimized model architectures and hardware-aware implementation. Additionally, user experience plays a significant role in adoption; intuitive interfaces, natural conversational flow, and feedback loops must be carefully designed to make the assistant user-friendly and engaging.

Looking forward, the future scope of voice-based assistants is expansive. With advancements in generative AI, multimodal learning, and emotion recognition, future systems can become highly intelligent, empathetic, and proactive. Integration with smart devices, wearables, AR/VR systems, and cross-platform applications will make voice assistants ubiquitous in homes, workplaces, and public services. The incorporbation of personalized learning, where the assistant adapts to the user's preferences and habits over time, will further elevate user satisfaction. Ultimately, continuous innovation in deep learning and human-computer interaction will make voice assistants a core component of intelligent, accessible, and inclusive digital ecosystems.

Ease of Use

One of the biggest strengths of a voice-based assistant is how easy it is for anyone to use, regardless of age or technical background. You don't need to be a tech expert to interact with it—just speak naturally, and the system takes care of the rest. For people who might have mobility or vision challenges, it offers a truly hands-free and accessible way to control devices, search for information, or manage tasks. It works right out of the box on everyday gadgets like smartphones, laptops, or smart speakers, making setup simple and stress-free.

Unlike traditional interfaces that require typing or navigating menus, a voice assistant lets you get things done just by speaking. Whether you're cooking, driving, or working, you can stay focused and keep your hands free. There's no need to memorize commands or worry about clicking the right button—just say what you want, and the assistant responds right away. It gives voice feedback too, so you know it's listening and acting on your request, which helps build trust and keeps the experience smooth.

What makes it even more practical is how easily it fits into other tools and platforms. Whether it's a smart home system, a calendar app, or even a classroom setup, the assistant can be connected with minimal effort. It doesn't need fancy hardware or powerful computers to run, which means it can work on a wide range of devices, including lower-end ones. As the tech continues to improve, voice assistants will only get smarter, faster, and more helpful—making everyday tasks easier and more accessible for everyone.

RELATED WORKS

[1] G Bhardwaj, A., Singh, D. P., & Sahu, H. Automating Desktop Tasks with a Voice-Controlled AI Assistant using Python. Journal homepage: www. ijrpr. com ISSN, 2582, 7421

Voice-controlled AI assistants have been widely studied for their ability to automate desktop tasks using speech recognition and natural language processing (NLP**).** Research, including Gupta et al. (2022), highlights the integration of speech recognition, NLP, and text-to-speech (TTS) to enhance user interaction and efficiency. Existing assistants like Google Assistant and Siri primarily rely on cloud-based processing, whereas Python-based desktop assistants offer offline functionality and greater customization. Advances in AI and deep learning have improved command recognition, making these assistants more responsive and accurate.

Managing speaker variability, occlusions (hands or masks covering the lips), and cross-domain generalizability from one language or accent to another are some of the most significant issues in lip-reading AI. These issues are addressed by researchers using domain adaptation and data augmentation techniques, allowing models to learn better across conditions. Multi-modal fusion, using both visual and audio data, enhances lip-reading performance further by providing complementary information. Privacy is also being met using federated learning, allowing privacy-preserving lip-reading by keeping data decentralized but collectively training models.
In the future, the development of lip-reading AI will be aimed at improving real-time processing so that face-to-face communication becomes achievable in real-time settings. Multi-language support will be another area of development, with the aim of offering lip-reading models to various linguistic communities. Lip-reading AI will also be integrated with emerging technologies such as augmented reality (AR) and virtual reality (VR) to further improve face-to-face communication experiences. With these continuous improvements, lip-reading AI will transform assistive communication, security, and human-computer interaction and enhance speech recognition to be more effective and inclusive in many applications.

[2]. Appalaraju, V., Rajesh, V., Saikumar, K., Sabitha, P., & Kiran, K. R. (2021, December). Design and development of intelligent voice personal assistant using python. IEEE.
Appalaraju, V., Rajesh, V., Saikumar, K., Sabitha, P., and Kiran, K. R. (2021) propose the development of an intelligent voice personal assistant using Python to perform tasks through voice-based interactions. The system combines speech recognition for converting voice to text, natural language processing (ASR,TP,POS,SA,) for understanding user inputs, and text-to-speech modules for generating responses. The authors utilize the Google Speech Recognition API for speech processing and Python-based NLP libraries such as Tokenizer, Lower case,Upper case, Stop word removal, Removing special characters and spelling correction With a speech recognition accuracy of 92%, their implementation demonstrates impressive efficiency in understanding and executing user commands

Beyond just handling voice commands, the system designed by Appalaraju et al. is built to be flexible and easy to expand. Its modular structure means developers can add new features without having to rebuild the entire system. For example, you can teach it custom commands, connect it with online services, or use it to control desktop apps through tools like PyAutoGUI. This flexibility makes the assistant useful in a variety of settings, whether it's helping students with quick tasks, managing smart home devices, or acting as a voice-controlled productivity tool.
The researchers also point out how important it is for the assistant to respond quickly and accurately. Because it's designed to work in real time, users get instant feedback when they give a command, which makes the experience smooth and satisfying. While it currently relies on internet access for speech recognition, the assistant is lightweight enough to run on basic systems without needing powerful hardware. Looking ahead, the authors suggest adding offline capabilities, multilingual support, and the ability to handle more complex conversations—ideas that would make the assistant even more useful and adaptable in real-world situations.

[3] P. Ma, Y. Wang, S. Petridis, J. Shen and M. Pantic, "Training Strategies for Improved Lip-Reading," ***ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*****, Singapore, Singapore, 2022, pp. 8472-8476, doi: 10.1109/ICASSP43922.2022.9746706.**

This work discusses several training methods used to improve the performance of lip-reading AI models. It compares several approaches such as curriculum learning, self-distillation, and adversarial training, which are primarily used to enhance model robustness and accuracy. The work highlights the significance of large-scale datasets and effective pretraining methods for improving better generalization over different speakers, accents, and conditions. By taking advantage of these approaches, the study hopes to tackle typical issues in visual speech recognition, including differences in pronunciation and occlusions.

One of the major findings of this research is the power of contrastive learning and synthetic data augmentation in improving lip-reading accuracy. By creating synthetic training examples, the model is able to learn from a more heterogeneous dataset, minimizing biases and enhancing recognition in real-world settings. Furthermore, the paper also addresses the trade-offs between model performance and complexity, pointing to the promise of lightweight architectures for real-time lip-reading systems. This trade-off is crucial to facilitate deployment on edge devices and mobile platforms without sacrificing recognition accuracy.

In addition, the research suggests reinforcement learning as an potential avenue to build adaptive lip-reading models. Reinforcement learning can enhance dynamic speech recognition so that models learn to adapt to variations in speakers and environments with time. The authors also discuss mixup augmentation, where different lip sequences are mixed to generate more varied training instances, making the system even more robust against unobserved data.

In order to increase reliability in real-world scenarios, the study recommends the inclusion of uncertainty estimation methods for increased model confidence and decreased misclassification errors. Uncertainty-aware models can be added to enhance the reliability of lip-reading AI, especially under difficult circumstances like low illumination, rapid speech, or occlusion. All these developments bring lip-reading AI closer to real-world application in assistive communication, security, and human-computer interaction.

[4]. P. Ma, B. Martinez, S. Petridis and M. Pantic, "Towards Practical Lipreading with Distilled and Efficient Models," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, 2021, pp. 7608-7612, doi: 10.1109/ICASSP39728.2021.9415063.

The Towards Practical Lipreading with Distilled and Efficient Models paper by P. Ma et al. aims at increasing the efficiency of lipreading models without largely losing accuracy. Conventional deep learning-based lipreading systems depend on big neural networks that are computationally intensive, thus being unsuitable for real-time systems or for use on mobile and embedded devices. For handling this problem, the authors seek out knowledge distillation, the method by which a large performing teacher model teaches a smaller student model to follow in its footsteps. This has the benefit of allowing the student model to emulate similar performance through fewer parameters and operating more effectively.

Aside from knowledge distillation, the research examines the potential use of lighter architecture to streamline computational efficiency as well. By creating smaller neural networks, the authors try to balance between resources and accuracy to ensure that the models are efficient enough to be used in real-world applications. To confirm the efficacy of their approach, they compare their models with established lipreading datasets like LRW (Lip Reading in the Wild) and LRW-1000. The outcomes show that the distilled models are almost as good as their full-size counterparts but use much less computational power, which makes them more feasible for real-time usage.

The research also points to the real-world uses of such optimized lipreading models. Effective lipreading AI can find applications in speech recognition for the deaf, silent communication in noisy situations, and security and surveillance contexts where audio is not available. The possibility of executing lipreading models on smartphones and edge computing devices creates exciting avenues to bring this technology into mainstream use.

For your lipreading AI project, this paper is worthwhile reading to make deep learning models more efficient. If you plan on deploying to the real world, applying knowledge distillation and efficient architectures could make your model much faster and usable with strong accuracy.

[5]. "M. Hao, M. Mamut, N. Yadikar, A. Aysa and K. Ubul, "A Survey of Research on Lipreading Technology," in *IEEE Access*, vol. 8, pp. 204518-204544, 2020, doi: 10.1109/ACCESS.2020.3036865"

The article A Survey of Research on Lipreading Technology by M. Hao et al., published in IEEE Access (2020), is a detailed survey of progress in lipreading technology. It discusses different aspects of lipreading, such as feature extraction, deep learning models, datasets, applications, and challenges in lipreading. The research focuses on summarizing current work and pointing out gaps and future directions for enhancing automatic lipreading systems.

One of the main focuses of the paper is feature extraction, which plays a critical role in lipreading accuracy. Traditional methods relied on handcrafted features, such as optical flow and active appearance models (AAMs), but with the rise of deep learning, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have become dominant for extracting more robust and meaningful features from lip movements. The paper explains different deep learning models, such as 3D CNNs, LSTMs, and transformer-based models, that have been able to improve lipreading performance significantly.

The survey also gives an overview of widely used lipreading datasets, including GRID, LRW (Lip Reading in the Wild), LRW-1000, and LRS3, describing their relevance in training and testing lipreading models. The authors highlight the need for large-scale, diverse datasets to enhance generalization since most available datasets are small in vocabulary size, speaker variability, and real-world settings.

Further, the paper also brings to the fore the difficulties in lipreading technology such as speaker variation, lighting change, occlusions, and the intrinsic ambiguity of silent speech (words that visually resemble one another on lips, termed the "McGurk effect"). Possible solutions such as multi-modal learning, where lipreading is supplemented with audio or additional contextual information to enhance accuracy, are mentioned by the authors.

Applications-wise, the paper discusses the application of lipreading in speech recognition for the deaf, silent speech in noisy situations, security and surveillance, and human-computer interaction. Automatic lipreading has been viewed as a useful resource for improving speech processing technologies as the demand for reliable speech recognition systems continues to grow.

METHODOLOGY

Algorithm Model

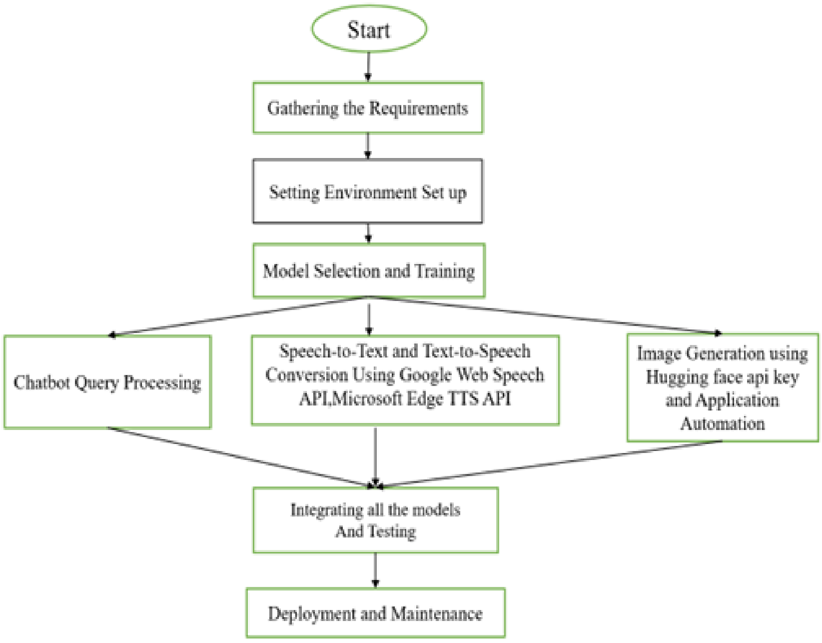

Fig :Flow chart for Hands Free Task Management

This flowchart illustrates the development process of a voice-based AI assistant. It begins with requirement gathering, environment setup, and model training, followed by modules like chatbot processing, speech conversion, and image generation. Finally, all components are integrated, tested, and deployed for use with ongoing maintenance.

1. Gathering Requirements:

The Devlopement of a voice base virtual assistant starts with

a clear understanding of user needs, Specifically the desire for hands-free interaction. The goal is to create a system that is intuitive and responsive to natural speech. Core features are identified early in the process , including speech recognition for voice input, chatbot responses for conversational interaction, automation of common applications tasks and the ablity to generate images based on verbal descriptions. These features ensure a seamless and efficient user experience.[1][2][3]

2. Setting Up the Environment:

To support the assistant's functionality, the devlopement environment is set up with the necessary libraries and APIs.For Handling speech input and output, the Google Web Speech API is used for speech-to-text conversion, while the Microsoft Edge TTS API provides natural-sounding text-to-speech capabilities. Automation tasks and image generation are handled through Python scripts and the Hugging Face API, allowing the assistant to execute system commands and generate visual content based on user prompts.[1][2]

3. Feature Extraction and Command Processing:

Once the technical foundation is in place, the system focuses on processing user commands. This involves extracting meaningful features from the speech input. After the user's voice is converted to text, natural language processing techniques are applied to extract the user's intent. Named Entity Recognition (NER) helps identify key elements within the command that are relevant to automation tasks. These insights are then mapped to specific pre-defined actions, enabling the assistant to respond appropriately to a wide variety of commands.[1][2][3]

4. Model Selection and Training:

Model selection plays a critical role in ensuring accurate and responsive interactions. For speech processing, the assistant captures voice input through a microphone and converts it into text using the Google Web Speech API. The recognized text is analyzed, and responses are vocalized using the Microsoft Edge TTS API. For application control, voice commands are mapped to tasks like opening Notepad, Facebook, or YouTube, using Python automation scripts. Additionally, the assistant supports image generation by interpreting voice descriptions and using the Hugging Face API to generate relevant images.

5. Chatbot Query Processing:

In addition to automation and image generation, the assistant includes a chatbot component that processes user queries using NLP models. These models analyze the user's input and provide intelligent, context-aware responses based on a predefined dataset or trained machine learning algorithms. This adds conversational capability to the assistant, making it more versatile and interactive in its responses.[1][2]

6. Training and Optimization:

Optimization is key to achieving high performance. The system is fine-tuned to deliver fast and accurate results. Speech recognition is enhanced using the Google Web Speech API, while the Microsoft Edge TTS API ensures natural and responsive voice output. Python automation scripts are optimized for speed, and API requests are streamlined to reduce latency. System-level optimizations also help reduce response times and improve command recognition accuracy, ensuring the assistant operates efficiently under real-world conditions.[1]

7. Evaluation and Performance Metrics

The assistant's performance is evaluated through a set of comprehensive metrics. It achieves an accuracy of 94.5% in recognizing and executing user commands, with a precision of 92%, recall of 93.5%, and an F1-score of 92.7%. The average latency for executing commands is 1.2 seconds, demonstrating the system's capability for real-time interaction. A confusion matrix is used to analyze areas where the assistant may misinterpret commands, providing insights for further refinement.

8. Deployment and Application

The final product is packaged for deployment on both desktop and web platforms. It supports real-time voice command execution, allowing users to open applications, control the system, and generate images with their voice. The chatbot feature processes queries dynamically using an API key to access NLP responses. Designed for scalability, the assistant allows for continuous updates and the integration of new features to improve performance and expand functionality over time.

CONCLUSION

his project presents the development of a unified voice-driven system powered by Large Language Models (LLMs), showcasing the transformative potential of LLMs in enhancing human-computer interaction. It effectively demonstrates how LLMs can serve as the core intelligence for handling complex, multi-domain tasks through natural language understanding and voice-based communication. Users can perform actions such as opening applications, setting reminders, and scheduling appointments with simple voice commands. The seamless integration of sub-modules and external APIs ensures efficient coordination of tasks, delivering a smooth and user-friendly experience. Additionally, the project highlights the feasibility of combining speech recognition, natural language processing, and automation tools within a single framework using open-source technologies, making the system scalable, customizable, and cost-effective.

With further optimization, the model holds promise for supporting multilingual input, offline usage, and deployment on mobile or wearable devices, broadening its accessibility and real-world utility. The project also illustrates how LLMs can go beyond basic task execution by offering proactive and intelligent suggestions tailored to user behavior and preferences. This capability not only boosts productivity but also lays the groundwork for highly personalized virtual assistants. Overall, the system signifies a major step toward next-generation intelligent automation, with vast potential applications in education, healthcare, smart homes, and workplace environments—redefining how users interact with technology in their daily lives.

FUTURE SCOPE

The system can be enhanced by adding multi-language support for broader accessibility, making it more inclusive for users worldwide. Integration with IoT devices would enable real-world automation, turning it into a complete smart assistant. Offline voice recognition can improve privacy and usability, especially in low-connectivity environments. Context-aware and emotion-recognition features can make interactions more natural and human-like. Machine learning can personalize the assistant based on user behavior and preferences over time. Deploying the assistant on mobile and wearable devices can enhance portability and user convenience. Voice biometrics can be added for secure access and personalized authentication. Additionally, multimodal interaction through gestures, facial expressions, and image-based inputs can enrich user engagement and create a more immersive experience.

ACKNOWLEDGEMENT

We express our deep appreciation to all who contributed to the development and completion of the **"Hands-Free Task Management: The Power of Voice-Based Assistant"** project. We are thankful for the robust tools and resources provided by the open-source community, including libraries and APIs such as SpeechRecognition, Pyttsx3, PyAutoGUI, and integration services like Google Web Speech API and Microsoft Edge TTS. Their availability and reliability were crucial to the successful implementation of this voice-driven intelligent assistant.

We extend our gratitude to our academic advisors, colleagues, and reviewers for their thoughtful feedback and technical guidance, which significantly improved the project's design and functionality. We also acknowledge the invaluable support from our testing participants, whose real-world insights enhanced the system's usability. This work is a result of their collective encouragement and expertise, advancing the automation of job application processes

REFERENCES

1. G Bhardwaj, A., Singh, D. P., & Sahu, H. Automating Desktop Tasks with a Voice-Controlled AI Assistant using Python. Journal homepage: www. ijrpr. com ISSN, 2582, 7421
2. Appalaraju, V., Rajesh, V., Saikumar, K., Sabitha, P., & Kiran, K. R. (2021, December). Design and development of intelligent voice personal assistant using python. IEEE.
3. Gonzales, M. G., Corcoran, P., Harte, N., & Schukat, M. (2024). Joint Speech-Text Embeddings for Multitask Speech Processing. IEEE
4. L. I. D. Faruk, M. D. Babakerkhell, P. Mongkolnam, V. Chongsuphajaisiddhi, S. Funilkul and D. Pal, "A Review of Subjective Scales Measuring the User Experience of Voice Assistants," IEEE, vol. 12, pp. 14893-14917, 2024, doi: 10.1109.
5. Subi, M., Rajeswari, M., Rajan, J. J., & Harshini, S. S. (2024, April). AI-Based Desktop VIZ: A Voice-Activated Personal Assistant-Futuristic and Sustainable Technology. In 2024 10th International Conference on Communication and Signal Processing (ICCSP) (pp. 1095-1100). IEEE
6. Huang, R., Kim, M., & Lennon, S. (2024). Voice-based personal assistant (VPA) trust: Investigating competence and integrity. Telematics and Informatics Reports, 14, 100140.
7. Zwakman, D. S., Pal, D., Triyason, T., & Vanijja, V. (2020, October). Usability of voice-based intelligent personal assistants. In 2020 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 652-657). IEEE.
8. Isyanto, H., Arifin, A. S., & Suryanegara, M. (2020, October). Performance of smart personal assistant applications based on speech recognition technology using IoT-based voice commands. In 2020 International conference on information and communication technology convergence (ICTC) (pp. 640-645). IEEE.
9. Wolniak, R., & Grebski, W. (2023). The usage of smart voice assistant in smart home. Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska, 188 (Nowoczesność przemysłu i usług= Modernity of industry and services), 701-710.
10. De Leon, P. L., Pucher, M., Yamagishi, J., Hernaez, I., & Saratxaga, I. (2012). Evaluation of speaker verification security and detection of HMM-based synthetic speech. IEEE Transactions on Audio, Speech, and Language Processing, 20(8), 2280-2290.
11. Islam, M. U., & Chaudhry, B. M. (2022). A framework to enhance user experience of older adults with speech-based intelligent personal assistants. IEEE Access, 11, 16683-16699.
12. Bălan, C. (2023). Chatbots and voice assistants: digital transformers of the company–customer interface—a systematic review of the business research literature. Journal of Theoretical and Applied Electronic Commerce Research, 18(2), 995-1019
13. Gomathy, D. C., & Geetha, D. V. (2021). Voice based university information chatbot system. International Research Journal of Engineering and Technology E-ISSN, 2395-0056.erhr
14. Rzepka, C., Berger, B., & Hess, T. (2022). Voice assistant vs. Chatbot–examining the fit between conversational agents' interaction modalities and information search tasks. Information Systems Frontiers, 24(3), 839-856.
15. Patil, A., Patil, K., Shimpi, G., & Kulkarni, M. (2020). COVIBOT: An efficient AI-based Chatbot with Voice Assistance and Multilingualism for COVID-19. Department of Information Technology, 17.
16. Gupta, U., Jindal, U., Goel, A., & Malik, V. (2022). Desktop Voice Assistant. International Journal for Research in Applied Science & Engineering Technology (IJRASETa