A Survey Paper On: "JARVIS – Artificial Intelligence Based Voice Assistant"

Prof. Shah .S.N.[1] , Salunkhe Sakshi Rajandra[2], Jadhav Avantika Rahul[3], Salunke Swaraj Nitin[4],
[1]*Head of Department of Computer Engineering*
[2,3,4]*UG Students*
Sharadchandra Pawar College of Engineering and Technology, Someshwarnagar, Pune

***Abstract*—The Jarvis Project is a state-of-the-art voice assistance system based on Android that presents a new dimension to user experience through intelligent voice applications. With smooth interaction with technology being the requirement of the day, Jarvis seeks to facilitate an intuitive and secure interface for the maximum efficiency of control over day-to-day class activities and communications. The system's core capabilities consist of three key features: a voice-activated password to allow for secure login, a content information download facility, and full automation of WhatsApp. Users can initiate voice message through WhatsApp with voice commands. The application also provides chat assistance for various topics that include general knowledge and specialized subjects.**

***Index Terms*—Speech Recognition, Machine Learning, Speech-to-Text, Text to-Speech, Artificial intelligence.**

I. INTRODUCTION

The above Software Requirements Specification (SRS) document presents software requirements for the Jarvis project, which is an Android-based voice assistance system. It should act like a blueprint for the design, development, and testing phases of the project so that all requirements are captured in every detail and must be understood and agreed upon by all stakeholders. To provide a voice-assisted experience that is sturdy, user-friendly, and secure. The system will facilitate voice-password authentication, content information downloads, WhatsApp automation, and a chat assistant for conversational purposes. The project aims to develop a complete voice solution, fully user-oriented, reasoning applications can enable users to interact with their mobile devices in a safe and efficient manner.

This SRS Document contains detailed specifications on the functional and non-functional requirements of the system being studied, covering technical and user-related aspects of the application. It also specifies the boundaries of the system, different use cases, and what users are expected to experience. The people responsible for the development, testing, and the other stakeholders would, therefore, refer to the SRS Document to assess whether these aforementioned specifications and objectives can be placed on the system. The Jarvis Project is to be developed with Android technologies while harnessing advancements in natural language processing (NLP) so as to guarantee uninterrupted relations with its users. The voice recognition system is to avail itself of sound biometric recognition techniques in securing access into the application. This chapterseeks to serve the dual purpose of giving the contextual introduction and the general aim of the software requirements specification for the Jarvis project.

II. METHODOLOGY

The system architecture for the smart glasses is designed to support real-time object detection and audio feedback for visually impaired users. The ESP32 microcontroller serves as the central processing unit, interfacing with multiple components including a camera module for image capture, ultrasonic sensors for distance measurement, and an audio output module for user feedback. Creating an assistant like JARVIS is a very technical, many-faced process involving speech recognition, NLP, Voice Authentication and Security, and control of systems. The method described below explains the stepwise procedure for the construction of the A.I. assistant.

[1] Voice Authentication and Security

Voice authentication and security serve primarily to make sure that only verified users can gain entrance to a system while keeping sensitive information away from any unauthorized individual accessing it. Voice authentication, which makes use of a person's vocal

traits, such as pitch, tone, and the pattern in which they speak, is necessary in isolating a speaker's identification No doubt, the advanced speech recognition systems assure the manipulation of commands with high precision. Currently, the modern voice assistants have also encompassed their security with the anti-spoofing, like liveness detection, which is based on deep learning methods to tell a real human voice from the recorded voices.

The most important parameter in which enumerated use are AES-256 and the TLS/SSL example for security of voice data at creation and transfer. There is also the option of on-device processing so as to avoid dependency on the cloud, which may result in loss of data. Some users would be restricted while others access the features wholly through role-based access.

[2]Google Text-to-Sprrch (GTTS)

Google Text-to-Speech (GTTS) is a Python library that converts text into spoken words. After generating the summary, GTTS converts the extracted text into speech and saves it as an MP3 file. This feature improves accessibility by enabling users, including those with visual impairments, to listen to the generated summary instead of reading it. Moreover, there are several languages as well as adjustable speech speed which GTTS offers to cater for more user satisfaction to a wider audience. This completely systematic methodology guarantees that our summarization system saves large text data along with a standard interactive audio output effective for user accessibility.

[3]NaturalLaungage Processing (NLP)

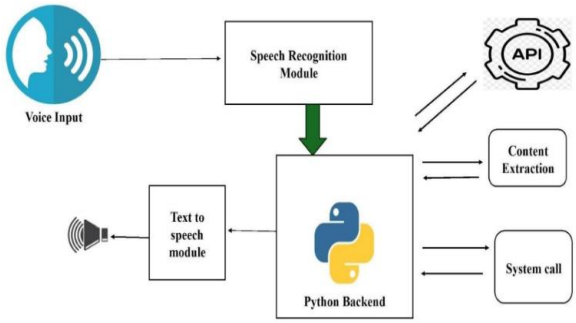
Natural Language Processing (NLP) is the central technology that enables any conversational system to empower machines to understand, comprehend, and communicate fully in human language. An NLP enabled conversational system incorporates several components: the ambience and noise, speech recognition, intent detection, context awareness, and answer generation. Thus, the usual course is to transcribe the spoken or written input into structured data by the speech-totext processes such as Whisper, Google Speech API, or DeepSpeech. Then Natural Language Understanding (NLU) analysis is performed on the corpus to extract the intent and relevant key entities through methods like Named Entity Recognition (NER), dependency parsing, and semantic analysis.

[4]WhatApp Automation and Integration

Automation and integration in WhatsApp offer a wide range of solutions that create tactful avenues for improving customer experience. It helps in creating better interaction for effective communication, which can lead to much better output for the operations of the business. Automated messages can be programmed as chatbots or integrated into the WhatsApp Business API. Companies can now automate what they previously did manually such as notifications, inquiries, and answers, or they can hire a WhatsApp agent, including installation into CRM ecommerce platforms or customer support tools. Automation takes the form of AI-based chatbots, bulk messaging, order processing, payment processing, and instant notifications.

III. SYSTEM ARCHITECTURE

Proposed framework for JARVIS project: follows a highly structured and systematic course of development for the voice assistant application.



IV.LITERATURE SUEVEY

[1] The paper is titled 'Biometrics: Personal Identification in Networked Society' (A.K. Jain, R.C. Dubey, and S.R.S. P. Misra). It sets out different complementary techniques of biometric authentication focusing on voice recognition systems that usually analyze unique vocal characteristics for secure access. The authors describe how voice recognition can be an alternative to existing traditional password uses with an added level of security. Insights on feasibility regarding using voice passwords in mobile applications are offered in the paper, perfectly matching the authentication utility in the Jarvis Project.

[2] The researchers summarize deep learning used for Natural Language Processing (NLP) concerning techniques to improve language understanding in

conversational systems. They discuss several deep learning models, including RNNs, LSTMs, and transformers, that can be used to enhance the NLP capabilities of spoken dialogue systems. The results of this paper would benefit the Jarvis Project very much concerning its ability to process and understand user commands in an intuitive conversation.

[3] Paper Title: A Survey of Chatbot Architectures and Technologies Authors: C. Jin, X. Zhang, and J. Zhao The paper gives in-depth descriptions of chatbot architectures and technologies, particularly as used in customer-service applications and virtual assistants. The authors discuss some critical systems for developing chatbots: rule-based systems, retrieval-based systems, and various generative models, outlining advantages and disadvantages for each. The discussion is essential for Jarvis Project considerations toward designing and deploying an effective conversational agent with whom users may engage in free-flowing, natural conversation.

[4] This research article, named Voice Interaction within Mobile Application Developing: Challenges and Best Practices, was authored A. Gabbur and D. G. Shenoy. It investigates the challenges imposed by the incorporation of voice interaction in mobile applications. Some design considerations dealt with are visualization of feedback, clear articulation of instructions, and handling error messages. These findings would be relevant to the Jarvis Project, which tries to incorporate voice interaction into the developing process through a sound, user-friendly and integrated approach.

[5] Paper Name: Automation of WhatsApp for Improved Communication in Android Applications Authors: S. Soni, R. Shukla and P. Verma The authors demonstrate that using WhatsApp Business API can help automate the communication tasks in applications built using Android environment. They also describe the aspect of how the incorporation of voices with the APIs of WhatsApp can make it possible for a user to perform complex tasks such as voice calls and video calls through voice commands. It contributes directly towards the automation features associated with Jarvis Project.

IV. CONCLUSION

This goes to show what might just become of advanced voice assistance integrated into a mobile app. The Jarvis Project is meant to meet the requirements of a user for a hands-free, efficient, and secure digital assistant. It aims at using voice commands for daily interaction purposes with mobile devices. The project was undertaken mainly with an emphasis on voice-based authentication, contents retrieval, WhatsApp API integration, and a conversational NLP-based chatbot which were critical aspects focused at different stages of development and implementation. These features allow users to make calls, send messages, gather information, and have conversations with little objection.

REFERENCES

[1] M. R. Zunzunwala, N. A. Vhora, N. A. Ansari, and S. A. Shaikh, "Intelligent Personal Assistant with Chatbot and Voicebot Using NLP Techniques," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 6, pp. 159-167, June 2020

[2] U. Ashfaq and F. Shaheen, "A Review of Machine Learning Techniques for Dialogue Management in Personal Assistant Systems," International Research Journal of Modernization in Engineering Technology and Science (IRJMETS), vol. 3, no. 6, pp. 209-212, June 2021.

[3] Vedant Kulkarni, Shreyas Kallurkar, Vipul Waikar, Saurabh Patil, - “Virtual Assistant Using Python”, Journal Of Emerging Technologies and Innovative Research, 2022, ISSN-2349-5162.

[4] Ujjwal Gupta, Utkarsh Jindal, Apurv Goel, Vaishali Malik – “Desktop Voice Assistant”, International Journal for Research in Applied Science & Engineering Technology (IJRASET).