Generative Language Model Technology Integrated into an IoT Device for the Development of a Voice Assistant

RICARDO YAURI[1,2], RAFAEL ESPINO[2]
[1]Facultad de Ingeniería Electrónica y Eléctrica,
Universidad Nacional Mayor de San Marcos,
Lima,
PERÚ

[2]Facultad de Ingeniería,
Universidad Tecnológica del Perú,
Lima,
PERÚ

Abstract: - The integration of artificial intelligence technologies into IoT devices has opened new possibilities for interaction with the environment through voice assistants, such as ChatGPT, improving interaction with smart devices in sectors such as home, health, and education. However, the adoption of these technologies faces challenges due to device heterogeneity, the need for interoperability, and concerns about data privacy and security. The objective of this research is to develop an IoT device that integrates artificial intelligence technologies and generative language models for a voice assistant, covering the design of a voice recognition system, the implementation of efficient communication with the model, the coordination between ESP32 microcontrollers and the integration of a voice synthesis system. The results show that the system can send queries to ChatGPT and receive responses in real time, validating its ability to handle natural language processing. Furthermore, speech synthesis, using Audio.h library and the MAX98357 module, have demonstrated effective text-to-audio conversion, while the integration of the INMP441 microphone and the Google Cloud Speech-to-Text platform ensures voice capture and processing. In conclusion, the operation of the IoT device and its real-time interaction with the ChatGPT API were validated to obtain an efficient text-to-speech conversion, being scalable for future improvements.

Key-Words: - Generative language model, artificial intelligence, ChatGPT, Internet of Things, voice assistant, LLM, ESP32.

Received: June 9, 2024. Revised: November 26, 2024. Accepted: December 26, 2024. Published: December 31, 2024.

1 Introduction

The integration of artificial intelligence (AI) technologies for use in IoT devices has generated new services to interact with the environment in which people live, one of the specific cases being the implementation of voice assistants, [1], [2] environmental monitoring or control, [3], [4]. In addition, currently, the use of generative language models allows a more natural and efficient interaction with hardware devices, facilitating tasks such as smart home control, and sectors such as health and education, [5], [6]. This integration of technologies contributes to the accessibility of information or data, allowing people to communicate with devices through voice commands, being a tool with the potential to automate processes and improve efficiency in industrial and commercial environments, [7], [8].

Despite the advantages of integrating artificial intelligence technologies into IoT devices, [9], there is no mass adoption in different areas of the industry, [10], [11] and environmental care, [12], [13]. One of the problems is the ability to interoperate due to the heterogeneity of devices and platforms. In addition, its use in critical processes related to safety or health generates challenges regarding privacy and data security, [14], [15]. These drawbacks highlight the need to continue developing research in the field of generative models, [16], integrated into daily life activities as a contribution to process automation, [17], [18].

2 Literature Review

In the review of the state of the art, studies have been identified that describe the development of technological solutions that integrate artificial intelligence and the Internet of Things, [19] and how they have transformed various industries, from home automation, advanced manufacturing and healthcare, [20], [21]. In addition, these solutions combine the ability of IoT devices to collect and transmit data to artificial intelligence services to be analyzed in real-time, [22], [23]. However, it is also mentioned that there are technical and security challenges that still need to be overcome to increase and ensure their reliable integration in critical applications, [24], [25].

Studies describe that the integration of generative language models (such as ChatGPT and Gemini) with IoT devices contributes to interact with processes and physical elements in multiple contexts, [26]. These advanced models allow IoT devices to collect and transmit data, understand and generate responses to complex queries in natural language, automating repetitive tasks, [27], [28]. Globally, innovative applications have been developed in areas such as smart home control, remote healthcare, and supply chain management, where the ability to process and respond to data in real time is critical, [29]. On the other hand, it is mentioned that the increasing availability of AI platforms, such as generative language models, together with the advancement in wireless communication technologies and cloud processing, has facilitated the creation of such systems, [30], [31].

In the case of voice assistants, research describes their presence in various areas such as medicine, education, and home automation. In medicine, these assistants facilitate patient care by providing quick access to information, [32]. In education, they are used to personalize teaching, providing instant answers to students' questions, [33], [34]. Home automation solutions are integrated in the home, where voice assistants allow remote control of devices, optimizing energy consumption, [35], [36].

Therefore, the following research question is raised: How can artificial intelligence technologies and generative language models be integrated into an embedded device for a voice assistant? To answer the question, the objective is to develop an IoT device that uses generative language, and techniques for accessing Web services and transforming text signals into speech and vice versa. To do this, it is necessary to carry out the following activities: design the voice recognition system, implement efficient communication with a generative language model, coordinate communication between the two ESP32 microcontrollers, integrate a voice synthesis system for the generation of auditory responses, and validate the integration of artificial intelligence technologies in IoT devices.

Regarding motivation, it is observed that there is a growing demand for smart devices in the Internet of Things (IoT) field, such as voice assistants, but the integration of generative language models in low-cost platforms, such as ESP32 microcontrollers, presents technical challenges due to processing and memory limitations. The main motivation of this research lies in overcoming these obstacles and demonstrating that it is possible to implement an effective voice assistant using limited resources. This research has value in that it provides a framework for the integration of generative language models in IoT environments, contributing to the existing knowledge in embedded systems design and human-computer interaction. The paper is organized into five sections: section 1 presents the introduction, section 2 with the methodological process, and section 3 shows the results. The paper ends with section 4 with discussions and section 5 with conclusions.

3 Methodology

This section shows the development of the IoT-based voice assistant system, composed of two ESP32 microcontrollers that operate together to provide a fluid and efficient experience. The first microcontroller is responsible for capturing the user's voice signal, converting the audio to text using a voice recognition service, and then sending the query to the generative language model, such as ChatGPT. Once the response is received, it is transmitted to the second microcontroller, whose function is to analyze the received text to detect possible orders or convert it into audio output using a voice synthesis system. (Figure 1).

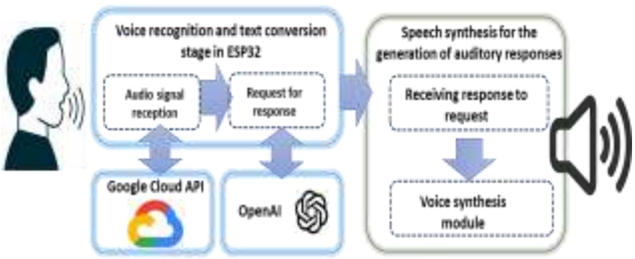

Fig. 1: Scheme for the system proposal

3.1 Architecture

The system architecture consists of two modules. The first ESP32 module is connected to an INMP441 microphone, which is a low-power device that provides high audio quality and energy efficiency and is responsible for capturing and converting the audio signal into text using the Google Cloud Speech-to-Text API and then querying the ChatGPT OpenIA API. This response is sent via serial communication to the second ESP32 which converts it into audio using a 2W 8-ohm speaker connected to a MAX98357 amplifier (Low power amplifier for audio applications in embedded systems), controlled via the I2S protocol (Inter-IC Sound) which is a digital interface standard used for the transmission of audio signals between electronic components). The integration of libraries such as Audio.h (used for text-to-speech synthesis in hardware systems) and CloudSpeechClient optimizes text-to-speech conversion and transcription handling (Figure 2).

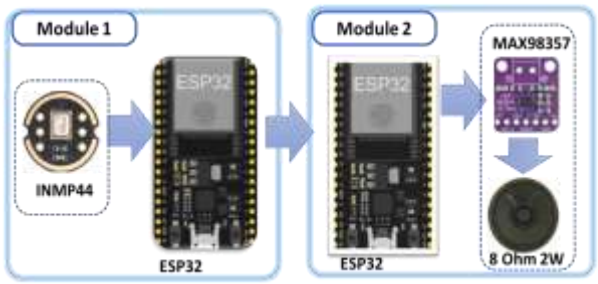

Fig. 2: System architecture

3.2 Communication with a Generative Language Model

The first stage of development focuses on implementing communication with a generative language model. In this phase, the connection between the ESP32-based IoT device and the advanced language model, such as ChatGPT, is established. This involves configuring the language model API to allow queries to be transmitted from the ESP32 and responses in real time (Figure 3).

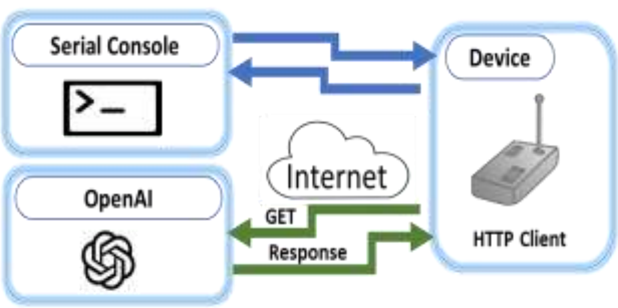

Fig. 3: System architecture Communicating with a generative language model

First, log in to the ChatGPT account and obtain the API Key from the control panel. In the upper right corner, select “View API Keys”. Then choose the option “Create the new secret key”, which will be copied to the program on the ESP32. It should be noted that the format of the Web request must follow what is indicated by the body format in the “raw” data type (Figure 4).

The algorithm establishes a Wi-Fi connection and sends a POST request to the ChatGPT API to get a response based on a query text (Figure 5). The WiFi.h and HTTPClient.h libraries are used to connect the ESP32 to a Wi-Fi network and manage HTTP communication. The request is generated using the serial console and the POST request is created in JSON format that includes the language model, the query text, and additional parameters (such as temperature and maximum number of tokens). The API response is printed to the ESP32 serial port.

Fig. 4: Access to the OpenAI API Key

The request to the ChatGPT API is sent in JSON format using the POST method. The fields in Table 1 are considered, where the model parameter specifies the language model to be used, the prompt is the query or text input that is sent to the model, temperature adjusts the creativity of the response (0 means more predictable responses) and max_tokens limits the length of the response generated in tokens.

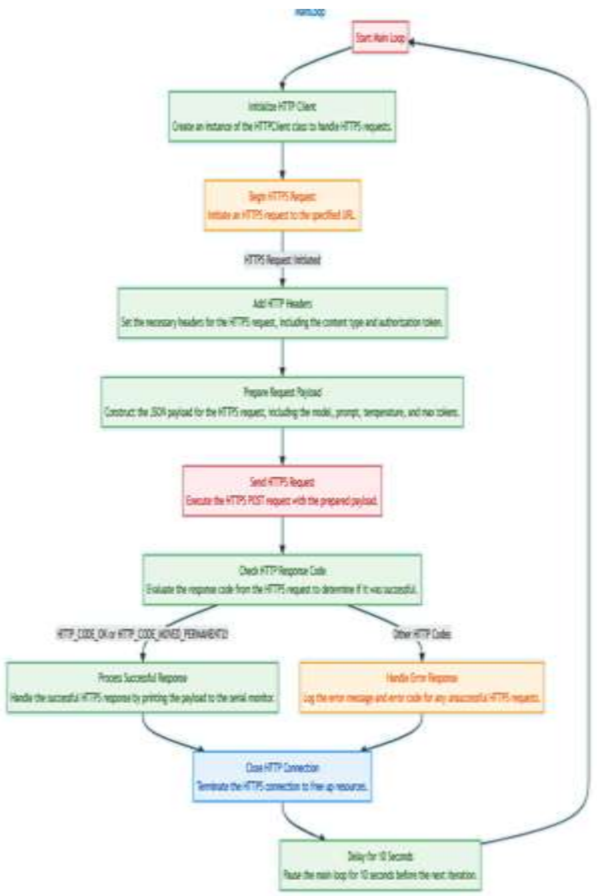

Fig. 5: Communication Algorithm with a generative language model

Table 1. Format for the request for response

Server	Subdomain: api Main Domain: openai Top-Level domain: com
Resource path	/v1/completions
Headerss:	Content-Type: application/json (Indicates that the request body is in JSON format). Authorization: Bearer YOUR_API_KEY (Authentication with API Token).
Payload:	{ "model": "gpt-3.5-turbo-instruct", "prompt": "Who are you", "temperature": 0, "max_tokens": 7 }

3.3 Integration of the Voice Synthesis Stage for the Generation of Auditory Responses

At this stage, the text generated by the language model is converted into audio played through a speaker. This is done by integrating the Audio.h voice synthesis library into the ESP32 microcontroller along with the query process to the ChatGPT API (Figure 6). To implement this stage, the following steps were taken: Integration of the voice synthesis library and implementation of instructions to send text and play the audio on the ESP32.

To perform the generation of auditory responses on the IoT device, a MAX98357 audio amplifier module, and a 2W, 8ohm speaker are integrated. The amplifier is connected to pins 25 (DIN module to I2S_DOUT ESP32), 26(I2S_LRC), and 27 (I2S_BCLK) for communication using the I2S protocol (Figure 7).

The algorithm for connecting to a Wi-Fi network for communication with the OpenAI API was implemented. In addition, the audio player module was configured by setting the volume to 100%. As a first step, the queries to be made are captured by telling the user to enter a question through the serial monitor, removing the last line break character. Then the module communicates with the OpenAI API by sending a POST request to obtain the response (using the text-davinci-003 model). The response is received and analyzed using the ArduinoJson library to extract the text generated by the model.

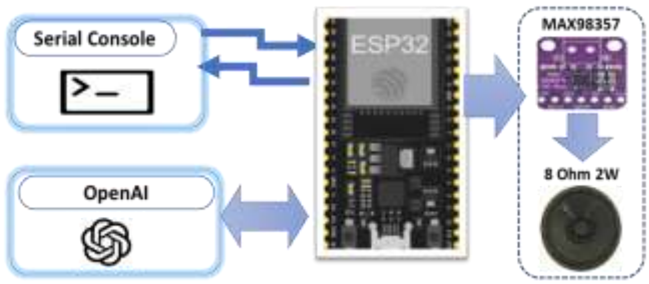

Fig. 6: Scheme of the Integration of the Voice Synthesis Stage

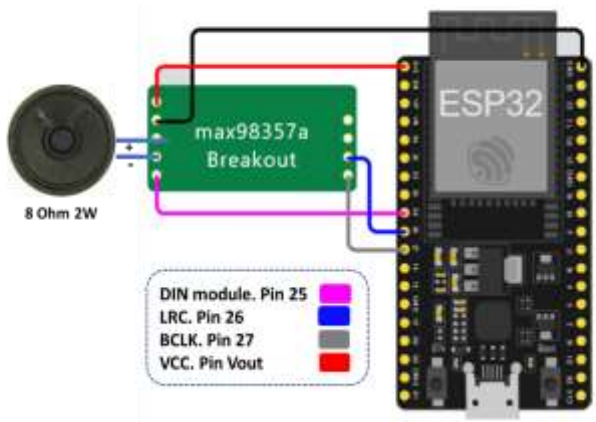

Fig. 7: MAX98357 Module Connections

Then, it converts the text into audio using an integrated voice synthesis system using the Audio.h library and the MAX98357 audio module. The generated response is played back with the audio.connecttospeech function. Finally, the process is restarted, clearing the Question variable to wait for the next user input (Figure 8).

3.4 Development of the Voice Acquisition Stage and Conversion to Text

The voice recognition and text conversion stage in the ESP32 microcontroller is conducted by integrating an electronic module that includes a microphone. This phase involves the implementation of a system that allows the device to interpret the user's voice commands, transforming them into text to be processed by the generative language model at a later stage. To do this, a library is integrated that allows the voice to be processed in a cloud service that analyzes the audio input, obtaining a string variable.

For voice acquisition, an INMP441 microphone was used using pin 26 for the BCLK or SCK (sets the rate at which audio data bits are transferred on the I2S bus), pin 22 for the WS (selects between the Left and Right audio channels), pin 34 for the DIN (data input pin for the ESP32) and pin 34 for the DOUT or SD (data output pin from the input device to the microcontroller) (Figure 9).

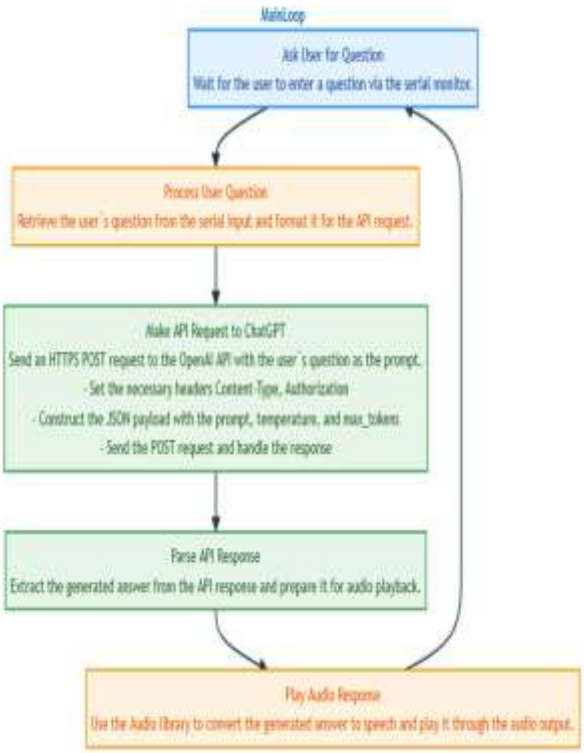

Fig. 8: Algorithm flowchart

Speech-to-Text API configuration is done through the Google Cloud Console platform. The Speech-to-Text API offers limited free use, for which a Project must be created in Google Cloud and then the Speech-to-Text API is enabled. This is done in the "APIs and Services" option, "Library", and "Speech-to-Text API" in the search bar and finally "Speech-to-Text API" is enabled. The algorithm at this stage acquires the data from the microphone and then encodes the audio using uncompressed, linear 16-bit PCM (Pulse Code Modulation). The audio data in Linear16 format is sent to the Google service, which processes it and converts it to text. Access to the Google service is done using the CloudSpeechClient.cpp library and the CloudSpeechClient::Transcribe(Audio* audio) function (Figure 10).

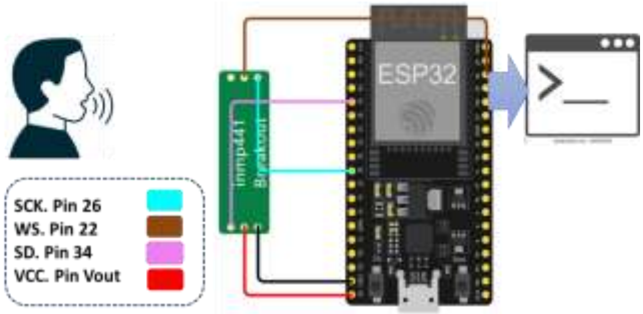

Fig. 9: Connection diagram for the voice acquisition module

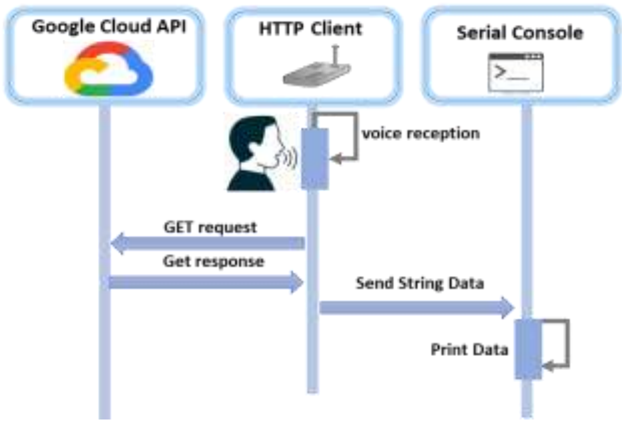

Fig. 10: Operation diagram for voice acquisition

3.5 Integration of the Stages of the System

To conduct the integration, the use of two components is considered, which will perform the tasks described above of acquiring the voice signal, converting it to text using the Google APIs, and querying ChatGPT with the generated text. While a second microcontroller is responsible for receiving the data in text form and performing the voice synthesis (Figure 11).

- Module 1. Performs speech recognition, sends a query to ChatGPT, and handles the response using LEDs to indicate the status of the process. The process starts with recording audio and sending it to Google Cloud to transcribe it to text. This text can then be sent to ChatGPT to generate a response. The CloudSpeechClient function handles the connection to Wi-Fi and the Google server, sending audio data in Base64 format, and receiving the audio transcription.

- Module 2. An ESP32 is configured to connect to a Wi-Fi network and use the I2S interface for audio synthesis and playback. "Serial2" serial communication receives commands or text, which are then converted to speech using Google Text-to-Speech.

4 Results

4.1 Integrating Communication with a Generative Language Model

The results of integrating communication with a generative language model demonstrate the effectiveness of the IoT device in interacting with ChatGPT. During testing, it was verified that the device can effectively send queries to the language model and receive responses in real-time, underlining its ability to handle complex natural language processing requests (Figure 12). This integration validates the technical approach adopted and establishes the flexibility of this stage. Although the device can send queries in real-time, limited qualitative analysis was performed. In addition, limitations in complex situations were addressed, necessitating more detailed studies in this area.

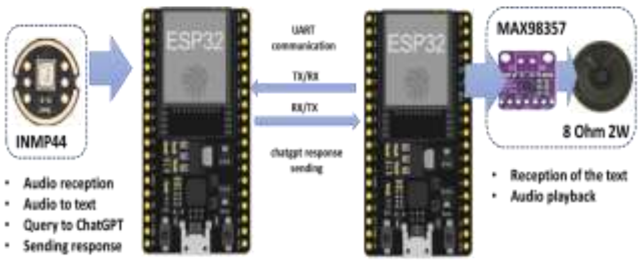

Fig. 11: Communication of the two modules of the system

4.2 Voice Synthesis Stage for the Generation of Auditory Responses

The speech synthesis results have proven to be effective in transforming the text generated by the language model into clear auditory responses. The use of the Audio.h library and the MAX98357 audio amplifier module, together with a 2 W, 8 ohm speaker, allow for high-quality sound reproduction via the I2S protocol. The operating scenario allowed validating the connection to the OpenAI API to answer queries using the serial port by extracting responses using the ArduinoJson library and converting the text to speech (Figure 13). Limited qualitative analysis was performed comparing sound quality, resulting in a preliminary assessment of audio quality. In this case, analysis is still pending on compatibility with other audio formats.

4.3 Voice Acquisition and Text Conversion Stage in ESP32

The implementation of a voice command interpretation system, using the INMP441 microphone and the Google Cloud Speech-to-Text platform, enabled accurate capture of the user's voice. Configuring the microphone to work with the I2S bus and encoding the audio in 16-bit PCM format ensured optimal data input quality (Figure 14). Reduced subjective comparisons were made with other voice capture devices and although basic metrics such as latency were considered, this requires more extensive evaluation in other noisy scenarios outside the scope of this research.

Fig. 12: Communication results with the generative model

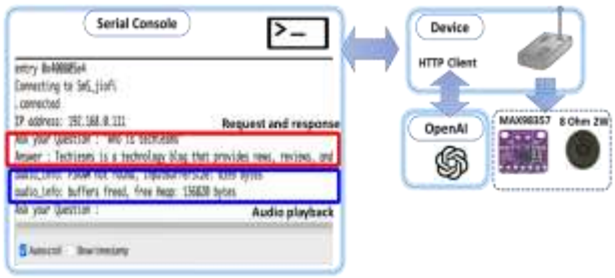

Fig. 13: Generating voice synthesis with ChatGPT responses

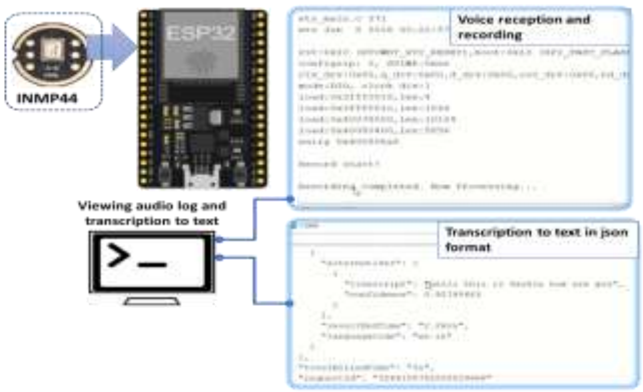

Fig. 14: Voice acquisition and text conversion stage in ESP32

In addition, CloudSpeechClient.cpp library to handle audio transcription facilitated access to Google’s advanced voice recognition capabilities and the sending of user queries by the generative language model.

5 Discussion

To evaluate the integration of the modules, pilot tests were carried out using the Wokwi simulation tool, which is an online simulation platform to evaluate all the stages described above. The system has proven to be effective in efficiently handling the acquisition and processing of voice signals. Module 1, responsible for voice recognition and communication with ChatGPT, has allowed obtaining the response through voice commands. On the other hand, Module 2 has been configured appropriately to handle voice synthesis from text.

Communication between modules via the serial port allows the transfer of text between modules, creating a modular architecture that optimizes the performance of the voice assistant and allows its scalability for future improvements (Figure 15)

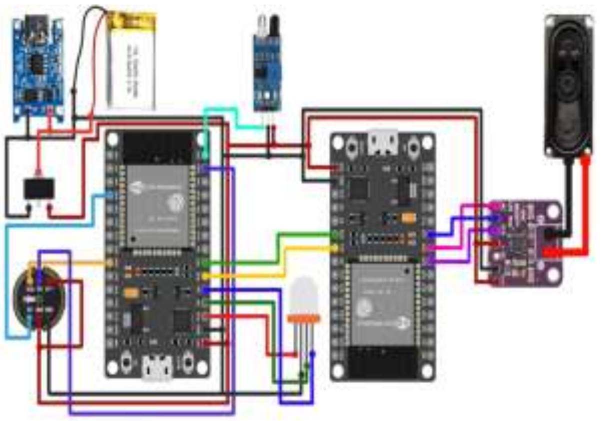

Fig. 15: Integrated voice assistant system

6 Conclusion

An IoT device was developed that integrates OpenAI's generative language technology, using techniques for accessing Web services and transforming text signals into speech and vice versa, through the design of a voice acquisition system, communication with a generative language model and serial communication between the two modules based on ESP32 microcontrollers.

The integration of the IoT device with a generative language model has proven its effectiveness when interacting with ChatGPT, sending queries, and receiving responses in real time. While the voice synthesis uses the Audio.h library and the MAX98357 module managed to transform the text into auditory responses. In the case of voice recognition with the INMP441 microphone and Google Cloud Speech-to-Text, accurate capture of voice commands was achieved. In addition, modular architecture is composed of two modules that handle voice recognition and synthesis, optimizing the performance of the voice assistant and guaranteeing efficiency and scalability for future improvements.

Future research may investigate the integration of machine learning capabilities into low-power IoT devices, which would allow the voice assistant to detect certain voice commands locally to perform critical activities in its environment. In addition, the analysis of the system's energy behavior and response latency could increase the device's autonomy and speed.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used Gemini to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the content of the publication.

References:

[1] A. Ermolina and V. Tiberius, “Voice-Controlled Intelligent Personal Assistants in Health Care: International Delphi Study,” *J. Med. Internet Res.*, vol. 23, no. 4, pp. 1–11, Apr. 2021, doi: 10.2196/25312.

[2] I. Froiz-Miguez, P. Fraga-Lamas, and T. M. Fernandez-Carames, “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, no. 1, pp. 63623–63649, 2023, doi: 10.1109/ACCESS.2023.3286391.

[3] R. Yauri, A. Castro, and R. Espino, “Automatic Leaf Health Monitoring with an IoT Camera System based on Computer Vision and Segmentation for Disease Detection,” *WSEAS Trans. Electron.*, vol. 15, no. 1, pp. 148–156, 2024, doi: 10.37394/232017.2024.15.17.

[4] I. Zakutynskyi, L. Sibruk, and A. Kokarieva, “IoT System for Monitoring and Managing

Public Transport Data," *WSEAS Trans. Syst.*, vol. 22, no. 1, pp. 242–248, 2023, https://doi.org/10.37394/23202.2023.22.25.

[5] C. Yin, Y. Mao, Z. He, M. Chen, X. He, and Y. Rong, "Edge Computing-Enabled Secure Forecasting Nationwide Industry PM2.5 with LLM in the Heterogeneous Network," *Electronics*, vol. 13, no. 13, pp. 1–16, Jun. 2024, doi: 10.3390/electronics13132581.

[6] J. Slamet, "Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions," *Discov. Artif. Intell.*, vol. 4, no. 1, pp. 1–15, Jul. 2024, doi: 10.1007/s44163-024-00143-2.

[7] C. Even, T. Hammann, V. Heyl, and C. Rietz, "Benefits and challenges of conversational agents in older adults," *Z. Gerontol. Geriatr.*, vol. 55, no. 5, pp. 381–387, Aug. 2022, doi: 10.1007/s00391-022-02085-9.

[8] M. Andrade, E. Wanderley, M. Azevedo, T. Medeiros, M. Silva, and I. Silva, "A Voice-Assisted Approach for Vehicular Data Querying from Automotive IoT-Based Databases," in *2023 Symposium on Internet of Things (SIoT)*, Sao Paulo, Oct. 2023, pp. 1–5. doi: 10.1109/SIoT60039.2023.10389856.

[9] P. Sanju, "Enhancing intrusion detection in IoT systems: A hybrid metaheuristics-deep learning approach with ensemble of recurrent neural networks," *J. Eng. Res.*, vol. 11, no. 4, pp. 356–361, Dec. 2023, doi: 10.1016/J.JER.2023.100122.

[10] C. Wang, Z. Yuan, P. Zhou, Z. Xu, R. Li, and D. O. Wu, "The Security and Privacy of Mobile-Edge Computing: An Artificial Intelligence Perspective," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 22008–22032, Dec. 2023, doi: 10.1109/JIOT.2023.3304318.

[11] J. Masud Cheema, M. Aqib, A. Elbeltagi, S. Rashid Saleem, S. Hussain, and A. M. Hayajneh, "TinyML Olive Fruit Variety Classification by Means of Convolutional Neural Networks on IoT Edge Devices," *AgriEngineering*, vol. 5, no. 4, pp. 2266–2283, Dec. 2023, doi: 10.3390/AGRIENGINEERING5040139.

[12] R. Yauri, E. Campos, R. Yalico, and V. Gamero, "Development of an Electronic Bird Repellent System using Sound Emission," *WSEAS Trans. Syst. Control*, vol. 18, no. 1, pp. 136–143, May 2023, doi: 10.37394/23203.2023.18.14.

[13] N. Tasmurzayev, B. Amangeldy, Y. Nurakhov, S. Shinassylov, and S. Dawit Bekele, "Development of an Intelligent Oil Field Management System based on Digital Twin and Machine Learning," *WSEAS Trans. Electron.*, vol. 14, no. 1, pp. 104–111, 2023, https://doi.org/10.37394/232017.2023.14.12.

[14] R. Gayathri, S. Maheswari, S. K. Mathivanan, B. D. Shivahare, R. R. Chandan, and M. A. Shah, "A comprehensive health assessment approach using ensemble deep learning model for remote patient monitoring with IoT," *Sci. Reports 2024 141*, vol. 14, no. 1, pp. 1–23, Jul. 2024, doi: 10.1038/s41598-024-66427-w.

[15] M. Sajid, K. R. Malik, A. Almogren, T. S. Malik, A. H. Khan, and J. Tanveer, "Enhancing intrusion detection: a hybrid machine and deep learning approach," *J. Cloud Comput.*, vol. 13, no. 1, pp. 1–24, Jul. 2024, doi: 10.1186/s13677-024-00685-x.

[16] S. L. Reeja, T. Soumya, and P. S. Deepthi, "A Study on Vision-Based Human Activity Recognition Approaches," *Smart Innov. Syst. Technol.*, vol. 373, no. 1, pp. 235–248, 2024, doi: 10.1007/978-981-99-6866-4_17.

[17] M. Razghandi, H. Zhou, M. Erol-Kantarci, and D. Turgut, "Variational Autoencoder Generative Adversarial Network for Synthetic Data Generation in Smart Home," in *ICC 2022 - IEEE International Conference on Communications*, Montreal, May 2022, vol. 2022-May, pp. 4781–4786. doi: 10.1109/ICC45855.2022.9839249.

[18] S. Joseph, R. Poorvaja, S. Shivaraj, V. Jeyanth and P. Shivesh, "IoT Empowered AI: Transforming Object Recognition and NLP Summarization with Generative AI," in *2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, Seoul, Dec. 2023, pp. 1–6. doi: 10.1109/CVMI59935.2023.10465077.

[19] L. Cotti, D. Guizzardi, B. R. Barricelli, and D. Fogli, "Enabling End-User Development in Smart Homes: A Machine Learning-Powered Digital Twin for Energy Efficient Management," *Futur. Internet*, vol. 16, no. 6, pp. 1–16, Jun. 2024, doi: 10.3390/fi16060208.

[20] A. Javed, A. Ehtsham, M. Jawad, M. N. Awais, A.-H. Qureshi, and H. Larijani, "Implementation of Lightweight Machine Learning-Based Intrusion Detection System on IoT Devices of Smart Homes," *Futur. Internet*, vol. 16, no. 6, p. 200, Jun. 2024, doi: 10.3390/fi16060200.

[21] A. Hennebelle, L. Ismail, H. Materwala, J. Al Kaabi, P. Ranjan, and R. Janardhanan, "Secure and privacy-preserving automated

machine learning operations into end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction," *Comput. Struct. Biotechnol. J.*, vol. 23, no. 1, pp. 212–233, Dec. 2023, doi: 10.1016/J.CSBJ.2023.11.038.

[22] M. A. Murti, A. R. A. Saputra, I. Alinursafa, A. N. Ahmed, A. Yafouz, and A. El-Shafie, "Smart system for water quality monitoring utilizing long-range-based Internet of Things," *Appl. Water Sci.*, vol. 14, no. 4, p. 69, Apr. 2024, doi: 10.1007/s13201-024-02128-z.

[23] M. Gaffurini, A. Flammini, P. Ferrari, D. Fernandes Carvalho, E. P. Godoy, and E. Sisinni, "End-to-End Emulation of LoRaWAN Architecture and Infrastructure in Complex Smart City Scenarios Exploiting Containers," *Sensors 2024, Vol. 24, Page 2024*, vol. 24, no. 7, pp. 1–23, Mar. 2024, doi: 10.3390/S24072024.

[24] T. Chinbat, S. Madanian, D. Airehrour, and F. Hassandoust, "Machine learning cryptography methods for IoT in healthcare," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, pp. 1–26, Dec. 2024, doi: 10.1186/S12911-024-02548-6.

[25] Y. Luo, J. Tao, Y. Zhu, and Y. Xu, "HSS: enhancing IoT malicious traffic classification leveraging hybrid sampling strategy," *Cybersecurity*, vol. 7, no. 1, pp. 1–13, Jun. 2024, doi: 10.1186/s42400-023-00201-9.

[26] S. Zhang, F. Ye, B. Wang, and T. G. Habetler, "Semi-Supervised Bearing Fault Diagnosis and Classification Using Variational Autoencoder-Based Deep Generative Models," *IEEE Sens. J.*, vol. 21, no. 5, pp. 6476–6486, Mar. 2021, doi: 10.1109/JSEN.2020.3040696.

[27] S. P. Sithungu and E. M. Ehlers, "From Concept to Prototype: Developing and Testing GAAINet for Industrial IoT Intrusion Detection," in *IFIP Advances in Information and Communication Technology*, vol. 703 IFIPAI, no. 1, Springer, Cham, 2024, pp. 453–468. doi: 10.1007/978-3-031-57808-3_33.

[28] S. Aramkul and P. Sugunnasil, "Intelligent IoT framework with GAN- synthesized images for enhanced defect detection in manufacturing," *Comput. Intell.*, vol. 40, no. 2, p. e12619, Apr. 2024, doi: 10.1111/coin.12619.

[29] H. Sharifpour, Y. Ghaseminezhad, M. Hashemi-Tabatabaei, and M. Amiri, "Investigating cause-and-effect relationships between supply chain 4.0 technologies," *Eng. Manag. Prod. Serv.*, vol. 14, no. 4, pp. 22–46, Dec. 2022, doi: 10.2478/EMJ-2022-0029.

[30] J. Liu, W. Jiang, R. Sun, A. K. Bashir, M. D. Alshehri, and Q. Hua, "Conditional Anonymous Remote Healthcare Data Sharing Over Blockchain," *IEEE J. Biomed. Heal. Informatics*, vol. 27, no. 5, pp. 2231–2242, May 2023, doi: 10.1109/JBHI.2022.3183397.

[31] S. Walter, "AI impacts on supply chain performance : a manufacturing use case study," *Discov. Artif. Intell.*, vol. 3, no. 1, p. 18, May 2023, doi: 10.1007/s44163-023-00061-9.

[32] P. Marques, A. Emami, G. Zhang, R. D. Lopes, and A. Razaghizad, "Impact of age and comorbid heart failure on the utility of smart voice-assistant devices," *Eur. Hear. J. - Digit. Heal.*, vol. 5, no. 3, pp. 389–393, May 2024, doi: 10.1093/ehjdh/ztae012.

[33] P. Sossenheimer, G. Hong, A. Devon-Sand, and S. Lin, "Voice Assistants' Responses to Questions About the COVID-19 Vaccine: National Cross-sectional Study," *JMIR Form. Res.*, vol. 7, no. 1, pp. 1–8, 2023, doi: 10.2196/43007.

[34] S. M. Alharthi, "Siri as an interactive pronunciation coach: its impact on EFL learners," *Cogent Educ.*, vol. 11, no. 1, pp. 1–20, Dec. 2024, doi: 10.1080/2331186X.2024.2304245.

[35] R. Zhong, M. Ma, Y. Zhou, Q. Lin, L. Li, and N. Zhang, "User acceptance of smart home voice assistant: a comparison among younger, middle-aged, and older adults," *Univers. Access Inf. Soc.*, vol. 23, no. 1, pp. 275–292, Mar. 2024, doi: 10.1007/s10209-022-00936-1.

[36] L. Reid and G. Sisel, "Digital care at home: Exploring the role of smart consumer devices," *Health Place*, vol. 80, no. 1, p. 102961, Mar. 2023, doi: 10.1016/J.HEALTHPLACE.2022.102961.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding was received for conducting this study

Conflict of Interest
The authors have no conflicts of interest to declare.