Design of Generative AI-based Speech-Recognizing KIOSK for Educational Institutions

Eun kyung Jung
Gyeongbuk ICT Advanced Talent Training Center
Deagu Catholic University
Gyeongsan, Rep. of Korea
f1rstf1y9@naver.com

Na yeong Kwon
Gyeongbuk ICT Advanced Talent Training Center
Deagu Catholic University
Gyeongsan, Rep. of Korea
cbxbc000zzz@gmail.com

Bo mi Lee
Gyeongbuk ICT Advanced Talent Training Center
Deagu Catholic University
Gyeongsan, Rep. of Korea
lbm000314@gmail.com

Sang jo Baek
Gyeongbuk ICT Advanced Talent Training Center
Deagu Catholic University
Gyeongsan, Rep. of Korea
skyboy822@naver.com

Ji soo Ryu
Dept. of Research & Development
Dreamideasoft Co. Ltd.
Deagu, Rep. of Korea
gnt700@dreamideasoft.net

Kee sung Kim[†]
School of Computer Software
Daegu Catholic University
Gyeongsan, Rep. of Korea
likee21@cu.ac.kr

Abstract: This paper discusses the development of an AI voice recognition-based kiosk for digitally marginalized groups, such as the elderly and users with physical limitations. By utilizing natural language processing (NLP) and generative AI models, the system interprets user intent and proposes a conversational system design that provides facility guidance and service information in public and educational institutions through voice.

Keywords—AI voice recognition, RASA, natural language processing (NLP), kiosk, digitally marginalized groups

I. Introduction

Modern society, alongside the rapid development of information technology, is experiencing a digital divide. Services provided by public and educational institutions often remain inaccessible to digitally marginalized groups, causing social alienation and inconvenience. As of 2023, the digital literacy level of Korea's four major vulnerable groups (people with disabilities, low-income individuals, farmers and fishermen, and the elderly) was reported to be 76.9% compared to the general population, with digital competence at just 65.1%[1]. This indicates that many individuals in these groups lack even the most basic skills for using GUI-based systems like PCs or mobile devices.

Traditional kiosk systems predominantly adopt touch-based GUIs, which are neither intuitive nor accessible for these groups, highlighting their limitations. This raises the need for a new type of kiosk system that provides an experience similar to interacting with a real person, rather than navigating complex GUIs.

In this research, we propose the design of a kiosk system utilizing generative AI-based voice recognition technology to enable natural dialogue with users. Through natural language understanding (NLU) and natural language generation (NLG), along with speech-to-text (STT) and text-to-speech (TTS) technologies, the system can accurately comprehend user intent and simplify complex menu navigation and information entry processes, significantly improving user experience.

Thus, the goal of this study is to develop an innovative kiosk system that allows diverse users, including the digitally marginalized, to easily access public services, thereby providing equitable technological benefits to all.

II. Related Works

A. Kiosk

A kiosk is an interactive terminal designed to provide information and services to users in various environments, such as public spaces, retail stores, and educational institutions. Historically, kiosks started as simple static structures offering limited information and have evolved into advanced systems delivering immersive user experiences. Modern kiosks are found in applications like self-checkouts at supermarkets, ticket machines at transport hubs, and information displays in museums. These developments reflect the growing demand for convenient and user-friendly interfaces in service provision. However, complex menu structures and non-intuitive user experiences remain barriers to kiosk use[2]. These issues indicate a lack of user-friendliness in current kiosk systems, particularly in addressing the needs of digitally marginalized groups.

B. NLP (Natural Language Processing)

Natural language processing (NLP) encompasses a range of technologies enabling machines to understand and generate human language. NLP's two main components are natural language understanding (NLU) and natural language generation (NLG). NLU allows systems to interpret user input by accurately deciphering intent and context, while NLG generates consistent and context-appropriate responses, facilitating effective communication between the machine and the user[3].

NLP has been successfully applied in areas such as virtual assistants, chatbots, and customer service applications. Systems like Amazon Alexa and Apple Siri use advanced NLP algorithms to facilitate seamless interactions with users, demonstrating the potential of these technologies to enhance user experiences.

C. YML

YML is a human-readable data serialization format, widely used for configuration files and data exchange between programming languages. In the context of interactive systems, YML can be leveraged to define and structure the dialogue flow between the user and the system. One of YML's key advantages is its simplicity and ease of use, which makes it highly popular among developers and data scientists alike.

In this research, YML plays a crucial role in managing the conversational flow of the AI-based kiosk system. By utilizing

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry Education(MOE)(No.: 2022RIS-006).

YML, developers can easily modify the dialogue paths, making the system more adaptable and responsive to user feedback. For instance, YML allows for dynamic adjustments based on user input, making it easier to update interaction patterns and improve user experience over time. This flexibility is particularly beneficial in creating customized solutions for users from diverse backgrounds, ensuring that the system can cater to the specific needs of digitally underserved populations.

D. *RASA*

RASA is an open-source framework for building conversational AI chatbots. It consists of two main components: Rasa NLU (Natural Language Understanding) and Rasa Core. Rasa NLU is responsible for understanding the user's intent and extracting key entities from the input, while Rasa Core manages the dialogue engine, allowing for more complex and customizable interactions[4].

In this research, RASA is utilized to enhance the natural language understanding capabilities of the kiosk system. The decision to use RASA is driven by its flexibility and open-source nature, enabling developers to customize the system according to the specific requirements of public and educational institutions. Rasa NLU helps in accurately capturing the user’s intent, which is crucial for providing relevant responses to users who may have limited digital literacy. Rasa Core, on the other hand, enables the system to handle multi-turn dialogues, making it more robust in guiding users through various service options.

Additionally, the combination of Rasa’s NLU and Core components allows for continuous improvement. As more user data is collected, the system can be fine-tuned to better understand regional accents, dialects, or colloquialisms, further enhancing accessibility for underserved users. This adaptability makes RASA a key element in developing an AI-powered kiosk system that aims to bridge the digital divide.

E. *Fine Tuning*

Fine-tuning is the process of retraining a pre-trained model on a small dataset tailored for a specific task, optimizing performance for a particular use case. In the case of a generative AI-based voice recognition kiosk, it is essential to fine-tune general natural language models (like GPT and BERT) to accurately process and respond to the specific domain of the kiosk system (such as public institutions, education, and retail environments). This process contributes not only to basic language understanding but also to providing tailored solutions that meet the specific requirements of various industries.

Fine-tuning enhances the model’s ability to comprehend specialized terminology and vocabulary used in these environments, significantly improving natural language understanding (NLU) and natural language generation (NLG) capabilities. This enables AI to interpret user input more accurately and generate appropriate responses aligned with the context of the conversation. In particular, fine-tuned models are equipped to provide more relevant and useful information in response to user inquiries specific to certain domains[5].

Moreover, leveraging fine-tuning allows generative AI to offer more relevant and personalized support, resulting in a more intuitive and efficient user experience. As the model’s responses gradually become more accurate and contextually appropriate, user satisfaction can be enhanced, while also increasing the reliability and utility of the kiosk system. Ultimately, this approach makes interactions with users more natural and seamless.

III. Methods

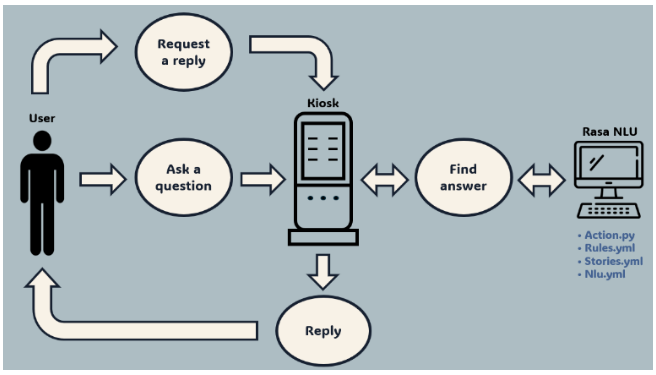


Fig. 1. Functional structure diagram

When a user asks a question by voice to a kiosk, the kiosk outputs the question content on the kiosk screen after voice recognition (Fig. 1). Thereafter, Rasa NLU (natural language understanding processing) is performed so that a keyword of a question may be extracted and a voice answer may be output through Speech To Text (STT) and Text To Speech (TTS).

Technology stacks used here include natural language processing engine, interactive AI, server, and front-end. Understand the concept of intention identification processing for user input using a natural language processing engine, and generate a natural and specific response using a Generative AI. By linking a natural language processing engine and a Generative AI, a user interface is implemented to receive user questions and show responses.

The 'Rasa Core', which is the management of the conversation flow of the learning server configuration Rasa used in this study, is based on four things. Rasa Core will be operated through Action.py , rules.yml, stories.yml, and nlu.yml. The description of each base file is as follows.

The Action.py file becomes the basis for running the RASA chatbot. It depends on the developer's capability or the environment in which you want to implement it. You can also run a function customized in action through the rule or domain, and only one action.py can exist.

The Rules.yml file sets specific rules. If Rasa specifies a specific rule and satisfies the intent of the rule, the specified result should be acted upon you will refer to the function at action.py . Rule.yml can exist only one.

The Stories.yml file contains the story of the rasa chatbot. In other words, if a question is asked close to the intent set in nlu.yml by setting a setting flow, the result action linked to the intention is made through the flow. You can set multiple intent->actions within a flow, or you can set entities to act in detail with similar intentions. Stories.yml files can be used by dividing them into several file names, and it is better to organize them by dividing them into similar topics or categories as much as possible.

Nlu.yml files often have intentions set through expected questions. Words with the same or similar meanings can be grouped by setting a synonym using intent, or similar sentence forms can be grouped by setting a regular expression (regex).

In addition, by setting a lookup, the country name can be grouped into a lookup: country. Nlu.yml files can be used by dividing them into several file names, and it is better to organize them by dividing them into similar topics or categories as much as possible.

If the user finds an appropriate answer to the question in the Rasa NLU, the answer is delivered by voice through a kiosk. If an unwanted answer is given, the user may request a re-answer from the kiosk. In this case, a new answer is provided to the user after going through the Rasa NLU (natural language understanding processing) process again.

In addition, there is a function of adopting an answer by scoring a score on the exact answer after grasping the intention of the user based on the NLU. This part is a method of delivering an action for a high-scoring intent value to the damin. The answers adopted by Damin are sent to the user. In addition, we will develop a function that allows simple daily conversations such as greetings and chatbots that enable conversations under various conditions.

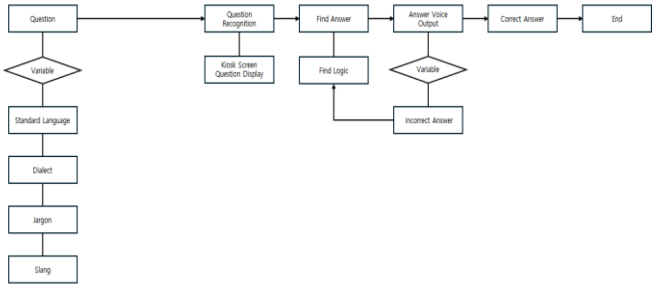


Fig. 2. Senario Flow

The precess of managing user inquiries through a kiosk begins with the Question Input Stage (Fig. 2.).

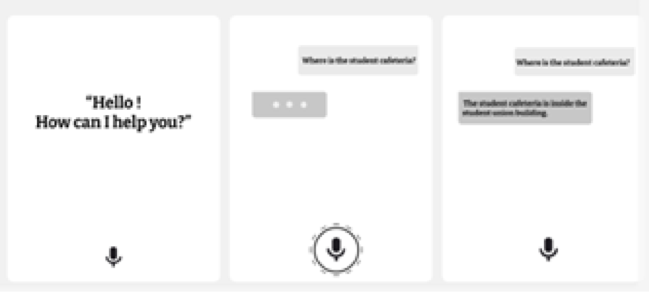
In this initial phase, users submit their questions using voice input, which can be articulated through either the kiosk itself or a connected voice recognition device. This input encompasses a wide range of language forms, including standard speech, various dialects, slang, and colloquialisms, reflecting the diverse ways people communicate. During this stage, the system diligently checks multiple variables to accurately comprehend the nature of the question being posed, ensuring that it can respond appropriately.

Fllowing the input, the process moves into the Question Recognition Stage. At this point, the system's primary focus is on accurately identifying the user's question. Advanced algorithms analyze the voice input to determine the intended meaning. Once the system successfully recognizes the question, it displays the text representation on the kiosk screen. This visual confirmation allows users to verify that their inquiry has been understood correctly, creating an interactive and user-friendly experience.

After the question has been confirmed, the process advances to the Answer Retrieval Stage. Here, the system undertakes the task of searching for an appropriate response to the user's query. It analyzes the content of the question and engages a sophisticated chatbot powered by RASA, which is designed to process and retrieve answers. This stage also includes a critical logic retrieval component, which enables the system to conduct additional logical processing. This is particularly important for questions that may require a more complex or nuanced answer, ensuring that the response is not only accurate but also contextually relevant.

Once the system has identified a suitable answer, the process transitions to the Answer Voice Output Stage. In this phase, the system delivers the response to the user in two ways: audibly, using Text-To-Speech (TTS) technology, and visually, by displaying the text on the screen. This dual output method enhances the user's understanding and ensures clarity. If the answer provided is incorrect or does not meet the user's expectations, the process can loop back to earlier stages for further reprocessing and refinement. This flexibility is crucial for maintaining a high level of user satisfaction.

However, if the system successfully delivers a correct and satisfactory response, and the user has no additional inquiries, the process concludes gracefully. This structured approach ensures that users receive accurate information efficiently, fostering a positive interaction with the kiosk system. Throughout each stage, the emphasis is on clarity, accuracy, and user engagement, making the entire experience both intuitive and effective.


"Hello !
How can I help you?"

Where is the student cafeteria?

Where is the student cafeteria?

The student cafeteria is inside the student union building.

Fig. 3. Kiosk Screen

The screen design of Generative AI-based speech recognition kiosks focuses on optimizing user interaction. Various design elements should be included to enhance the user experience. The design of generative AI-based speech recognition kiosks should focus on optimizing user interaction, incorporating various design elements.

First, when a user approaches the kiosk, a welcome message should immediately be displayed to grab the user's attention and create a sense of friendliness. A message like "Hello! How can I assist you today?" (Fig. 3) can provide a positive first interaction with the user[6]. This increases the usability of the kiosk and helps users feel comfortable interacting with the technology. Second, the voice recognition status display is an essential element to visually communicate that voice recognition is activated. A microphone icon can be used to indicate that voice recognition is in progress, allowing the user to recognize that the voice input is functioning correctly[6]. This visual feedback plays a crucial role in building user trust in voice-based interfaces.

The conversation flow should include a feature that displays the user's questions or requests in real-time text form on the screen. This allows users to immediately confirm the results of voice recognition and provides an opportunity to make corrections if necessary. Such text-based feedback improves the accuracy of voice recognition and helps users express their intentions more clearly. This feature is especially beneficial for users with accents or speech impediments, ensuring that everyone has the opportunity to communicate clearly.

Visual feedback also plays an important role in user-friendly design. By providing visual cues through loading animations when processing user requests, the system demonstrates to users that it is responding. This contributes to enhancing the user experience and conveying the system's stability to users[6]. they can make corrections before the system processes the input, thus improving the overall accuracy and effectiveness of the interaction.

A user-friendly interface should be designed to be easily understood and used by users of all ages and technical skill levels[7].It is important to ensure that even users with limited technical experience can use the kiosk smoothly through intuitive and concise design. To achieve this, users should be able to easily recognize and navigate the system through an intuitive design. Namely, The user interface should avoid complex procedures or technical jargon, instead presenting information in a way that is easy for anyone to understand. Key features and options should be clearly displayed, and visual feedback, along with simple instructions, should help users feel confident that they are following the correct steps.

By comprehensively considering design elements such as welcome messages, speech recognition status indicators, conversation flow, and visual feedback, a user-friendly and intuitive conversational kiosk screen can be implemented. This design approach will improve interactions between users and the system, maximizing the efficiency of speech recognition-based systems.

IV. Conclusion

The generative AI-based voice recognition kiosk system proposed in this research is anticipated to significantly enhance digital accessibility for users experiencing the digital divide. By streamlining administrative services in public and educational institutions, this system is set to transform user convenience, enabling individuals who may be less familiar with technology to access essential services more easily. The comprehensive approach to voice data collection, which includes capturing dialects, slang, and other linguistic nuances, alongside guidance based on RASA NLP, will be instrumental in bridging the information gap. This will not only cater to a wider audience but also ensure that the kiosk system can effectively respond to the diverse communication styles of users, thereby promoting inclusivity and reducing barriers to entry.

Moreover, the generative AI model's ability to learn from user interactions will foster a more intuitive interface, alleviating challenges posed by complex menu structures and non-intuitive user experiences. As the system evolves, its adaptability to various user needs will enhance usability, ensuring that even those with minimal digital skills can navigate services with confidence. This evolution is not just about technology; it’s about creating an environment where everyone can engage with digital services meaningfully.

The proposed system will benefit from continuous technological updates and improved compatibility, which will be crucial in maintaining its relevance and effectiveness. The ongoing development of advanced algorithms will enable more natural and fluid interactions between users and kiosks, fostering a sense of familiarity and comfort. Tailored solutions applicable across various sectors will be prioritized, ensuring that the kiosk system addresses the specific needs of diverse industries, from healthcare to retail. This proactive approach to innovation will ensure that the system remains relevant, scalable, and responsive to the ever-evolving technological landscape.

In addition to enhancing access to essential services, the incorporation of tailored services for the digitally marginalized will empower these users, giving them the tools they need to navigate the digital world with greater ease and independence. By focusing on user-centered design and inclusive practices, this system can significantly impact the lives of those who have traditionally faced barriers in accessing technology.

The potential applications of this system extend beyond public institutions to encompass financial institutions, commercial facilities, and various other sectors. When linked with comprehensive digital literacy programs, the kiosk system can enhance usability and promote greater engagement with technology. By actively fostering digital literacy alongside the implementation of this technology, we can create a synergistic effect that empowers individuals across all ages and social classes, contributing to a more equitable distribution of technological benefits and fostering a more informed society[7].

In conclusion, the outcomes of this study are expected to contribute to making public services more accessible to all, including the digitally marginalized, by providing equitable technological benefits that meet the diverse needs of users. Future research should concentrate on verifying the system's effectiveness and exploring avenues for continuous improvement through extensive data collection and analysis. This will not only ensure that the kiosk system adapts to user feedback but also lay the foundation for offering a more inclusive and innovative user experience. By actively addressing the gaps in digital accessibility, this research seeks to reinforce the commitment to bridge the digital divide and promote equity in access to technology across society. Through such efforts, we can envision a future where technology serves as a unifying force, fostering inclusivity and improving the quality of life for all members of society.

References

[1] National Information Society Agency, “2023 The Report on the Digital Divide,” pp. 7, 2024.

[2] S. Y. Hong and J.-H. Choe, “A study on the kiosk UI reflecting the elderly's characteristics, ” *J. Korea Contents Assoc.*, vol. 19, no. 4, pp. 556-563, Apr. 2019.

[3] C.-S. Lee and H.-J. Baek, “A Study on The Need for AI Literacy According to The Development of Artificial Intelligence Chatbot,” *The Journal of the Korea institute of electronic communication sciences*, vol. 18, no. 3, pp. 421–426, Jun. 2023.

[4] Prof. Yogesh S Sapnar, Dnyaneshwari P Kodlinge, Sakshi B Jogde, Kimaya V Bhosale, and Simran S Kudale, “RASA Chatbot Using AI,” *International Journal of Advanced Research in Science, Communication and Technology*. Naksh Solutions, pp. 348–353, 12-Apr-2022.

[5] H. S. Kim, “Development of chat web service functions for administrative and public institutions in the cloud environment using the implementation of RAG technology data learning automation, ” M.S. thesis, Dept. of AI IT Convergence, Soongsil Univ., Seoul, South Korea, Jun. 2024.

[6] Z. Wang, “Research on mobile application interface design strategy under the background of aging society, ” The Frontiers of Society, 2024.

[7] Dream Idea Soft, “2024 Gyeongbuk ICT Advanced Talent Training Center Project Corporate Collaboration Task Application,” 2024