Intelligent Assistant Based on Voice Driven Technology

Jinghong Lai *

School of International Digital Economy Institute, Minjiang University, Fuzhou, China

* Corresponding Author Email: 2024EUM27@students.euc.ac.cy

Abstract. The application of artificial intelligence technology in the field of education is becoming increasingly widespread, particularly in foreign language teaching and learning, which has emerged as a hot topic in research and teaching practice. Students often rely on mere rote memorization to learn Japanese, memorizing vocabulary, grammar, and texts. Their learning methods are outdated and rigid, leading to significant psychological pressure on students. However, nowadays, people have adopted a new learning approach, utilizing artificial intelligence to transform learning methods and further enhance students' vocabulary. This article primarily explores the research on intelligent assistants in both daily life and learning. In the future, artificial intelligence will be integrated with learning methods and adapt to the current trends of the era, thereby providing learners with more efficient, innovative, and personalized learning approaches and methods. With the application and popularization of Artificial intelligence (AI) voice assistants, the privacy protection of individual users needs further improvement and optimization, the user experience also needs to be gradually improved.

Keywords: Artificial intelligence voice assistant; Voice interaction; Artificial intelligence.

1. Introduction

In 1952, Alden H. Lemmer, a researcher at American Telephone and Telegraph (AT&T), developed the first speech recognition system. In the same year, Bell Labs manufactured a 6-foot-tall automatic number recognition machine named "Audrey", which could recognize the pronunciation of numbers 0-9 with an accuracy rate of over 90%. It was an important early achievement in speech recognition technology. In 1966, Joseph Weizenbaum of the Massachusetts Institute of Technology developed ELIZA, which is considered the first referenced chatbot in the history of technology. The application of artificial intelligence technology in the field of education is becoming increasingly widespread. Among them, Artificial intelligence (AI) voice assistants have become crucial tools for personalized and adaptive learning. They can accurately recognize and understand learners' voice inputs and even provide timely feedback. Meanwhile, the application of artificial intelligence in the field of foreign language teaching and learning has become a hot topic in research and teaching practice. Students often rely on mere rote memorization to learn Japanese, memorizing vocabulary, grammar, and texts. Their learning methods are outdated and rigid, leading to significant psychological pressure on students. However, nowadays, people have adopted a new learning approach that utilizes artificial intelligence to transform learning methods, further enhancing students' vocabulary and improving their learning outcomes.

The application of AI voice assistants has not only transformed traditional teaching modes but also influenced learners' learning behaviors, cognitive styles, and learning motivations [1]. Through voice interaction, AI can reduce learners' cognitive load, enhance the convenience of information acquisition, enrich personalized learning experiences, and foster students' autonomous learning abilities [2, 3]. With the development of technology, AI voice assistants can clearly recognize different needs of users, thereby greatly improving the efficiency of handling affairs. Artificial intelligence has a wide range of applications. For instance, nowadays, PowerPoint presentations, academic papers, and copywriting can all be written using artificial intelligence. With the continuous development of AI, functions such as autonomous driving and unmanned stores, which are replaced by AI, have continuously emerged in our lives. Artificial intelligence has also made significant contributions to the design and development of games. For instance, the non-player characters (NPCs) we encounter in our daily gameplay are what we commonly refer to as "artificial intelligence". These

"artificial intelligences" are all controlled by artificial intelligence. These NPCs are also ranked, roughly from low to high. Some NPCs are even stronger than real people, which indicates that the research on artificial intelligence is worth pursuing.

Voice skills are used for the development of applications on smart voice devices in IoT scenarios, such as the voice devices provided by Amazon Alexa and Google Assistant. They interact with users through voice channels, and developers need to download specific voice skills, such as "unlock the door", And integrate with the applications of related IoT devices to achieve the function of controlling devices with voice. That is, when the user issues a voice command, the smart device will recognize the voice instruction, match it with the corresponding voice skill, and invoke the skill to implement the corresponding service and functional logic [4]. In terms of precise voice activation, Xiaomi Auto adopts five-zone voice interaction in the car, which can accurately recognize voice commands from different positions inside the car. Even commands issued by passengers in the middle of the rear seats can be accurately captured. Outside the car, Xiaomi Automotive has arranged an array consisting of 8 microphones, which enables 360° all-round voice activation. Users can give commands at a distance of about 30 meters from the vehicle at normal volume, and Xiaoai can clearly recognize them. Some domestic voice intelligent assistants have advantages in connecting with the hardware ecosystem.

To better enable AI voice assistants to provide students with formula derivation, concept explanation, and interactive exercises, and to further integrate the capabilities of AI voice assistants into the intelligent tutoring system, thereby utilizing deep learning and knowledge graph technology to achieve automated teaching, learning path optimization, and intelligent assessment. Moreover, people should strengthen the theoretical research and top-level design of intelligent learning, promote technological development and practical application, and improve the quality of talent cultivation. This provides learners with more learning motivation and further optimizes foreign language teaching methods [2]. The research on AI voice assistants is primarily aimed at providing users, especially college students, with a more advanced, convenient, efficient, and intelligent experience. Furthermore, it aims to apply the numerous advantages of AI voice assistants to enhance students' learning efficiency in the classroom. Next, this article mainly introduces the concepts, advantages, disadvantages, and current applications of semantic understanding, speech recognition, and speech synthesis technologies. It also discusses the various challenges faced by AI voice assistants and their applications in daily life.

2. Technical Analysis

2.1. Semantics Understanding

Semantic understanding is a technology that recognizes and analyzes speech [5]. Language understanding primarily focuses on language as the object, conducts quantitative analysis of language information with the help of computers, and provides language descriptions that can be commonly used between humans and computers. It includes two parts: natural language understanding and natural language generation. The implementation steps of semantic understanding technology are as follows: extract the main keywords, then comprehend the context one by one, and finally analyze the user's intention. In the process of semantic understanding, only by effectively solving problems such as lexical, syntactic, and grammatical ambiguity, background knowledge ambiguity, and emotional thought concealment, can the meaning of the user's voice be accurately analyzed [5]. With the rapid development of artificial intelligence technologies such as deep learning, coupled with the application of big data and cloud computing, a vast amount of information in the information age is being utilized efficiently, and empiricism based on statistics has gradually become the mainstream method. The improvement of ambiguity resolution in natural language has been greatly enhanced, thereby significantly enhancing the accuracy and real-time performance of intent understanding.

2.2. Speech Recognition

Speech recognition is responsible for collecting users' voices and recognizing them. Its implementation steps are as follows: recording, noise reduction, and converting unstructured speech into structured text and computer language [6]. After sampling the simulated speech signal to obtain waveform data, the speech recognition system inputs it into the acoustic feature extraction module to extract appropriate acoustic feature parameters for use in acoustic model training [7]. Most of the currently mainstream acoustic models employ the Hidden Markov Model (HMM). The state transition characteristics of this model align with the short-term stationarity of human speech. The dynamic programming algorithm associated with it can effectively segment and classify variable-length speech sequences, and it is less affected by emotions, states, environments, and regions.

2.3. Speech Synthesis

Speech synthesis technology involves determining the user's intention after speech recognition and semantic understanding, converting the result into text, converting the text into speech, and broadcasting the speech to the user [7]. The process of speech synthesis is divided into the front-end and back-end. The main function of the front-end module is to preprocess the text to be synthesized and convert it into a symbolic description of pronunciation, transforming the text into a phoneme sequence containing prosodic information [5]. The synthesis backend module converts the symbolic description of the pronunciation output by the frontend into speech waveform. The mainstream speech synthesis backends can be divided into two categories: waveform concatenation speech synthesis and statistical parametric speech synthesis [6].

3. Discussion

Although AI voice assistants can adopt a slower speaking speed, provide more detailed explanations, and offer more demonstrations, and for advanced learners, they can reduce basic explanations, provide more challenging questions, and encourage students to engage in deep thinking, over time, students may become overly dependent on AI voice assistants [8]. In mathematics learning, if students answer multiple questions correctly in succession, the AI voice assistant can automatically increase the difficulty level appropriately. Conversely, it will provide relatively simpler exercises. In reading comprehension assessments, the AI voice assistant can guide students to think about the core content of the article through conversational questioning, and assess their comprehension ability [2]. Utilizing some functions of AI voice assistants can help teachers explore the deeper meanings contained in texts more deeply, further expand and discover more valuable teaching content, and thus further improve teaching quality. In noisy environments, background noise can affect the accuracy of voice recognition by voice assistants to some extent, thereby reducing the user's experience. At the same time, dialects and accents from different regions also pose a significant challenge to AI voice assistants, necessitating stronger adaptability. In today's globalized world, intelligent assistants need to support multiple languages, and achieving accurate recognition in a multilingual context is also quite challenging. Today's Doubao is not very perfect. For example, it cannot generate images for users, and there is a phenomenon that when facing some basic mathematical problems, the answers it gives are wrong, but the process is correct. Another problem is that when facing some difficult physics or mathematical problems, most of the answers it gives are wrong. For gamers who have played the game, the popular game "Peace Elite" now incorporates the voice assistant "DeepSeek". In the game, there is a mode where you can team up with such NPCs, who can not only engage in conversation with you but also fight alongside you. However, it currently cannot clearly recognize what you say. Basically, people need to repeat what you say at least three times before it can clearly recognize what you said. It may even significantly reduce the player's game experience [9].

Future AI voice assistants will be deeply integrated with educational big data platforms to provide a more precise and personalized learning experience. Future AI voice assistants will not merely be a voice interaction system, but will be able to deeply integrate with virtual and augmented reality

technologies to offer a more immersive and interactive learning experience. Future AI voice assistants will adopt more advanced privacy computing technologies, such as federated learning and differential privacy, to ensure the security of user data. The application of voice assistants will gradually expand to different fields such as traffic safety, intelligent interaction, medical health, and information security in the future. In the future, voice assistants will integrate multiple interaction methods such as voice, gestures, and facial expressions. For instance, in the control of smart homes, users can utilize settings to enable voice commands, gestures, and facial expressions to control the voice assistant, thereby obtaining personalized services [10-12]. Voice assistants are not far from achieving seamless integration across platforms and devices. This means that users will be able to use the same voice assistant on different devices, and the voice assistant will be able to synchronize users' data privacy and preferences, thereby providing an efficient and consistent user experience.

4. Conclusion

In summary, AI voice assistants have broad application prospects in the field of education. Their innovations in personalized learning, adaptive teaching, and intelligent evaluation will reshape future learning methods. By integrating with other advanced technologies such as AR/VR, big data, and explainable AI, AI voice assistants will be able to create a more intelligent learning environment, enabling learners to have a more efficient, personalized, and equitable educational experience. As time goes by, the world is constantly developing in a positive direction. Currently, the development of AI voice assistants has reached a relatively advanced stage, but there are still some unresolved issues, as well as some shortcomings and limitations. In summary, AI voice assistants hold great potential for future development and application. In the near future, artificial intelligence and AI voice assistants will continue to improve.

Combining learning methods with artificial intelligence not only aligns with the trend of the times, but also provides learners with an innovative, efficient, and personalized learning path and method. Data privacy and security improvements: With the large amounts of data collected and processed by AI voice assistants, further optimization is needed for privacy protection and ethical issues. Voice emotion analysis and personalized learning: Through voice emotion analysis technology, AI voice assistants will be able to recognize learners' emotional states and adjust teaching strategies in real time. With the development of deep learning technology, natural language processing will become more precise, enabling a smoother human-machine conversation experience. AI voice assistants will better understand user emotions through affective computing, providing more humanized interactive services.

References

[1] Sun Q. AI Voice Assistants and Intelligent Tutoring: Impact on and Optimization of Students' Learning Behavior. Research on Artificial Intelligence Education. 2025, 1(1).

[2] He L. Innovation and challenges brought by AI in advanced Chinese language classes at Japanese universities. Aichi Shude's Undergraduate Essays Collection. 2025, 15(3).

[3] Wang J K, Qin D H, Bai F B, Li L L, Kong L R, Xu C. Overview of research on the integration technology of speech recognition and large language models. Computer Engineering and Applications. 2025, 61(6).

[4] Ji S L, Wang Q Y, Chen A Y, Zhao B B, Ye T, Zhang X H, Wu J Z, Li Y, Yin J W, Wu Y J. State-of-the-Art survey of Open-source Software Supply Chain Security. Journal of Software. 2021.

[5] Tang F. Research on the Application of Intelligent Voice Technology in Call Centers. China New Communication. 2021, 23(17).

[6] Yang C Y, Zhuang L. Speech synthesis technology and its application in financial scenarios. China Financial Computer. 2021 (6).

[7] Wu R. Implementation of an automatic customer marketing system based on intelligent voice calling. 3D Imaging Technologies—Multi-dimensional Signal Processing and Deep Learning. 2025 (3).

[8] Ahmed A. A Critical Review of AI Voice Assistants' Role in Enhancing EFL Learners' Willingness to Communicate. Communicative figurations. June 02, 2025.

[9] Sun X X, Shen T Y, Jiang Q L and Jiang B. Research on the Impact of an AI Voice Assistant's Gender and Self-Disclosure Strategies on User Self-Disclosure in Chinese. Postpartum Follow-Up Phone Calls. 2025, 15(2).

[10] Huang L, He Y, Ma Z J, Mi H P, Yao Y. Analysis of research trends in smart home from the perspective of human-computer interaction. Journal of Computer Aided Design and Graphics. 2023 (2).

[11] Zhang L H, Lin Y. Application of Artificial Intelligence in Surgical Robotics. Chinese Journal of Medical Instrumentation. 2025.

[12] Tao J H, Chen J L, Li Y W. Overview of Speech Emotion Recognition. Signal Processing. 2023.