IntelliJoyCare: A Realistic Interactive Audiovisual System for AI-based Elderly Care Companionship

Wenjie Ma

Sichuan University, Chengdu, 610000, China

ma22912142@163.com

Abstract. In recent years, China's aging population has become increasingly severe, with the mental health of the elderly becoming a major concern in aging society. Data shows that 78% of the elderly believe that care from their children is the most effective way to alleviate depression. Based on this, this project is oriented towards the social hot topic of elderly mental health, rooted in the field of humanistic medicine, combining artificial intelligence technology with humanistic healing, and innovatively producing a smart app that integrates detection, treatment, and relief into one, breaking through the limitations of time and space. With the help of virtual human image technology, it builds a communication bridge between children and the elderly, fills the gap in the elderly's need for companionship, and assists in elderly mental health care.

Keywords: Aging, Mental health, AI technology, Intelligent interaction.

1 Introduction

In 2022, the National Health Commission issued a notice, deciding to extensively carry out actions for elderly mental health care across the country from 2022 to 2025, based on the organization and implementation of elderly mental health care projects. The country's attention to the mental health of the elderly continues to rise. The rapid development of AI technology empowers the field of humanistic medicine, facilitating the exploration and solution of elderly mental health issues through intelligent channels. The project aims to create a new humanistic therapy for emotional issues in the elderly, achieving dual goals of emotional connection and medical assistance.

2 Innovative Background

2.1 Support from Elderly-related Policies

The country's emphasis on the mental health of the elderly has been increasing. As early as 2019, the "Notice of the General Office of the National Health Commission on Implementing the Elderly Mental Health Care Project" required continuous improvement of the mental health of the elderly. In 2022, the National Health Commission issued a

Z. Zhan et al. (eds.), *Proceedings of the 2024 10th International Conference on Humanities and Social Science Research (ICHSSR 2024)*, Advances in Social Science, Education and Humanities Research 858,
https://doi.org/10.2991/978-2-38476-277-4_94

notice, deciding to extensively carry out actions for elderly mental health care across the country from 2022 to 2025, requiring at least one community or village in each county (city, district) nationwide to have a point of care for elderly mental health care by the end of the "14th Five-Year Plan" period.

The "Notice of the State Council on Printing and Issuing the '14th Five-Year Plan' for National Aging and Elderly Care Services" emphasizes that China's elderly population is large, aging rapidly, and the demand structure of the elderly is shifting from survival-oriented to development-oriented. There are still imbalances and inadequacies in elderly care services and aging services, particularly in rural areas, insufficient home and community-based elderly care, shortage of professional talents especially nursing personnel, the need for strengthening scientific and technological innovation and product support, and the need to enhance the coordinated development of the industry. The country strongly supports the development of technology-assisted elderly care, and the project aligns with this major trend.

2.2 Favorable Development Trend in the Digital Medical Industry

Digital medicine refers to the application of modern digital information technology in the medical process, which is the development direction and management goal of public healthcare. Analysts from the China Industry Research Institute predict that the global digital medical market size will reach $286.35 billion in 2023 and $365.67 billion in 2024. Currently, the digital medical industry chain from upstream to downstream has formed a relatively mature operating model, and digital medicine will continue to develop further on this basis.

2.3 Space for Development in the Combination of AI Technology and Medicine

Artificial intelligence has developed rapidly in recent years and has been widely used in various fields such as medicine, computer science, and neuropsychology due to its strong interdisciplinary nature. Nowadays, the combination of artificial intelligence and medicine mainly manifests in areas such as image recognition, intelligent diagnosis, personalized medicine, and the development of medical robots, providing advanced technological support and infinite possibilities for the future of medicine. However, the integration of artificial intelligence with clinical practice still requires further exploration. This research and development focus on integrating AI technologies such as robotics, speech recognition, and image recognition into the medical treatment process, providing humanistic care, assisting in alleviating negative emotions related to psychological health in the elderly such as anxiety and unease, and thereby complementing targeted medical interventions to achieve therapeutic effects and improve quality of life.

3 Pain Points and Needs Analysis of Elderly Mental Health Problems

3.1 Effective Ways to Alleviate Depression as Perceived by the Elderly

Analyzing the needs of the elderly from the perspective of Maslow's hierarchy of needs, they can be roughly summarized into five aspects. Due to their relatively detached social reality, the elderly have needs in physiological, safety, social, esteem, and self-actualization domains. Among them, social needs encompass various types of emotional relationship fulfillment. According to the "2022 National Depression Blue Book" jointly created by the People's Daily Health Client, 78% of elderly patients believe that care from their children is the most effective way to alleviate depression.

3.2 Information Needs of Children about the Health Status of the Elderly

Surveys show that the empty-nest rate of the elderly in China is approximately 83%[1]. Due to the rapid development of society, the culture of competition, and increasing societal pressures, it has become challenging for children and descendants to provide sufficient companionship to their family members. This leads to a contradiction between the psychological needs of the elderly and the objective life realities of their children and descendants. Children cannot always access detailed information about the health status of the elderly. Filial piety, as one of the core elements of Chinese traditional virtues, has been an integral part of Chinese historical development. This project directly addresses these pain points by providing children with channels to understand the health status of the elderly in practical terms and helps in emotional terms by facilitating the transmission of filial piety and emotional bonding.

4 Project Introduction

4.1 Sound Therapy and Comfort Care

4.1.1 Sound Therapy

Sound therapy, originating from the Yellow Emperor's Inner Canon, specifically mentions "Gong, Shang, Jiao, Zhi, Yu" corresponding to the five organs of the human body based on the Five Elements theory. Different sound attributes have different effects on the body[2]. In later developments, examples of sound therapy applications abound, as seen in "Meridians, Chapter Ten," which describes the effects of different sounds on the body. Modern sound therapy has been widely used clinically, such as in the treatment of tinnitus, vocal fold nodules (VFN), speech disorders, and stimulating comatose patients, as well as preventing delirium in ICU continuous blood purification patients[3].In this project, incorporating sound therapy enhances the physical health of the elderly and helps them resist the risk of illness.

4.1.2 Comfort Care

Comfort Care Comfort care, known as "hospice care" in Europe and the United States and translated as "Caring Care," "End-of-life Services," or "Comfort Care" in Singapore[4], Taiwan, and mainland China, is now officially referred to as "comfort care" as per the "Practical Guidelines for Comfort Care (Trial Implementation)" issued by the National Health Commission in 2017. It encompasses end-of-life care, palliative care, and relief treatment, focusing on terminally ill patients and their families, providing multidisciplinary care, physical, psychological, and spiritual support, pain and symptom management, improving quality of life, ensuring a peaceful and dignified death, and achieving peacefulness for the departed, peace of mind for the living, and comfort for the observers. The Worldwide Palliative Care Alliance (WPCA) designated Singapore and China's comfort care services at level 4a in the "World Atlas of Palliative Care," indicating the gradual integration of comfort care into mainstream medical services. This project aims to help the elderly enjoy their later years, promote the development and widespread adoption of comfort care in China.

Applying appropriate comfort care interventions can alleviate varying degrees of anxiety and pain. Digital health is widely used in comfort care, but there is still room for improvement domestically. This project strengthens the combination of modern technology and comfort care, raises awareness among patients and society, and provides compassionate care. Recent research at the Southern Medical University Affiliated Hospital indicated that the positive impact of comfort care on reducing pain, improving emotional responses, and enhancing satisfaction levels[5].

4.2 Product Features

4.2.1 Generating Unique Sounds and Real-time Interaction with Elderly Individuals

The app can train based on voice and image materials uploaded by users to generate personalized voice and animated virtual characters for communication. Compared to traditional treatment methods and medical hardware, the app's integration into smart devices makes it more portable and easy to use. With the widespread use of smartphones today, even the elderly can easily grasp how to use it, enabling them to "chat anytime, anywhere." Furthermore, the question-and-answer function simulating the voice and appearance of loved ones greatly enhances the app's interactivity, partly compensating for the regret of not having loved ones or friends constantly present.

4.2.2 Collecting Data, Monitoring, and Improving Elderly Mental Health

The app will collect and analyze user data during communication through professional psychological analysis, extract key issues, provide timely feedback to the backend and their children, and analyze changes in users' mental health status and other multidimensional data features. This data will be visualized and analyzed to research and deduce universal therapeutic improvement strategies. It can also provide timely guidance and assistance to elderly individuals struggling with depressive emotions when they need emotional care, bridging the gap caused by geographical distance and alleviating their pessimistic and negative emotions to some extent. Additionally, it allows the elderly's children and family members to better monitor their daily lives, pay attention to their physical and mental health, and take timely measures against diseases caused by negative emotions, thereby creating better conditions for their later years.

4.3 Introduction of Innovations

4.3.1 Integration of AI Technology with Humanistic

Medicine Building upon the BERT-ViT2 model, Sadtalker image virtual digital human technology, So-vits voice model training technology, and corpus analysis large model, this product further integrates with medical humanities research, actively developing AI-driven therapy for elderly healing through human-machine interaction.

4.3.2 Corpus Database

Collecting dialogues between multiple elderly individuals and their families to create a dataset, conducting corpus analysis, generating a large corpus model through machine learning techniques, and training the chatbot in various scenarios to establish its identity. The corpus database will provide data for related companies in the elderly care industry, promoting industrialization in elderly care.

4.3.3 Conveying Elderly Needs

Based on the highlighted social background of elderly depression and the limitations in time and space for children to communicate and connect emotionally with their elders, this project is the first to apply AI technology to elderly care, creating an emotional connection exclusively between the elderly and their children in real space. The daily lives, emotional needs, and health data of the elderly will be transmitted to their children through the app.

4.4 Advantages Analysis

4.4.1 Strong Interaction with AI Technology in Medical Therapy

Considering the target audience as the elderly, the product's design focuses on elder-friendliness as a breakthrough point. The product reduces the difficulty for the elderly to use the app through simple interfaces, large fonts, and audio accompaniment.

This product combines artificial intelligence technology with large-scale corpus generation models to create an online communication method that is distinct from flat and single-dimensional communication. It maximizes the reproduction of the voices and appearances of elderly individuals and their children, enhancing the user experience.

4.4.2 Personalized Customization, Targeted Companion

This product creates a unique communication mode for each user, focusing on personalized companionship from support to assistive therapy, fully implementing personalized care. It combines targeted medical methods to achieve therapeutic effects and improve quality of life.

In addition to providing compassionate care, the product will also interface with medical databases to evaluate the physical and psychological health status of elderly individuals. It will provide health tips based on their status and report assessment results to the backend, providing data for developing different health care plans for elderly residents.

4.5 Application Technologies

4.5.1 So-vits Voice Model Training Technology

So-vits-svc (also known as Sovits) is an artificial intelligence voice conversion software developed based on projects like VITS, SoftVC, and VISinger2. It uses the SoftVC content encoder to extract source audio speech features and inputs them into VITS (a high-performance speech synthesis model combining variational inference, standardization flow, and adversarial training). Through latent variables rather than spectral concatenation, it links acoustic models and vocoders, supporting both cloud and local training to replace original text with synthesized human voices. For local training, remove accompaniments and reverberations from audio materials using Ultimate Vocal Remover v5 for over half an hour, slice voice materials into 2-15s segments, package them into a dataset, and then import them into Sovits for training[6].

After training, this project selects the trained model, imports the desired original voice audio for inference conversion, and obtains the required voice model.

4.5.2 Sadtalker Image Virtual Digital Human Technology

Sadtalker is an open-source model that automatically synthesizes animations of characters speaking based on image and audio files. The model recognizes and applies facial movements such as opening mouths, blinking, and moving heads based on provided image and audio material, generating 3D motion coefficients (head posture, expressions) for 3DMM from audio. It implicitly modulates a novel 3D perceptual facial rendering to create head motion videos for speech. Sadtalker explicitly models audio and different types of motion coefficient connections, proposes the ExpNet model, and learns facial expressions from audio by extracting motion coefficients and 3D rendering facial movements. Sadtalker synthesizes different styles of head movements through PoseVAE.

4.5.3 BERT-ViT2 Model

BERT-ViT2 is an advanced cross-modal model that further extends the capabilities of BERT-ViT. It combines the latest natural language processing and computer vision technologies to achieve more precise and rich interactions between text and speech. In usage, developers prepare corresponding voice material data, train and infer using the

BERT-ViT2 model, and parse the model's output to obtain the desired task. Finally, based on the input text information, they obtain real, fluent, and fast speech audio output.

4.5.4 Corpus Analysis for Generating Large Models

Firstly, collect language data generated by participants interacting with AI humanoid robots. Then, use machine learning methods for pre-training to learn language patterns and generate unique contextual neural network models through supervised fine-tuning and command fine-tuning, aiming to make AI robots respond personalized and specialized. Simultaneously, cloud storage of participants' (elderly individuals with mental health issues) language data is used to create a corpus database, facilitating effect tracking, post-optimization, and providing crucial evidence for future potential problem-solving conversations.

4.5.5 Voice Emotion Recognition

Nowadays, there is a trend towards multidisciplinary integration in voice depression detection technology, with machine learning methods beginning to be applied in voice depression detection, which is expected to drive solutions to related problems. Common modalities for depression detection include electroencephalography (EEG), imaging, text, and voice signals. Among these, voice signals have the advantage of being easily obtained and having fewer usage restrictions, making research on depression detection based on voice signals a current hot topic. The project applies voiceprint recognition technology and attempts to analyze the mood index of elderly people through multiple dimensions such as intonation, speech length, and problem descriptions based on their voice during conversations[7].

Voice emotion recognition is crucial for human-computer interaction. The project plans to introduce a Dynamic Window Transformer (DWFormer), which is a new architecture that utilizes the importance of time by dynamically splitting samples into windows. Existing experimental results have shown that the model's performance with DWFormer is better than previous state-of-the-art methods, leading to further improvements in the accuracy of voice emotion recognition[8].

Additionally, the project will introduce Vision Transformers (ViTs) to replace traditional CNNs for Speech Emotion Recognition (SER), making it easier to train when sample data is scarce. Subsequently, the model will be extended to speaker VGG CCT, which utilizes the ViT self-attention mechanism in the architecture. This mechanism uses the representation of the speaker (speaker embedding) to partially compensate for the fact that each person may express emotions in different ways[9].

4.6 Areas for Project Improvement

4.6.1 Technological Challenges

The primary technological challenge of this project lies in the feasibility of voice training. Designing and implementing an operational program that closely matches the voices of elderly individuals and their children, and actually achieves the desired effect, requires multiple technical adjustments and support for the base model. Furthermore, the corpus needs to consider special cases such as dialects. Elderly individuals from different regions may communicate in local dialects, necessitating personalized customization to meet their needs effectively.

4.6.2 Lack of Relevant Legal Safeguards and Constraints

As a new attempt in the emerging field of combining AI technology with medicine, ethical issues are of paramount importance. Nowadays, with the proliferation of malicious incidents due to information leaks, users naturally worry about issues like personal information leakage when providing necessary information. Additionally, most people tend to adopt a wait-and-see attitude towards emerging technologies. How to use emerging AI technology to build trust between businesses and users is also worth considering. Developers should apply professional expertise to ensure the normal operation of procedures, strengthen confidentiality of personal information, enhance product usability, aesthetics, convenience, and humanization in design and production. Relevant legal provisions should also be gradually established and improved.

5 Application Scenarios

This project will start as a pilot in Chengdu and then gradually expand to western regions and nationwide. Leveraging partnerships with multiple nursing homes like Chengdu Taibao Home, the project will initially be launched and tested in Chengdu Taibao International Nursing Community in Sichuan Province. After the trial, relevant experimental data will be collected, and feedback from elderly users will be recorded. Based on the trial results, improvements and enhancements will be made to the product's functionality, appearance, and usage patterns to better align with user experiences. Subsequently, building on breakthrough achievements, the project will collaborate with other nursing homes in China, sign strategic agreements, and elevate the development of humanistic medicine to new heights and broader scope.

6 Conclusion

Humanistic medicine is emerging in the era of artificial intelligence, bringing both opportunities and risks. During the project's advancement, technological and ethical challenges are critical issues beyond the humanistic medical theme. However, with the stable development and favorable trends of smart healthcare, the future of humanistic medicine holds limitless possibilities.

References

1. Wang Ziyuan. Research on the intelligent companion product design of "Youyi" empty nesters[D]. Wuhan Textile University, 2023. DOI: 10.27698/d.cnki.gwhxj. 2022.000386.
2. YAO JACQUELINE. Research on the origin of the five notes "gong, shang, jiao, zhi, yu" and the issue of the five tones in the Yellow Emperor's Inner Canon[D]. Shanghai University of Traditional Chinese Medicine, 2023. DOI: 10.27320/d.cnki.gszyu.2020.000879.
3. Yue Yan, Lin Lunwer, Liao Shungi, Deng Ming, Lu Lei, Xiao Guojin. Effect of "Wu Yin An Shen" of Choosing Time Technique on Delirium and Sleep in Intensive Care Unit Patients with Contin-uous Blood Purification[J]. J Chengdu Univ TCM, 2023, 46(04):53-57.
4. GAO X Y, HU L P, ZHAO Y, et al. Palliative care service development experience in Singapore and its implications for China[J]. Chinese General Practice, 2023. [Epub ahead of print].
5. WANG Xiaoqiu, SU Xiaomin, CHEN Jinying, YE Wanling, LIU Qiaoyi, LIAO Jingsheng, HE Miaozhu, ZHANG Lin. Application effect of community hospice care in end-stage cancer patients[J]. Contemporary Chinese Medicine, 2023, 30(18):184-188+192.
6. ZHANG W, CUN X, WANG X, et al. SadTalker: Learning Realistic 3D Motion Coefficients for Stylized Audio-Driven Single Image Talking Face Animation[J]. 2022.
7. LIU Zhentao, XIANG Chunni, LIU Chenling, ZHONG Baoliang, HUANG Hai, PENG Zhikun, LYU Zhu, DING Zhong. Survey on Depression Detection Research Based on Speech Signals[J]. JOURNAL OF SIGNAL PROCESSING, 2023, 39(4): 616-631. doi: 10.16798/j.issn.1003-0530.2023.04.003.
8. CHEN S, XING X, ZHANG W, et al. DWFormer: Dynamic Window transFormer for Speech Emotion Recognition[J]. 2023.
9. AREZZO A, BERRETTI S. SPEAKER VGG CCT: Cross-corpus Speech Emotion Recognition with Speaker Embedding and Vision Transformers[J]. 2022.