Validation of a Thai artificial chatmate designed for cheering up the elderly during the COVID-19 pandemic
[version 3; peer review: 1 approved, 2 approved with reservations]

Somrudee Deepaisarn¹, Ek-uma Imkome², Konlakorn Wongpatikaseree³, Sumeth Yuenyong³, Ploi Lakanavisid⁴, Rangsiman Soonthornchaiva², Panida Yomaboot⁵, Angkoon Angkoonsawaengsuk¹, Napawan Munpansa⁵

¹Thammasat University, Klong-luang, Pratumtane, 12120, Thailand
²Department of mental health and psychiatric nursing, Faculty of Nursing, Thammasat University, Klong-luang, Pratumtane, 12120, Thailand
³Department of Computer Engineering, Mahidol University, Salaya, Nakhon Pathom, 73170, Thailand
⁴Faculty of Medicine, Burapha University, Chon Buri, Chon Buri, 20131, Thailand
⁵Department of Psychiatry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok Noi, Bangkok, 10700, Thailand

Abstract

Background

The COVID-19 pandemic severely affected populations of all age groups. The elderly are a high-risk group and are highly vulnerable to COVID-19. Assistive software chatbots can enhance the mental health status of the elderly by providing support and companionship. The objective of this study was to validate a Thai artificial chatmate for the elderly during the COVID-19 pandemic and floods.

Methods

Chatbot design includes the establishment of a dataset and emotional word vectors in which data consisting of emotional sentences were converted into the word vector form using a pre-trained word2vec model. A word vector was then input into a convolutional neural network (CNN) and trained until the model converges using sentence embedding and similarity word segmentation. Sentence vectors were...
generated by averaging each word vector using an averaged vector method. For approximate similarity matching, the Annoy library was used to create the indices in tree sorting. Data were collected from 22 elderly and assessed by the Post-Study System Usability Questionnaire (PSSUQ).

**Results**

The study revealed that 72.73% of the respondents found the chatbot easy to learn and use, 63.64% of the respondents found the chatbot can autonomously determine the next course of action, and 59.09% of the respondents believed that troubleshooting guidelines were provided for overcoming errors. The accuracy of the chatbot providing a reasonable response is 56.20±13.99%.

**Conclusions**

Most users were satisfied with the chatbot system. The proposed chatbot provided considerable essential insights into the development of assistance systems for the elderly during the coronavirus pandemic (COVID-19) and during the period of national disasters. The model can be expanded to other applications in the future.

**Keywords**
artificial chatmate, Elderly, mental health, COVID-19, health promotion
Corresponding authors: Ek-uma Imkome (ek-uma@nurse.tu.ac.th), Konlakorn Wongpatikaseree (Konlakorn.won@mahidol.ac.th)

Author roles: Deepaisarn S: Data Curation, Investigation, Methodology, Validation, Writing – Review & Editing; Imkome Eu: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Wongpatikaseree K: Data Curation, Formal Analysis, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; Yuenyong S: Formal Analysis, Software, Writing – Original Draft Preparation, Writing – Review & Editing; Lakanavisid P: Data Curation, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; Soonthornchaiva R: Supervision, Validation, Writing – Review & Editing; Yomaboot P: Conceptualization, Investigation, Methodology, Validation, Writing – Review & Editing; Angkoonsawaengsuk A: Formal Analysis, Validation, Writing – Review & Editing; Munpansa N: Validation, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Copyright: © 2024 Deepaisarn S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Deepaisarn S, Imkome Eu, Wongpatikaseree K et al. Validation of a Thai artificial chatmate designed for cheering up the elderly during the COVID-19 pandemic [version 3; peer review: 1 approved, 2 approved with reservations] F1000Research 2024, 11:1411 https://doi.org/10.12688/f1000research.127431.3

First published: 30 Nov 2022, 11:1411 https://doi.org/10.12688/f1000research.127431.1
Introduction

The COVID-19 pandemic has severely affected global public health, economics, education, and society. During the pandemic, technology’s use for companionship attracted considerable research attention. In addition to physical healthcare, ensuring the population’s mental health during the pandemic is critical. Therefore, this research developed personalized healthcare systems to support the elderly in their daily tasks and promote active and healthy aging. Validity, reliability, and sustainable measures are essential for the healthy aging of the population (Martinho et al., 2022).

As a result of the impact of COVID-19 outbreaks in Thailand since early 2020, the Ministry of Higher Education, Science, Research and Innovation promoted the health status and economic underpinnings of national strategic projects by using integrated development schemes related to four elements: a) creation of health technologies and applications for promoting community health, b) promotion of goods in community, c) promotion of tourism and d) promotion of circular economy. For this strategy, Thammasat University has applied the technology, knowledge, and expertise to help people in the healthcare domain. Every risk is assessed considering possible effects so that the risk can be addressed appropriately. The pilot project reveals that an efficient health technology system should be developed to deduce strategies to engage, persuade, and motivate older adults to promote positive behavioral changes and increase their awareness of other health conditions. The state quarantine and the physical distancing regulations of the first wave of the pandemic in Thailand significantly impact the elderly’s mental health status. These issues need to be addressed.

Considering the Thai culture, older adults with problematic health conditions are typically assumed to spend most of their time alone in a private household, resulting in loneliness. Such people depend on other family members and require personal assistance to complete daily tasks and general chores. Thus, technological influence is essential in improving the preventive measures against mental health problems in the elderly population.

Artificial Intelligence (AI) has been incorporated into many sciences and arts fields, including healthcare and support systems (Kusal et al., 2022). Furthermore, AI technologies have been used in hospitals as well as health and well-being sectors for mental health screening (Otero-González, Pacheco-Lorenzo, Fernández-Iglesias, & Anido-Rifón, 2024). The implementation of AI in healthcare can provide cost-effective solutions for improving the health of people in modern society (Väinänen, Haataja, Vehviläinen-Julkunen, and Toivanen, 2021). A chatbot can interact with the elderly and clarify their needs and emotions across the database and analytic system. Chatbots can replace the workforce required in many sectors, including chat. Chatbot exhibits considerable potential for integration into clinical practice and healthcare services. In medical and health research, chatbots have been used for food suggestions (Fadhil, 2018), smoking cessation (Wang et al., 2018), and cognitive behavior therapy (Fitzpatrick et al., 2017). Similar approaches have supported focused groups with health problems (Kumar and Rosé, 2014). Chatbot assists young adult or teenager groups in answering socially sensitive questions related to sexual health, drugs, and alcohol consumption. Users can “talk” to a programmed chatbot without fewer concerns and anxiety about sharing personal information with others (Crutzen et al., 2011). A well-designed chatbot can also give accurate guidance to the users. Thus, the first barrier to be overcome is the user's trust in the system. This can only be achieved if the system can ensure that the private information is non-identifiable and non-disclosable by any other means apart from undergoing a screening procedure that benefits the user and is authorized by that user only.

Furthermore, chatbot systems provide a considerable advantage to people who want to obtain health information compared with conventional search engines (Bickmore et al., 2016) for enhancing their understanding of their health conditions, improving health-related behaviors, and making decisions for treatments.

Chatbot applications in primary healthcare are more cost-effective than other solutions, such as medical nursing and medical robots. Many people/patients can access such tools using various chatbot platforms, including text and voice interfaces (Kreps and Neuhauer, 2010). Voice interfaces are more efficient, especially for mental health counseling.
Such systems provide suggestions for improving users’ health and quality of life. Therefore, in this study, “Ai-aun”, the artificial chatmate, was developed to promote the health of elderly people, especially during the COVID-19 pandemic and floods. The chatbots assess the similarity of the content cited in a spoken sentence and identified as an essential part of speech. Ai-aun can help the elderly adapt to stressful situations during the COVID-19 pandemic by maintaining good mental health, promoting healthy routines, and preventing dementia. The objective of this study was to validate the Thai artificial chatmate, “Ai-aun”, for the elderly people during the COVID-19 pandemic and during common disasters, such as floods. However, the use of the chatmate is not limited to such emergencies.

Methods
Chatbot architecture
In this study, a natural interface was used. The focus was on developing computers to understand human nature. Originally developed for linguistics, psychology, and computer science, natural interfaces can be applied to various fields where humans can conveniently communicate with computers/machines and convey empathy. Virtual reality is creating a virtual image or a simulation of an event using a computer system based on the interpretation of an ordinary human language. In this work, the development of the voice interface system allows humans to speak or operate efficiently with mobile devices.

Chatbot mechanism can respond to and aid users in comprehending the level of their health status and assist the users with the decision-making process to take care of themselves and improve their health. In such systems, user voice data are the inputs (i.e., voice interface). Next, the input is translated into text information. The association between the input text and the database is mapped using a trained model. This model is used to produce the output in the form of text. This result is translated back into voice and promptly conveyed to the user. The process is repeated for all user-triggered inputs, and each triggered input is regarded as an independent conversation.

The Ai-aun chatbot was developed using natural language processing, virtual reality, and hybrid AI systems. Trust in the system, and its usability is the central cores of this chatbot. In this study, a novel framework was created for the chatbot to produce a reasonable response to a user's message regarding content and emotional characteristics. To build the Ai-aun chatbot, three consecutive modules: namely emotional word vector, sentence embedding and similarity, and approximate similarity matching, were adopted. Figure 1 shows the framework with all components of the chatbot. The elderly can take care of themselves in six general areas: general health, medication, diet, exercising, sleep and rest, and family relationship. The chatbot can provide timely reminders to the elderly to take their medicine, exercise, and sleep to ensure they stay in good health, especially during the COVID-19 pandemic.

Algorithms and system design
The chatbot responds to the elderly user’s conversation to improve their mental health. The database system records and monitors all input information so that the authorized administrators can access it. The system can generate a report for administrators for analysis. After using the system, users are asked to complete a trust survey containing 15 questions to understand their trust, confidence, and the quality of conversation. The system performance was assessed using Qualtrics

Figure 1. Conceptual framework of the chatbot.
An effective interface should provide simple and easy-to-use features, well-structured, and robust outputs. Higher scores indicate higher trust and usability levels. This is the conventional way of validating a chatbot. A statistical technique was also applied to assess the proposed artificial chatmate reliably. Ai-aun comprises three modules: emotional word vector, sentence embedding and similarity, and approximate similarity matching.

**Module 1: Emotional word vector**

To convert a word into a vector for further analysis, data pre-processing and word2vec training are required.

The data pre-training was obtained from the 30 elderly volunteers on the U2T project. The research teams checked for accuracy (general health; 193 data sets; family; 115 datasets; medication; 122 datasets; diet; 137 datasets; exercise; 118 datasets; sleep and rest and other activities; 151 datasets). After obtaining the accurate data set, the researcher teams submitted data to the three experts for the process of content validity testing. The dataset was addressed as the expert comment for the training sessions.

In the pre-processing steps, raw input data are prepared. That is, text messages are proposed into a suitable format for analysis. These steps include word segmentation and stop word removal because the sentences were appropriate for content validity by three experts and the research teams before launching the language in NLP. Additionally, we need to check and remove some sentence errors during text pre-processing. This process was adjusted to suit the structure of the Thai language. Note that the variation of the user’s pronunciation can result in misinterpretation by the model and cause unintended outputs (i.e., amiss expression of emotion). This can be prevented with a proper pre-processing of data.

- **In Word segmentation**, texts that appear in the sentence are divided into individual words. Binary classification is applied to search for a unique character as the starting point of each word in an input sentence. Cutkurn was selected as the Thai word tokenization library for constructing this system. Cutkurn is based on a recurrent neural network (RNN) model.

- **Stop word removal** eliminates the words of minor significance in the sentence so that these words do not affect the sentence’s overall meaning. Examples of Thai-specific words in this category include the prefix or suffix of phrases expressing politeness or emphasis, conjunctions, and some ignorable adverbs.

Next, the word2vec functions allow an appropriate model to be trained from the raw input by trying to map the vector representations of words using the process of word embeddings (Mikolov et al., 2013). Skip-gram and Common Bag of Words (CBOW) are used. Skip-gram considers the meaning of a word to predict multiple consecutive words and vice versa for the CBOW technique.

The visual information in conversational agents may take affective states and thus is an appropriate indication of emojis in texts. The objective is to use emojis in the study for better interpretability of emotional user states, especially for short polysemous phrases of both positive and negative emotions in many scenarios. Understanding willingness, mood, and affect will help conventional agents respond to reliability, validity, and appropriateness. Additionally, Emojis are users’ reported labels of mood and affect to convey emotions in the context of communication interactions.

The working process of the emotional word vector module consists of two main steps.

- **Step 1**: In this step, emotional information is inserted into the pre-trained word vector to add the semantic meaning of the emotion into the vector of individual words, typically pre-trained on a large, standard data corpus. Therefore, the syntax and syntactic aspects of each word are accurately captured, for example, in a way that the vector of the word “dog” is far away from the vector of the word “doctor”. However, the emotional dimension is not captured satisfactorily. Users find that the vectors for the words “sad” and “happy” are closer because these are adjectives. Thus, if the user applies this pre-trained vector to the emotional chatbot, the results will be inaccurate in terms of emotion.

- **Step 2**: Word vectors from the trained word embedded models were re-adjusted via a Convolutional Neural Network (CNN) to improve the emotional dimension that should be optimized for each word. The first layer and word embedded layer constitute the first module. The weight of the word embedded layer is continuously optimized until converged in order to determine the emotional dimension.
Module 2: Sentence embedding and similarity

The vector obtained from Module 1 is at the word level, which is converted into a vector of sentence embedding and similarity, as illustrated in Figure 2.

Words from the sentence were segmented using the Python library developed for Thai word segmentation, called “Cutkum”, which was developed by Treeratpituk (2017). Next, each word is converted into a word vector using the emotional word vector model created in Module 1. Then, a sentence vector is generated from averaged word vectors. Finally, the vector similarity between the input sentence and all the sentences in the response database is computed using cosine similarity according to Equation (1), where $A$ and $B$ are the sentence embedding vectors to be compared.

$$\text{Similarity} = \cos \theta = \frac{A \cdot B}{|A||B|}$$

Module 3: Approximate similarity matching

Similarity matching consists of generating an index followed by the sorted-tree search. A sentence vector is compared to the existing library. Hence, an accurate answer can be suggested. However, real-time response is required for a chatbot to be sufficiently fast. This research uses the approximate nearest neighbors (Annoy) library, where the vectors are rearranged in the response according to the database’s index. This exhibits a fast similar-items retrieval data structure that is ideal for use with chatbots with numerous sentences.

The process of maintaining the system to protect against errors

For the chatbot system and training, three experts prepared and validated content’s Ai-aun’s response to the older adults. Afterward, the research teams check the content validity before training the chatbot. The AI-AUN response was programmed and did not create a new sentence for a response while chatting with the older adults as the generative AI. Therefore, there are no errors in the sentence response, especially the medication sentence response. Additionally, the process of maintaining the system to protect the errors was: a) the elderly and the primary caregiver recorded the medication on the application and double-checked for non-high-alert drugs, and high-alert drugs were excluded from the study. The medication data was automatically backed up on the website, b) the research teams asked the primary caregiver to check for medication taken daily, and c) the researcher teams set system checks every week.

Noteworthy, the chatbot will issue alerts based on the caregiver’s time. There are two scenarios in which the reminder system may fail: a) the device is not connected to the Internet: the reminder will not be triggered if the device is not connected to the Internet, and b) the device is not turned on at reminder Time: If the device is not turned on at the scheduled reminder time, the alert will not activate.
Tool validation

Ethical approval

Ethical approval for the study was obtained from the Ethics Review Committee for Research Involving Human Research Participants, Thammasat University, Thailand (COA No. 115/2564) on November 4, 2021. The scope, risks, and benefits of the study were explained to all participants. Before data collection, written and verbal consent was obtained directly from participants. Participation was voluntary, and participant anonymity and confidentiality were ensured.

Data collection

The elderly people were selected as participants through purposive sampling on November 11, 2021. People aged above 60 years who speak Thai as their first language and can read Thai were selected. However, the ability to read Thai was a soft requirement because this study focused on communication through the voice interface. Those who could not participate in at least 80% of the experiment, ongoing high-alert drug therapy, and had health problems that hindered participation were excluded.

Twenty-two elderly participants, 18 women, and 4 men were selected from the population in Bang Krabue, Sam Khok, Pathum Thani, Thailand. The participants were interviewed to collect the sentences they would use to chat with the Ai-aun chatbot. Data collection was finalized with the saturation of the data. After that, the dataset was evaluated by the researchers (including computer scientists, physicians, psychologists, and psychiatric nurses) and submitted to three validators (i.e., a psychiatrist, a nurse who worked in the aging field, and a nursing lecturer in the aging field) to examine the content validity index. To qualify the datasets for the emotional word vector, sentence embedding and similarity word segmentation were performed to generate sentence vectors, followed by approximate similarity matching. Subsequently, the time spent with the chatbot, the rate of questions per minute, and the usability of the chatbot by the elderly were tested after 15 days of using the Ai-aun chatbot on a mobile device.

Data analysis

- Demographic data were recorded, including age, gender, marital status, education, career, income, personal illness, social networking sites, and time on social networks.

- A Post-Study System Usability Questionnaire (PSSUQ) was used to measure the usability variable, which consists of 15 questionnaires to assess users’ satisfaction with the chatbot interface in terms of language, graphics, and simplicity of the user interface (Lewis, 1995). Typically, PSSUQ scores range from 1 (strongly agree) to 7 (strongly disagree). The higher value means more excellent usability. The Content Validity Index (CVI) is 0.86.

- Training dataset for the conversational chatbot agent: A sample scheme of questions and responses was utilized for the dataset validation. An appropriate answer was selected from the volunteering elderly from the community for training and testing the chatbot model. The CVI for the dataset is 0.95.

- The time spent with the chatbot, recorded in the time log, was analyzed concerning the rate of questions per minute.

All parameters mentioned above were reported using descriptive statistics.

Technical performance test for accuracy

The accuracy of the chatbot providing reasonable responses was systematically tested by ten trained technicians. Each individual spoke 100 random questions based on how they would talk generally. The elderly conversations and answers from the chatbot were collected. One out of the four emotional states, i.e., positive, neutral, frustrated, and negative, was determined for each sentence. The predicted scores for such emotional states were calculated.

Results

The sample consisted of 22 elderly people. The mean age was 68.86±2.56 years old. Out of the total participants, 81.82% were women, and 18.18% were men. People with physical illnesses accounted for 72.73% of the samples. Diabetes mellitus is the most common chronic illness (45.45%) among the participants. All samples stay with their family members. In the daytime, the family members leave for work. 40.90% of them live alone and can perform daily activities.
The time spent with the Ai-aun chatbot was approximately 0.20-0.30 s, and the rate of the user asking questions to Ai-aun was 0.28 questions per minute. The Ai-aun chatbot had a moderate usability level (see the summary Table 1). After 15 days of using the Ai-aun chatbot, 59.09% of the participants were delighted with the system, indicating that it had all the functionalities and capabilities they expected. Additionally, the accuracy of the chatbot in terms of providing a reasonable response was 56.20±13.99%. The predicting scores for the emotional states ranged between 0.50 and 1.00, with the mean predicted score of 0.79±0.17.

The results revealed that 72.73% of the participants found the chatbot easy to learn and use, and 63.64% of participants thought the chatbot was sufficiently autonomous in determining its next course of action. In case of a system error, the system also provided troubleshooting guidelines according to 59.09% of participants.

| Variable                  | Measure                              | Result               |
|---------------------------|--------------------------------------|----------------------|
| 1. The time spent with the chatbot | Time log                            | ~0.20-0.30 s         |
| 2. Rate of questions per minute | Number of questions/times taken      | 0.28 min⁻¹           |
| 3. Usability              | PSSUQ                                | Moderate usability   |

Table 1. Variables, measures, and the results of the study.

The application can be downloaded from the Google Play Store.

Figure 3. Ai-aun (U2T, 2022).
Discussion

In this study, the Ai-aun chatbot was developed to cheer up elderly people during the COVID-19 pandemic. The time spent with the chatbot, the rate of questions per minute, and the usability measured using the PSSUQ were the variables used in this study. The time spent was 0.2-0.3 s. The average rate of questions was 0.28 min⁻¹. The dataset for the Ai-aun chatbot was designed for short spoken sentences as the input, and the average rate of questions per minute was fast. The usability was at a moderate level. The participants were happy that the Ai-aun system had all the functionalities and capabilities they expected from a chatbot, and the application device was easy to use. Furthermore, a young girl’s voice was adopted along with a friendly, easy-to-use interface (U2T, 2022). Therefore, familiarizing participants with the Ai-aun chatbot was essential. Elderly people were satisfied with the Ai-aun chatbot (Mehta et al., 2022). Moreover, a chatbot provides smooth conversation, including greeting, salutation, compromising, support, and guidance. For example, a simple sentence that sounds relaxing is added, such as “Have a nice day”, “Don’t worry…I will always be with you”, and “I love talking to you”. These sentences encouraged elderly people to interact with the Ai-aun. In addition, Ai-aun provides a feature for medication, rest, exercises, diet schedule, and reminders. The system was consistent with the mobile application developed by Puig (2021), +Approp, to support older and HIV-infected patients. It provides clinical history (e.g., vaccines and medication taken), educational feedback, and reminders of medical appointments. Positive responses were obtained regarding the usability and satisfaction of mobile healthcare applications.

The Ai-aun chatbot was developed by a paradigm suitable for accepting natural language input and processing computer instructions to enable users to achieve their health goals (Abdul-Kader and Woods, 2015; Fadhil, 2018; Kusal et al., 2022). Participants enjoyed chatting with Ai-aun (U2T, 2022) (Figure 3). A positive experience and familiarity with the mobile application result in the satisfaction of use.

This study proposed a mobile chatbot system to support elderly people, which encourages the elderly to access and get familiarized with the technology to support their physical and mental health. The agent provides proactive functions, sending messages to help elderly users educate and relax in general health, taking medication, diet, family relationships, exercise, rest, and sleep. Reminiscence strategies are considered challenging when related to design interfaces, conversation logic, and meaningful metrics directed at elderly adults, which benefit from screening and educating them (Kusal et al., 2022; Moreno, Sánchez-Anguix, Alberola, Julián, & Botti, 2024; Otero-González, Pacheco-Lorenzo, Fernández-Iglesias, & Amido-Rifón, 2024). Having enhanced the technological experience of the elderly, they can interact directly with Ai-aun confidently and successfully receive responses related to promoting their health conditions. Four emotional states of the conversation were determined to reflect the user’s actual sentiment. The effectiveness of the response from the chatbot system in terms of progress, interactions, participation, and satisfaction was assessed. Therefore, the system can be expanded to incorporate more functions that align with the users’ needs and applied to more participants.

For the current configuration, three consecutive modules: emotional word vector, sentence embedding and similarity word segmentation, and approximate similarity matching, were used to elicit a response from the database (Ng, Chia, Yap, & Goh, 2022). The accuracy of the chatbot providing a reasonable response was 56.20±13.99% which depends on the dialogue between the individual and chatbot. Such accuracy is typical for a chatbot that is expected to handle various sentences and sentiments. The accuracy can be improved by increasing the training dataset size to cover a broad variety of conversations. Based on this study, Ai-aun’s model – of the elderly language phenomenon should be studied further. The elderly’s visual and auditory impairment may be an obstacle to the kind of chatbot models.

Conclusion

This study revealed a moderate usability level. Several users were satisfied with the chatbot system, which provided basic information. They can be used as guidance in developing care and assistance systems for the elderly in the community during the coronavirus pandemic (COVID-19) and in national disasters, such as flood situations. Therefore, the chatbot can be potentially expanded for use in various scenarios. Future development of the Ai-aun chatbot is toward personalized healthcare and remote precision medicine for recognizing and monitoring the health status and behavior of the elderly. The chatbot should allow their family and the authorized caregivers to be in touch with the important updates related to the elderly health.

Future directions

In future work, the proposed chatbot system should be improved to consider higher-complexity conversations and to cover the requirements of people with more complicated behaviors, lifestyles, and health conditions. Various contexts and environments also affect the chatbot responses. We intend to incorporate the proposed chatbot with the improved models that can provide real-time data by directly monitoring the precise health conditions of individuals and the elderly. The system can be improved by (1) including additional users (family) that can provide additional content and rewards to the elderly person; and (2) establishing cognitive plans that can be evaluated.
Limitation of the study
The elderly may have benefited during the chatbot use, especially in physical and mental health aspects. These valid chatbots via mobile phone are very easy to learn and use, but the present study’s error tolerance rate, delays and bugs was not measured. This limitation existed but did not have a significant effect on the results. However, it is recommended to address these limitations in future studies.

Data availability
Underlying data
Figshare: Chatbot Ai-aun, https://doi.org/10.6084/m9.figshare.19744933.v15 (Imkome, 2022).

This project contains the following underlying data:

- Raw data.xlsx
- The training dataset for the chatbot.docx
- Demographic data.docx
- The results of the interviews.docx
- Source Code Aiaun.pdf
- AiAunApp
- ChitchatAiaun

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgments
This study was supported by Thammasat University Research Unit in the innovation of Mental Health and Behavioral Healthcare: staff, laboratory research room, and research equipment.

References
Abdul-Kader SA, Woods JC: Survey on Chatbot Design Techniques in Speech Conversation Systems. Int. J. Adv. Comput. Sci. Appl. 2015; 6(7). 2156-5270.

Bickmore TW, Utami D, Matsuyama R, et al.: Improving access to online health information with conversational agents: A randomized controlled experiment. J. Med. Internet Res. 2016; 18(1): e1. PubMed Abstract | Publisher Full Text

Corritore C, Marble R, Wiedenbeck S, et al.: Measuring online trust of websites: Credibility, perceived ease of use, and risk. Paper presented at the 11th Americas Conference on Information Systems, Omaha, Nebraska. 2005.

Crutzen R, Peters G-JY, Portugal SD, et al.: An artificially intelligent chat agent that answers adolescents’ questions related to sex, drugs, and alcohol: An exploratory study. J. Adolesc. Health. 2011; 48: 514-519. PubMed Abstract | Publisher Full Text

Fadhil A: Can a chatbot determine my diet: Addressing challenges of chatbot application for meal recommendations. Comput. Intell. 2018. arXiv:1802.09100. Reference Source

Fitzpatrick KK, Darcy A, Vierhile M: Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. J. Med. Internet Res. 2017; 4(2): e19. PubMed Abstract | Publisher Full Text

Google Code: word2vec. [cited 2020 Jul 17]. Reference Source

Imkome E: Chatbot Ai-aun. figshare. [Dataset]. 2022. Publisher Full Text

Judson TJ, Odisho AY, Young JJ, et al.: Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. 2020. Reference Source

Kolenik T, Gams M: Intelligent Cognitive Assistants for Attitude and Behavior Change Support in Mental Health: State-of-the-Art Technical Review. Electronics. 2021; 10(11): 1250. Publisher Full Text

Kreps GL, Neuhauser L: New directions in ehealth communication: Opportunities and challenges. Patient Educ. Couns. 2010; 78(3): 329-336. PubMed Abstract | Publisher Full Text

Kumar R, Rosé CP: Triggering effective social support for online groups. ACM Transactions on Interactive Intelligent Systems. 2014; 3(4): 1-32. Publisher Full Text

Kusal S, Patil S, Choudrie J, et al.: AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions. IEEE Access. 2022; 10: 92337-92356. Publisher Full Text

Lewis JR: Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Int. J. Hum.-Comput. Interact. 1995; 7(1): 57-78. Publisher Full Text

Liu WD, Chuang KY, Chen KY: The Design and Implementation of a Chatbot’s Character for Elderly Care. 2018 International Conference on System Science and Engineering (ICBBSE). 2018; pp. 1-5. Publisher Full Text

Manickam RN, Mu Y, Kshirsagar AV, et al.: Hospital readmissions reduction program (hrrp). 2019. Reference Source
Dear Editor,

I am grateful for the opportunity to evaluate the study that proposes a mobile chatbot system to support the elderly by encouraging them to access and familiarise themselves with technology to improve their physical and mental health.

The COVID-19 pandemic has affected physical health and disrupted daily life worldwide by changing individual and social attitudes and behaviors. In such conditions, technological applications are becoming mainstream worldwide to continue social life. Artificial intelligence in healthcare is one of the most popular topics today, and it can provide cost-effective solutions in healthcare. Moreover, chatbot systems offer significant advantages to individuals seeking health information than traditional search engines. They help people understand their health status, improve health-related behaviors, and make informed treatment decisions.

1. The study emphasizes the importance of elderly individuals in Thai culture. It was observed that older people with health problems in Thai culture usually spend most of their time in a private home, which leads to loneliness. Such people depend on other family members and require personal assistance to complete daily tasks and general chores. Therefore, the role of technology is crucial in developing preventive measures against mental health problems in the elderly population.

2. The study used the Post-Study System Usability Questionnaire (PSSUQ) to help users measure their satisfaction with the usability of computer systems and the structure of the preliminary principal factor analysis was confirmed.

3. It was determined in the results section that family members left home for work, and 40.9% of the participants stayed alone at home. After using the Ai-aun chatbot for 15 days, 59.09% of the participants reported being satisfied with the system, which shows that it meets their expectations. In addition, the reasonable response rate of the chatbot was found to be 56.20%. The results revealed that 72.73% of the participants found the chatbot easy to learn and use, and 63.64% thought it was sufficiently autonomous in determining its next course of action. This study shows a moderate usability level of the chatbot.

4. However, the study emphasizes that the elderly may have visual and auditory impairments, which may limit their usage of the chatbot. Also, the fact that the study did not evaluate
whether the elderly benefit, especially in terms of mental health, during chatbot use with the results of the stress scale before and after chatbot use can be counted among the limitations.

5. The clinical implications mention that the Ai-aun chatbot can contribute to personalized health services by being developed to recognize and monitor the health status and behaviors of the elderly for future development.

In conclusion, in addition to physical health during the pandemic, it is crucial to improve the mental health of the elderly, who are a more vulnerable group. This study will contribute significantly to the literature on developing care and assistance systems for the elderly in the community during disasters such as the COVID-19 pandemic.

Gözde Bacık Yaman

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Mental Health, mental disorders, stress, perceived stress, behavioural changes in COVID-19, videoconference fatigue at COVID-19, mental health in the elderly

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
Dear authors,

The reviewer provides the following comments:

(1) In this study, how did the author define the scope of the elderly? Although this article clearly states that the age is over 60 years old. However, is it suitable for older adults with dementia?

(2) In addition, does the subject live alone? What was their original living situation?

(3) The reviewer worries that chatbot reminders to take medicine may easily cause medical accidents if the system makes errors. How do you avoid the execution of non-safety projects?

That is all. Thanks

Is the work clearly and accurately presented and does it cite the current literature?  
Yes

Is the study design appropriate and is the work technically sound?  
Yes

Are sufficient details of methods and analysis provided to allow replication by others?  
Yes

If applicable, is the statistical analysis and its interpretation appropriate?  
Yes

Are all the source data underlying the results available to ensure full reproducibility?  
Yes

Are the conclusions drawn adequately supported by the results?  
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Smart healthcare, Web application, Human-computer interaction

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
Dear reviewer,

Thank you for your thought provoking review. The point-by-point responses are as follows:

1. The scope of older adults, which we defined on page 5, lines 29-33 of the data collection part as "People aged above 60 years who speak Thai as their first language and can read Thai, was selected. However, the ability to read Thai was a soft requirement because this study focused on communication through the voice interface. Those who were unable to participate in at least 80% of the experiment and had health problems that hindered participation were excluded."

This chatbot is not provided for older adults with dementia. The subjects of this study were healthy adults in the community, as we state on page 3, line no. 4-6, "Ai-aun can help the elderly adapt to stressful situations during the COVID-19 pandemic by maintaining good mental health, promoting healthy routines, and preventing dementia." We focus on preventing dementia.

2. The original living situation of the subject was added in the parts of the result as "All samples stay with their family members. In the daytime, the family members leave for work. 40.90% of them live alone and can perform daily activities."

3. We added the process of maintaining the system to protect the errors. We explained the chatbot safety in terms of the medication response and maintaining the system to protect against errors.

"For the chatbot system and training, three experts prepared and validated content's Ai-aun's response to the older adults. Afterward, the research teams check the content validity before training the chatbot. The AI-AUN response was programmed and did not create a new sentence for a response while chatting with the older adults as the generative AI. Therefore, there are no errors in the sentence response, especially the medication sentence response. Additionally, maintaining the system to protect the errors was: a) the elderly and the primary caregiver recorded the medication on the application and double-checked for non-high-alert drugs, and high-alert drugs were excluded from the study. The medication data was automatically backed up on the website, b) the research teams asked the primary caregiver to check for medication taken daily, and c) the researcher teams set system checks every week.

Noteworthy, the chatbot will issue alerts based on the caregiver's time. There are two scenarios in which the reminder system may fail: a) the device is not connected to the Internet: the reminder will not be triggered if the device is not connected to the Internet, and b) the device is not turned on at reminder Time: If the device is not turned on at the scheduled reminder time, the alert will not activate."

Additionally, we added "ongoing high-alert drug therapy" to the data collection parts.
Sinan Chen
Kobe University, Kobe, Japan

The reviewer appreciates the level of detail that the authors provide with accompanying data. Regarding the content described in this article, the reviewer made the following comments:

1. Is this conversation robot personalized? For example, how did you obtain pre-training data for emotional word vectors? Additionally, the author should provide a representative history of the conversation.

2. Although the author used PSSUQ to evaluate system usability, it is unclear which fine-grained aspects of improving users' mental health conditions are specifically reflected.

3. Regarding speech synthesis and recognition technology, what is the error tolerance rate? Will there be delays and bugs? How can the system provide 24-hour virtual care for the elderly who cannot use mobile phones?

That is all. Thanks

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.
Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Smart healthcare, Web application, Human-computer interaction

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

**Author Response 07 Jan 2024**

**Ek-uma Imkome**

Dear Dr. Sinan Chen,

We appreciate your time and valuable comments on our manuscript. We have carefully reviewed your suggestion. We provide a point-by-point response below:

(1) Is this conversation robot personalized? For example, how did you obtain pre-training data for emotional word vectors? Additionally, the author should provide a representative history of the conversation.

**Authors:** Thank you for pointing out the critical issue. Per your comment, we added more information on the revised manuscripts on page 6, line no. 19-24.

“The data pre-training was obtained from the 30 elderly volunteers on the U2T project. The research teams checked for accuracy (general health; 193 datasets; family; 115 datasets; medication; 122 datasets; diet; 137 datasets; exercise; 118 datasets; sleep and rest and other activities; 151 datasets). After obtaining the accurate data set, the researcher teams submitted data to the three experts for the process of content validity testing. The dataset was addressed as the expert comment for the training sessions.”

(2) Although the author used PSSUQ to evaluate system usability, it is unclear which fine-grained aspects of improving users' mental health conditions are specifically reflected.

**Authors:** Thank you very much for your concern. The mental health improvement state is shown in the research results of another manuscript that is in the process of publication.

(3) Regarding speech synthesis and recognition technology, what is the error tolerance rate? Will there be delays and bugs? How can the system provide 24-hour virtual care for the elderly who cannot use mobile phones?

**Authors:** Thank you very much. These valuable comments address in limitation of the manuscript as below
“The elderly may have benefited during the chatbot use, especially in physical and mental health aspects. These valid chatbots via mobile phone are very easy to learn and use, but the present study's error tolerance rate, delays, and bugs was not measured. This limitation existed but did not have a significant effect on the results. However, it is recommended to address these limitations in future studies”.

That is all. Thanks

**Competing Interests:** No competing interests were disclosed.

---

Reviewer Report 19 June 2023

https://doi.org/10.5256/f1000research.139936.r176964

© 2023 Kusal S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sheetal Kusal

Symbiosis Institute of Technology, Pune, 412115, India

The paper presents a Thai chatbot for the mental well-being of elderly people. The paper is very interesting to read. Although the authors have tried to make it technically sound, some points must be clarified to make the paper more technically sound.

Points to be discussed:
1. Some sentences in the paper are repetitive. Could you remove the repetitiveness of sentences?

2. Literature review must be done thoroughly. It is advisable to make an LR table. The authors can refer to the format mentioned in the citation¹.

3. In the text pre-processing, the authors have mentioned only segmentation and stop words removal. In NLP, text pre-processing involves many steps to make text appropriate for processing. So, have the authors performed them or which steps? And if not performed, why only stop word removal? Clarify.

4. The authors have mentioned emotion information inserted into the pre-trained word vectors. The authors are expected to elaborate more on how emotion information is inserted into pre-trained word vectors.

5. More discussion is expected on results and performance.

References
1. Kusal S, Patil S, Choudrie J, Kotecha K, et al.: AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions. IEEE Access. 2022; 10: 92337-92356 Publisher Full Text

**Is the work clearly and accurately presented and does it cite the current literature?**
Partly

**Is the study design appropriate and is the work technically sound?**
Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**
No

**If applicable, is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**
Partly

**Are the conclusions drawn adequately supported by the results?**
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Natural language processing, emotional detection, conversational agents, Artificial Intelligence.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

**Author Response 07 Jan 2024**

**Ek-uma Imkome**

Dear Dr. Sheetal Kusal,

Thank you for your time and valuable feedback on our manuscript. We have carefully reviewed your comments. As suggested, we provide a point-by-point response as below:

1. Some sentences in the paper are repetitive. Could you remove the repetitiveness of sentences?

   **Authors:** Thank you very much. We remove the repetitiveness of sentences on page 6 (line no. 4-6), page 10 (Line no. 4,5,7)

2. Literature review must be done thoroughly. It is advisable to make an LR table. The authors can refer to the format mentioned in the citation1.
Authors: Thank you for your attention. The Literature review is our work and on the process of publication.

3. In the text pre-processing, the authors have mentioned only segmentation and stop words removal. In NLP, text pre-processing involves many steps to make text appropriate for processing. So, have the authors performed them or which steps? And if not performed, why only stop word removal? Clarify.

Authors: Thank you very much. We addressed your comment in the Algorithms and System Design section (Page 6; Line no. 21-22 ) as below:

“because the sentences were appropriate for content validity by three experts and the research teams before launching the language in NLP. Additionally, we need to check and remove some sentence errors during text pre-processing”.

4. The authors have mentioned emotion information inserted into the pre-trained word vectors. The authors are expected to elaborate more on how emotion information is inserted into pre-trained word vectors.

Authors: Thank you very much. We added the inserted emoji into pre-trained word vectors on page 7 (Line no. 4-10)

“The visual information in conversational agents may take affective states and thus is an appropriate indication of emojis in texts. The objective is to use emojis in the study for better interpretability of emotional user states, especially for short polysemous phrases of both positive and negative emotions in many scenarios. Understanding willingness, mood, and affect will help conventional agents respond to reliability, validity, and appropriateness. Additionally, Emojis are users’ reported labels of mood and affect to convey emotions in the context of communication interactions”.

5. More discussion is expected on results and performance.

Authors: Thank you very much. We added new references on the part of the discussion to support our results in the text.

“The agent provides proactive functions, sending messages to help elderly users educate and relax in general health, taking medication, diet, family relationships, exercise, rest, and sleep. Reminiscence strategies are considered challenging when related to design interfaces, conversation logic, and meaningful metrics directed at elderly adults, which benefit from screening and educating them (Kusal et al., 2022; Moreno, Sánchez-Anguix, Alberola, Julián, & Botti, 2024; Otero-González, Pacheco-Lorenzo, Fernández-Iglesias, & Anido-Rifón, 2024)”.

References
1. Kusal S, Patil S, Choudrie J, Kotecha K, et al.: AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions. IEEE Access. 2022; 10: 92337-92356
2. Otero-González, I., Pacheco-Lorenzo, M. R., Fernández-Iglesias, M. J., & Anido-Rifón, L. E. (2024). Conversational agents for depression screening: A systematic review. International Journal of Medical Informatics, 181. https://doi.org/10.1016/j.ijmedinf.2023.105272

3. Moreno, J. C., Sánchez-Anguix, V., Alberola, J. M., Julián, V., & Botti, V. (2024). An intelligent conversational agent for educating the general public about HIV. Neurocomputing, 563. https://doi.org/10.1016/j.neucom.2023.126902

Competing Interests: The authors have not disclosed any competing interests.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com