

Open Access Full Text Article

ORIGINAL RESEARCH

Harnessing Technology for Pandemic Support: Developing a COVID-19 Chatbot in Vietnam

Exploiter la technologie pour soutenir la pandémie : développer un chatbot COVID-19 au Vietnam

Vinh Nguyen-Nhu^{1,2}, Tho Thanh Quan³, Thien Duc Duong¹, Son Trung Huynh¹, Ngan Thi Kim Le¹, Duy Phung Quoc Vo¹, Liem Thanh Nguyen¹, Bao Pham Gia Nguyen¹, Hue Thi Hong Nguyen¹, Phuc Hoang Vo¹, Tung Ngoc Nguyen¹, Chinh Quang Doan³, Minh Tran Cong Bui³, Sy Duong-Quy⁴

¹: Faculty of Medicine, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam

²: Department of Respiratory Functional Exploration, University Medical Center, Ho Chi Minh City, Vietnam

³: Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh City, Vietnam

⁴: Bio-Medical Research Center, Lam Dong Medical College, Da Lat, Viet Nam

ABSTRACT

Background. In the context of limited human resources during a pandemic, chatbots serve as vital tools for disseminating essential information. This study aimed to outline the design process of a chatbot tailored to the public's demand for timely and visually accessible COVID-19 information in Vietnam. **Methods.** Task-Oriented Dialogue Systems were used to build the Chatbot. After retrieving data according to the key label, Chatbot's content was curated and managed using a NoSQL database and MongoDB database management system. We incorporated Support Vector Machine, Named Entity Recognition and K - nearest neighbor to train the model. **Results.** The developed database comprises 12 modules containing over 1000 sets of questions and answers, addressing prevalent COVID-19 concerns. The Chatbot dataset was strictly reviewed and validated to ensure its accuracy and suitability for community education to be scientifically accurate and appropriate for community education. A preliminary survey involving 245 medical students revealed encouraging potential for the application of the Chatbot. **Conclusion.** We have initially constructed a chatbot containing comprehensive content pertaining to COVID-19. Despite limitations in financial and human resources, our Chatbot serves as a foundation for future research aimed at refining this tool to meet the demands for medical information search.

KEYWORDS: COVID-19; Chatbot; SARS-CoV-2; Pandemic; Vietnam.

RÉSUMÉ

Introduction. Dans le contexte de ressources humaines limitées en période de pandémie, les chatbots constituent des outils essentiels pour diffuser des informations essentielles. Cette étude visait à décrire le processus de conception d'un chatbot adapté à la demande du public d'informations COVID-19 opportunes et visuellement accessibles au Vietnam. **Méthodes.** Des systèmes de dialogue orientés tâches ont été utilisés pour créer le Chatbot. Après avoir récupéré les données selon l'étiquette clé, le contenu de Chatbot a été organisé et géré à l'aide d'une base de données NoSQL et d'un système de gestion de base de données MongoDB. Nous avons incorporé Support Vector Machine, Named Entity Recognition et K - le plus proche voisin pour entraîner le modèle. **Résultats.** La base de données développée comprend 12 modules contenant plus de 1 000 ensembles de questions et réponses, répondant aux préoccupations courantes liées au COVID-19. L'ensemble de données Chatbot a été strictement examiné et validé pour garantir son exactitude et son adéquation à l'éducation communautaire afin qu'il soit scientifiquement précis et approprié pour l'éducation communautaire. Une enquête préliminaire auprès de 245 étudiants en médecine a révélé un potentiel encourageant pour l'application du Chatbot. **Conclusion.** Nous avons initialement construit un chatbot contenant un contenu complet relatif au COVID-19. Malgré les limites des ressources financières et humaines, notre Chatbot sert de base à de futures recherches visant à affiner cet outil pour répondre aux demandes de recherche d'informations médicales

MOTS CLÉS: COVID-19; Chatbot; SARS-CoV-2; Pandémie; Vietnam.

Corresponding author:

Vinh Nguyen-Nhu. University Medical Center, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam

E-mail: vinhnguyenmd@ump.edu.vn

INTRODUCTION

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus, which has been recognized as a global pandemic by the World Health Organization (WHO). COVID-19 infection can progress severely and requires medical care, especially in individuals with underlying conditions such as heart disease, diabetes, chronic lung disease, or cancer. Early intervention in the community by providing real-time, evidence-based information for people who are seeking help for their COVID-19 knowledge and information is essential to protect infectious patients and reduce the burden on the healthcare system, which is always overloaded everywhere during the pandemic. Studies proved that understanding the disease and how the virus spreads is the best way to prevent and slow the infection rate [2]. Given the limited human resources and the vast amount of information and knowledge related to COVID-19 that surpasses the consultant's capacity to respond to information seekers in a brief time, a chatbot may serve as an alternative method to address this issue. The fact that chatbots can offer rapid and pertinent information regarding disease and necessary local information may contribute to the reduction of mortality rates and expenses associated with COVID-19 treatment.

In January 2020, following the declaration of the novel coronavirus as a pandemic, the WHO, Centers for Disease Control and Prevention (CDC) and United Nations Children's Fund (UNICEF) utilized chatbots to disseminate information about COVID-19 [1-4]. These chatbots provide up-to-date news and information about the virus, including symptoms of infection and measures individuals may take to protect themselves and others; the information is presented in various formats such as text, video, and graphics. Additional chatbots have also been developed to assist communities in alleviating the impact of COVID-19, such as "COVID-19 Preventable" in Thailand and "COOPERA" in Japan [5,6].

In Vietnam, the Information Technology Department - Ministry of Health launched the "Q&A Chatbot about COVID-19" on their electronic information page in 2020.⁷ This chatbot provides the latest comprehensive information on COVID-19, including government directives, disease prevention and control measures by the Ministry of Health. Updated statistics on the situation, the number of domestic and international patients, the origin and mechanism of transmission of the virus, infection symptoms, and a self-assess infection scoring system for infection are also offered. Additionally, the chatbot includes a list of healthcare facilities nationwide that are qualified to receive and treat COVID-19 patients and synthesizes news on COVID-19 [7].

Nevertheless, the interaction between this chatbot and users is limited to text messages exclusively, without the capability for image or video formats. This limitation may cause confusion for users in certain situations, such as when dealing with medical devices or following breathing exercise instructions. Furthermore, the data available for this chatbot is still limited in terms of its comprehensiveness and diversity. Therefore, this research aims to develop the design process of a chatbot that caters to the public's need for prompt and visually accessible information regarding COVID-19, provides continuous updates from various authoritative sources and is suitable for the social context of Vietnam.

METHODS

The Chatbot model is collaboratively developed by an integrated team of lecturers and students from the Ho Chi Minh City University of Technology, focusing primarily on software design, and from the University of Medicine and Pharmacy at Ho Chi Minh City, being responsible for designing the COVID-19 dataset. The process is conducted through several steps:

Determining the model

To establish a highly efficient and effective interactive conversation between users and computers, many factors are needed. These include comprehending user intents, providing prompt responses, and, most significantly, possessing a system capable of fulfilling these needs. Among the systems for building chatbots, Task-Oriented Dialogue Systems (TODS) have many advantages compared to other systems and can meet the mentioned requirements.⁸ Therefore, the TODS was chosen as the framework to build the COVID-19 Chatbot. TODS employs two methodologies: the pipeline and the end-to-end. Pipeline is an approach that consists of multiple sub-modules with different tasks, and the system will operate by passing data through each module. For the end-to-end, only one module is responsible for processing user information. The pipeline approach is suitable for constructing the COVID-19 Chatbot (Figure 1).

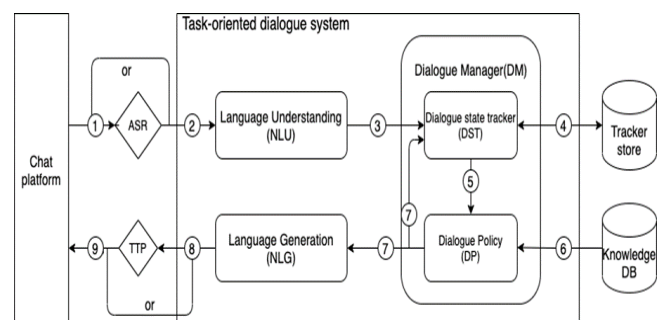


FIGURE 1. Structure of the TODS System.

The structure of a pipeline-based TODS system consists of four main components:

- **Natural Language Understanding Block (NLU):** When the user's text message is inputted, the NLU functions to analyze and present it in pre-defined structured data. In Chatbot software, the structured data format includes Intents and Entities, and each Chatbot will have its specific settings.
- **Dialogue State Tracking Block (DST):** This block is responsible for monitoring and updating the state of the conversation based on the tracked dialogue state.
- **Dialogue Policy Block (DP):** Based on the conversation state from the Dialogue State Tracking block, the DP will determine the next action of the system.
- **Natural Language Generation Block (NLG):** Based on the action determined by the Dialogue Policy block, the Natural Language Generation block is responsible for providing a response to the user in natural language.

Building up content

MongoDB, a NoSQL database system was used as database management system to build up Chatbot's content. The data were structured as question-answer pairs. In the former part, questions were formulated based on aspects and concerns of the public about COVID-19. These sources included social media, articles, television, live broadcasts, textbooks, courses on COVID-19, and guidance documents provided by the Ministry of Health and reputable medical associations.

After constructing the set of questions, corresponding answers were generated based on reliable and reputable sources. This was done to ensure that each answer fulfilled the criteria of being correct, complete, and easy to understand. In this way, the initial set of questions and answers was formed. The upcoming dataset was then edited, reviewed, tested, and finalized to create the most comprehensive dataset. A dataset of over 1000 questions and answers has been established and will be consistently updated and expanded as needed. The data are categorized into 12 modules covering the corresponding issues related to COVID-19, as listed below:

General information and epidemiology: The chatbot provided initial, basic information about COVID-19 and a more detailed understanding of this pandemic. Additionally, information regarding anticipating public concerns, variants of the virus and the mechanisms of transmission was also provided.

State regulations: The state regulations are flexible in response to epidemic situations. Therefore, the data in this part was updated to align with the "new situation" state of the whole country. The data is sourced from official information in the most recent documents from the Ministry of Health, Ho Chi Minh City Department of Health, and the Ho Chi Minh City Center for Disease Control (HCDC), focusing on isolation regulations and other policy regulations.

Prevention measures: One of the contributing factors

in preventing a pandemic is the effective implementation of preventive measures to control the spread of the virus. This is also one of the most concerning and searched-for topics during the pandemic. Hence, constructing a chatbot scenario based on this is appropriate and highly necessary. This module focuses on developing content for this scenario around the Vietnamese "5K model" (Khu trang (facemask)- (Khu khuan) disinfection- (Khoang cach) distance- (Khong tu tap) no gathering- (Khai bao y te) health declaration).

Health declaration and contact information: In relation to this module, Chatbot provides users with the latest addresses and information sourced from the Ministry of Health regarding the regulations governing healthcare for insured patients. Furthermore, users are provided with district-level government/healthcare contact phone numbers and guidance on obtaining oxygen. These vital pieces of information become extremely crucial for everyone during this outbreak, especially in severe situations.

COVID-19 symptoms: It is crucial to monitor symptoms to prevent severe cases, hospitalizations, and fatalities. Citizens should be able to recognize signs of deterioration to use appropriate medication and promptly seek medical assistance. A dataset was created to help users with inquiries regarding mild, moderate, and severe symptoms of COVID-19 and underlying health conditions that pose a risk for severe progression of COVID-19.

COVID-19 testing: The demand for testing is continuously increasing, and testing is not only conducted at healthcare centers but also at homes. This leads to a need for information on how to perform the tests and interpret the results. Chatbot's testing scenario emphasizes the provision of essential information. The current dataset will focus on rapid antigen testing and RT-PCR testing, the two most interesting categories to the community. For each category, questions based on key components were formulated according to characteristics, target groups, locations, and result interpretation.

COVID-19 diagnosis: In the current situation of the pandemic, with a considerable risk of close contact, users will have many concerns about whether they have been infected with COVID-19. Therefore, the issue of diagnosis is crucial. To optimize responses for users, the diagnosis module incorporates two-way communication between the Chatbot and users to gather comprehensive information and provide the most suitable risk assessment and guidance. The interaction can also be seen between those with symptoms and users without symptoms, as the diagnosis module will be developed in two directions for the Chatbot to provide them with answers. The diagnosis tree was developed to solve this issue based on initial symptoms reported by individuals to suggest the likelihood or risk of infection.

Symptomatic users will be divided into two streams: Those who have been tested and those who have not. The former group will be further categorized according to the type of test, i.e., rapid antigen, or RT-PCR. With a negative test result, users will be further asked about epidemiological factors and previous contacts. The user's risk of COVID-19 infection and recommended next steps are determined based on symptoms, test results, epidemiological factors, and previous contacts. For the asymptomatic group, the questioning will focus on the suspicion of contracting the disease through inquiries other than symptoms.

Vaccination: Aiming to achieve nationwide community immunity, the policy focuses on implementing a vaccination strategy that provides free vaccines to the public. The aim of the country is to vaccinate the entire eligible population. Given this situation, there is a high demand for a comprehensive understanding of vaccines, and it is crucial to provide people with accurate information. The vaccine-related topics include vaccine characteristics, target groups, vaccines related to specific groups, side effects and complications, pre- and post-vaccination precautions, and common misconceptions about vaccines.

Treatment: Given the widespread misuse of COVID-19 treatment medications, the dissemination of unverified information in online communities and the disregard for the guidance of medical professionals, the Ministry of Health has promptly issued official documents providing regularly updated instructions on how to treat infectious patients at home. The data on home-based drug treatment included medication packages supplied by the Ministry of Health, indications and contraindications for different patient groups, and common mistakes encountered when using medication.

Health promotion: The pandemic has significantly impacted the health and lives of many people. It is understandable that there is increased concern and attention towards one's health. Therefore, health promotion was chosen as one of the main modules, addressing the public's questions regarding the use of traditional medicine methods, exercises (breathing exercises), and psychological issues.

Medical equipment: Data on equipment is essential, especially when most infectious cases are self-monitored at home. People need to know how to use healthcare equipment at home. Sets of questions and answers were developed focusing on the correct usage and common mistakes when using devices such as SpO₂ monitors, blood pressure monitors, thermometers, and oxygen tanks.

Post-COVID issues: Following the decline of the COVID-19 pandemic in Ho Chi Minh City and the whole country, people's attention has shifted from preventive measures, diagnosis, and COVID-19 testing to focusing on methods to enhance overall health and addressing post-COVID-19 issues. Chatbot has also

adapted to be more aligned with these new circumstances, with a particular emphasis on post-COVID issues – a topic of great current interest. Post-COVID is a new issue, and there are many aspects that people are not yet aware of or do not understand clearly. Thus, for this module, the Q&A dataset primarily comprises basic information about post-COVID health concerns, including definitions, symptoms, methods to improve post-COVID health (lung function, mobility, and endurance), and the timing and location for post-COVID consultations.

Processing and querying data

The user's messages were processed through the following steps: (i) preprocessing the data prior to inputting it into the classification model, (ii) storing the data received from the system's models into the database, and (iii) retrieving the data after receiving the key label for the question from the models and scripts.

Model training

Data were divided into two parts, with the training part comprising 85% of the content and the testing part encompassing the remaining 15%. We utilized the Support Vector Machine (SVM), Named Entity Recognition (ER) and K - nearest neighbor (KNN) to train the model.

System design and implementation

The Chatbot interface is deployed on a web platform and consists of two components. The first is a web-based user interface (UI); the second is a backend that handles the conversation using the Botkit framework. The UI and the backend will communicate with each other through the WebSocket protocol. Upon receiving a message from users, the Chatbot will transmit it to the backend via the Application Programming Interface (API) and follow the execution flow.

Ethical Considerations in the Study

This research adheres to ethical principles in medical research and computer software development. The information in the Chatbot COVID-19 software has been verified and ensured for accuracy by the expert panel. The Chatbot COVID-19 software respects and protects the data of individuals/organizations. It does not infringe on the privacy rights of others, and it does not collect user personal information. This proposal is written by the research team and has not been published in any other work. This proposal has been read, edited, and approved by all members of the team and the supervising instructors. Our study does not collect personal information from survey participants, and participants face no physical or mental health risks. The study was approved by the Ethics Committee of the University of Medicine and

Pharmacy at Ho Chi Minh City.

RESULTS

A database comprising 12 modules with over 1000 sets of questions and answers, addressing most of the issues related to COVID-19 that are of interest to the public. The number of pairs of questions and answers for each module is shown in Figure 2.

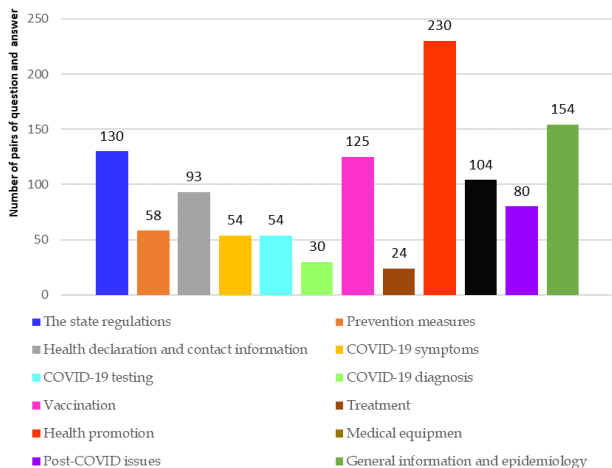


FIGURE 2. Number of pairs of question and answer of each module in the dataset.

The Chatbot dataset has been reviewed and confirmed by the Training Center for Family Physicians of the University of Medicine and Pharmacy at Ho Chi Minh City. It has been determined to be scientifically accurate and appropriate for community education to assist the public during the COVID-19 pandemic.

Applying machine learning and deep learning techniques allows Chatbot to effectively understand the users' intentions and information. The development and refinement of the Chatbot system on the website platform <https://covid19chatbot.com> has been completed. The system has been tested on 245 medical students from the Faculty of Medicine, University of Medicine and Pharmacy at Ho Chi Minh City, yielding promising results (Figure 3-5).

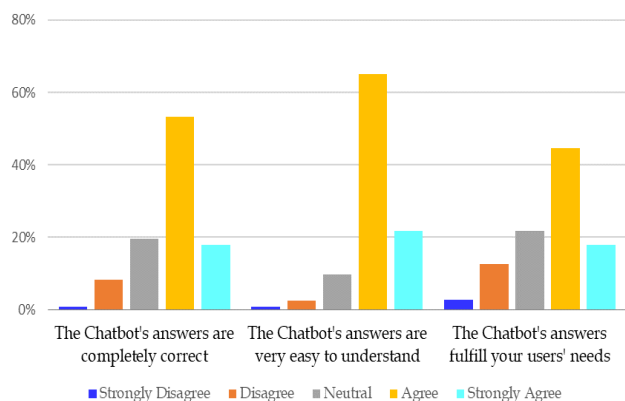


FIGURE 3. Response rates about satisfaction with the content of the Chatbot (n=245).

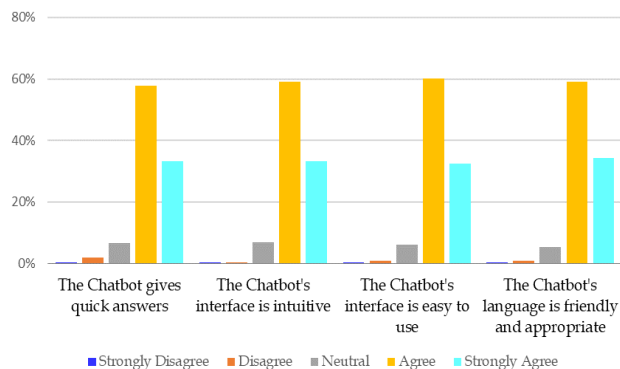


FIGURE 4. Response rates about satisfaction with the interface of the Chatbot

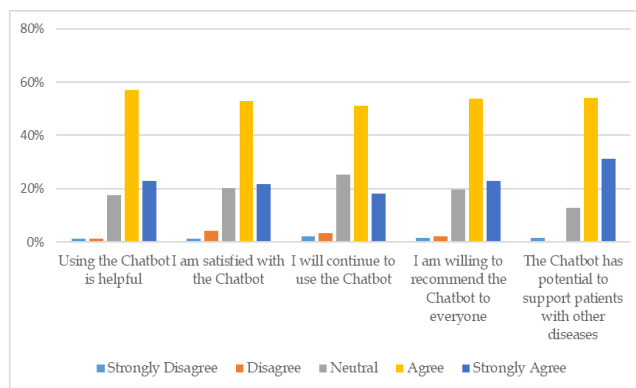


FIGURE 5. Response rates about satisfaction with the Chatbot in general.

DISCUSSION

Since the COVID-19 pandemic, artificial intelligence (AI) technology in healthcare has gained increasing attention and application. The literature on using chatbots in the health sector has been steadily increasing [1]. Recent research emerging since the beginning of the pandemic in 2020 indicates that chatbots play a distinctive role in the aftermath of the COVID-19 pandemic [9]. When crafting content for a chatbot in a critical emergency, emphasis should be placed on maintaining an iterative and agile approach, leveraging complementary forces to triangulate data, and striking a balance between global knowledge and a local focus [1].

Chatbot software development represents a cornerstone and a step forward in healthcare. The software will enable people to access accurate and timely information about the COVID-19 pandemic while alleviating the burden on the healthcare system during that pandemic time. This chatbot operates in the Vietnamese language and engages in bidirectional communication with users via diverse mediums such as text, images, and audio, partly contributing to the successful controlled pandemic in Vietnam.

Several challenges and limitations have been identified through the design and development process,

both in terms of data construction and software development. Regarding data construction, the database constructed remains insufficient in size. As a result, Chatbot may not provide all the information that users need. This lack of compatibility may lead users to believe that our Chatbot does not offer enough practical benefits, causing them to turn to other search engines. Additionally, ensuring the reliability and accuracy of data is still a challenge. While information is used from reputable and official sources like WHO, CDC, and HCDC, the database is quite extensive with various modules. This can lead to minor errors during editing and data entry. Therefore, Chatbot requires a professional team to review the information regularly. Another challenge is the issue of maintaining regular updates of information. Due to the complex and rapidly changing pandemic situation worldwide and in Vietnam (e.g., the emergence of new virus variants, new vaccines, and new policies from the government and the Ministry of Health), verifying and updating the latest information frequently and continuously regarding the pandemic has been challenging. Besides, there is some inconsistency in information about the pandemic between Vietnam and the rest of the world, as well as between different regions within Vietnam. This includes differences in regulations and policies for pandemic prevention, such as isolation and vaccination policies. Finally, the language used in the current platform remains limited. The Chatbot, for instance, still does not provide enough data on the languages used in different regions (local dialects, ethnic languages, internet slang, and so forth).

In terms of software development, there is a limitation in question recognition capability. As the Chatbot is programmed to respond to simple questions that are already present in the compiled dataset, it may not be able to recognize complex, lengthy, multi-faceted questions, or questions not included in the dataset. This is an area that needs improvement and requires a team with greater expertise in software development. The second challenge lies in the inadequate availability of high-level expertise required for software development, upgrades, maintenance, and error correction during operation.

Lastly, the accessibility and promotion of Chatbot still have many limitations. Our users can only access Chatbot through the provided website, and most people are uninformed about it. Thus, it is essential to integrate the COVID-19 Chatbot into popular platforms such as social networks and mobile applications. This will increase its visibility and provide

more convenient conditions for users.

Financial support and additional extensive research and development for the Chatbot are necessary to address these challenges. This includes collaborating with a team of experts in higher-level software development tasks such as building, maintaining, debugging, and continuously improving the software. The Chatbot also needs financial support to facilitate its integration with popular social networks and messaging apps such as Messenger and Facebook, as well as to support promotional campaigns and communication initiatives to make it more accessible to users.

CONCLUSIONS

Given the unpredictable complexity of the pandemic, there is a great need for advanced and innovative approaches to managing the crisis. We have developed a chatbot on the topic of COVID-19 to assist users in accessing authoritative information about health and the pandemic, countering misinformation, evaluating individual risks, tracking exposure to SARS-CoV-2 and health symptoms, and disseminating information on disease prevention measures as well as relevant state regulations.

Although our preliminary survey shows that most users are satisfied with this Chatbot, there are limitations and challenges as the development process is still in its initial stages. Therefore, multidisciplinary collaboration and support in terms of funding and workforce are needed to develop Chatbot further and make it more refined.

ACKNOWLEDGEMENTS

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM and the University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam for supporting this study.

FUNDING

University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam.

CONFLICT OF INTERESTS

Non.

REFERENCES

1. Rambaud K, van Woerden S, Palumbo L, et al. Building a Chatbot in a Pandemic. *J Med Internet Res.* 2023;25:e42960.
2. World Health Organization. HealthBuddy: a new chatbot to engage with communities in Europe and central Asia on COVID-19. 2020.

3. World Health Organization. WHO Health Alert brings COVID-19 facts to billions via WhatsApp. 2020.
4. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. *NPJ Digit Med.* 2020;3:65.
5. Almalki M, Azeez F. Health Chatbots for Fighting COVID-19: a Scoping Review. *Acta Inform Med.* 2020;28(4):241-247.
6. Rodsawang C, Thongkliang P, Intawong T, Sonong A, Thitiwatthana Y, Chottanapund S. Designing a Competent Chatbot to Counter the COVID-19 Pandemic and Empower Risk Communication in an Emergency Response System. *Outbreak, Surveillance, Investigation & Response (OSIR) Journal.* 2020;13(2):71-77.
7. Ministry of Health's Portal (MOH). Home - Cổng thông tin Bộ Y tế. Ministry of Health's Portal (MOH) Home - Cổng thông tin Bộ Y tế (moh.gov.vn).
8. Chen H, Liu X, Yin D, Tang JJA. A Survey on Dialogue Systems: Recent Advances and New Frontiers. 2017;abs/1711.01731.
9. Amiri P, Karahanna E. Chatbot use cases in the Covid-19 public health response. *Journal of the American Medical Informatics Association.* 2022;29(5):1000-1010.