Original Article

JHSMR Journal of Health Science and Medical Research

Usability Assessment of Wowbot: A Chatbot Platform for Healthcare Applications

Supawadee Naorungroj, D.D.S., M.Sc., Ph.D.[1,2], Kittiwara Pupong, D.D.S., MS.[3], Watcharawalee Tangkuptanon, B.Sc., M.Sc., Ph.D.[4], Samerchit Pithpornchaiyakul, DDS, M.Sc.[5,6], Pissamai Wattanasit, RN, Ph.D.[7], Jaranya Hunsrisakhun, DDS, Ph.D.[2,5,6]

[1]Department of Conservative Dentistry, Faculty of Dentistry, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.
[2]Research Center of Excellence for Oral Health, Faculty of Dentistry, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.
[3]Maelan Hospital, Maelan, Pattani 94180, Thailand.
[4]Department of Interdisciplinary Engineering, Faculty of Engineering, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.
[5]Department of Preventive Dentistry, Faculty of Dentistry, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.
[6]Improvement of Oral Health Care Research Unit, Faculty of Dentistry, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.
[7]Department of Pediatric Nursing, Faculty of Nursing, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.

Received 29 September 2025 ● Revised 8 October 2025 ● Accepted 9 October 2025 ● Published online 22 January 2026

Abstract:

Objective: Chatbots have emerged as valuable tools in healthcare; however, existing platforms' features are not designed to support user engagement and data collection for healthcare research. This study aimed to assess the usability issues of Wowbot, a chatbot platform developed as a plugin to the Botnoi platform, through usability testing with chatbot developers.

Material and Methods: Usability testing was conducted with five chatbot developers, who were recruited to perform six tasks using platform functions: dialogue creation, push messages, application programming interfaces (APIs) in creating photo frames and score comparisons, data export, agent-mode chat, and natural language processing (NLP) for parameter recognition. One expert user served as a control. The think-aloud method, semi-structured questionnaires, and exit interviews were used to collect data. Effectiveness was assessed using performance scores, task completion time, and incomplete task incidence, while efficiency was evaluated by comparing task completion time with the expert.

This paper was from the 19th International Dental Collaboration of the Mekong River Region Congress (IDCMR, October 15-17, 2025)
Contact: Jaranya Hunsrisakhun, DDS, Ph.D.
Department of Preventive Dentistry, Faculty of Dentistry, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.
E-mail: hjaranya@hotmail.com

J Health Sci Med Res
doi: 10.31584/jhsmr.20261302
www.jhsmr.org

Results: Four out of five participants completed all tasks. The overall task completion rate within two attempts was 74.5%, with lower rates in photo frame creation (20%), score comparison (40%), and message pushing (40%). Participants required more time to complete tasks than the expert (188 vs. 130 minutes). Nevertheless, users reported high satisfaction (mean 4.2/5) and acknowledged the platform's potential to enhance engagement in healthcare applications.

Conclusion: The Wowbot platform showed potential usefulness for healthcare applications, with new features that may enhance user engagement. However, task completion challenges highlight the need for further refinement, particularly in API design and documentation, to optimize user experience and platform adoption.

Keywords: chatbot platform, oral health education, think-aloud, usability testing

Introduction

Chatbots are conversational software systems designed for a wide range of applications, such as providing information, answering questions, acting as virtual assistants, and performing specific tasks. These systems can operate based on predefined rules or employ advanced natural language processing (NLP) and artificial intelligence (AI) to understand and respond to user inputs conversationally and even manage complex dialogues. In recent years, chatbots have been increasingly utilized in various healthcare sectors to support health education, promote health, engage patients, improve access to care, and collect health data[1,2]. However, not all chatbot platforms are tailored for healthcare applications.

Maintaining user engagement and retention via push notification messages is a key component of health chatbots. The push message or broadcast function enables asynchronous, one-to-many communication from the chatbot to users. This feature allows for the proactive dissemination of information, alerts, or prompts to either the entire user base or targeted segments, independent of user-initiated interactions. It serves to enhance engagement, deliver time-sensitive information, and facilitate mass communication within the chatbot ecosystem. In the healthcare industry, this feature facilitates the implementation of conversational chatbots to promote behavioral modification and provide health education programs[3-5]. However, existing platforms for building Thai chatbots fall short in supporting healthcare and oral health education (OHE), as restricted push functions hinder them for the LINE application, limit data export capabilities, lack features tailored to healthcare research, and have weak Thai NLP.

Earlier Thai chatbots built on Zhang's AI chatbot behavior change model showed promising results in delivering 21- and 30-day OHE programs for caregivers; however, their application was limited by the constraints of the platforms on which they were developed[6]. Each day, the participants received a push message conversation about oral health knowledge and practice for young children. Both OHE chatbots were built on the Chatfuel platform[2], which is a user-friendly platform that requires no coding, with a built-in feature of push notification messages[7]. Nevertheless, Chatfuel is more advanced and optimized for Facebook Messenger and WhatsApp rather than the Line application, which is a more widely used messaging platform in Thailand. Additionally, the data export functionality in chatbot platforms is crucial for comprehensive system analysis and improvement. It enables in-depth analytics, compliance with regulations, debugging, and optimization of conversational flows. This feature supports research,

integration with external tools, and thorough auditing of operations. Ultimately, it drives continuous improvement and maximizes the chatbot's effectiveness in meeting user needs and organizational goals. Moreover, its Thai NLP capabilities are inadequate for push message and data export functions, and the introduction of tiered pricing structures poses additional challenges for a non-profit organization with constrained budgets. In addition, the introduction of new application programming interfaces (APIs) that generate picture frames and enable score comparison for change monitoring could greatly enhance user engagement. These features would allow users to submit their images and easily compare scores over time, potentially serving as valuable gifts or awards.

Currently, the Botnoi platform is widely used to develop Thai AI chatbots. More than 10,000 chatbots in a wide range of businesses, including the health sector, operate on this platform. This is primarily because the Botnoi platform supports the Thai language and has regular upgrades for its content. The platform is built on a drag-and-drop interface and supports API and plugin connections, facilitating more natural and adaptive conversations. Their chatbots can be utilized across various applications, websites, or messaging systems[8,9]. For OHE, such platforms could be particularly useful for general dentists and healthcare providers, as they offer an accessible way to extend education and behavioral guidance to their patients beyond the clinic. However, without functions specifically tailored to health research and education, they may find these platforms insufficient for sustained engagement, data collection, and integration with patient care.

To address the limitations of existing chatbot platforms, the Wowbot platform was developed as an extension of the Botnoi platform, integrating functions specifically tailored for healthcare and OHE while maintaining the core system architecture. These include push messaging, data export, real-time live chat, an analytics dashboard (featuring charts displaying total user count, user retention, popular user input, and demographic information), parameter storage, and APIs for photo frames and score comparison. The development of these features was informed by a user-centered design process involving a focus group discussion with 24 healthcare professionals. Building on this foundation, the objective of the present study was to quantitatively and qualitatively evaluate the usability of the Wowbot platform for healthcare applications. Usability testing was conducted with chatbot developers, who represent the primary user group responsible for constructing health chatbots that can subsequently be used by patients and healthcare providers.

Material and Methods

Participants

The study protocol (EC6408-055) was approved by the Human Research Ethics Committee, Faculty of Dentistry, Prince of Songkla University. When investigating usability issues among homogeneous users, Nielsen has proposed that the first three users will aid in the discovery of problems exponentially. After the fifth user, the data is hypothetically saturated[10]. Thus, five chatbot developers who are current trainees at the Botnoi Group company[9] were enrolled in this study. One chatbot developer who is also a chatbot journey expert was recruited as the control participant. All participants tested the chatbot platform on their computers.

Data collection

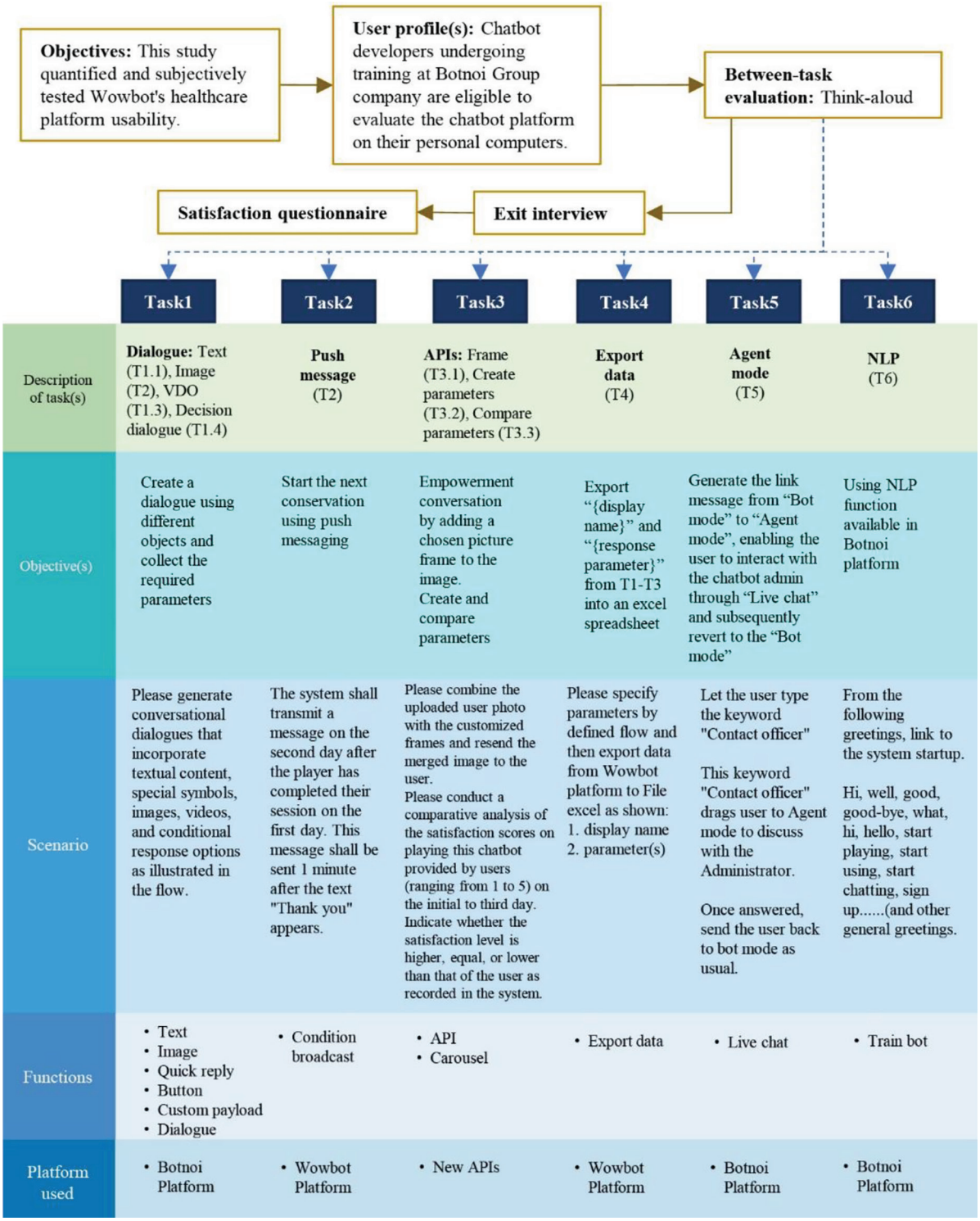
Table 1 shows the metrics gathered during and following tasks to investigate the usability concerns related to the Wowbot platform. The data were obtained from three diverse sources: think-aloud responses, self-administered surveys, and exit interviews to triangulate the results.

The Zoom application, a cloud-based web conferencing platform, was used to record participants' computer screens and faces while performing tasks. The participant was sent the chatbot platform's handbook for self-study two days before testing. On the testing day, participants watched a video explaining the think-aloud strategy. They were asked to verbalize their thoughts and feelings while performing tasks for both negative and positive aspects. A researcher explained to all participants what the rules and six tasks were using a comprehensive dialogue flow diagram (Figure 1). Six representative tasks were selected for the usability evaluation: (1) dialogue creation, (2) sending push messages, (3) use of APIs for creating photo frames and score comparison, (4) data export, (5) agent-mode chat, and (7) NLP for chatbot training. These tasks were chosen because they represent the core functions newly introduced by the Wowbot platform, and they are the essential activities that chatbot developers must be able to perform when building health chatbots for patient education and healthcare research. To ensure testing was conducted in a realistic context, task scenarios were adapted from the 21-Day FunDee chatbot[11], previously developed on the Chatfuel platform, and applied to evaluate performance on the Wowbot platform. It is noteworthy that parameter setting was a prerequisite for score comparisons and data export tasks. Despite no time constraints, participants could terminate tasks at will. Total task duration was recorded, excluding optional breaks. Evaluation assistance was prohibited. However, they were able to assess the chatbot platform's handbook.

After the rules were explained, participants were then individually assigned to Zoom sessions. Two calibrated evaluators observed and documented each participant's interaction with the Wowbot platform in an observational form. A task performance scoring system was used to evaluate users' ability to complete intervention-related tasks, ranging from 0 to 5. A score of 5 was awarded for completing tasks correctly, regardless of whether a manual was used. Scores of 4 to 1 reflected correct task completion within 2 to 4 attempts, respectively, with or without a manual. A score of 0 indicated failure to complete the task. This scoring method allowed for a standardized assessment of task proficiency across varying levels of user support and learning curves.

Table 1 Measures of usability

Usability attributes	Data source		
	Think-aloud and observation	Self-administered questionnaire	Exit interviews
Effectiveness	Performance score Completion rate Incomplete task incidents		
Efficiency	Comparing time completion between the expert and participants		
Satisfaction		Learnability Helpfulness Affect	Perceived task difficulty Guideline helpfulness Perceived value of the platform Intent to use and recommend to others



T=task, APIs=application programming interfaces, NLP=natural language processing

Figure 1 Overall evaluation framework for the usability test of the Wowbot platform

Following the completion of the task performance, participants were given a self-administered questionnaire to evaluate their satisfaction. The questionnaire, adapted from the Software Usability Measurement Inventory (SUMI), employed a 5-point Likert scale (1=Strongly disagree; 5=Strongly agree) in three dimensions: learnability, helpfulness, and affect. To ensure content validity, the modified SUMI questionnaire was reviewed by a panel of five experts in digital technology, dental public health, and health communication. Each item was assessed for relevance, clarity, and representativeness using a 4-point scale. The Content Validity Index (CVI) for individual items ranged from 0.80 to 1.00, indicating acceptable to excellent content validity across the instrument. Although internal consistency reliability was not re-evaluated in the current study, published validation data have demonstrated that the SUMI instrument possesses strong reliability, with Cronbach's alpha coefficients reported for its subscales: learnability=0.84, helpfulness=0.81, and affect=0.88. These established values support the robustness of the SUMI framework from which our modified version was derived[12]. Furthermore, an exit interview was conducted. The participants engaged in a discussion regarding several aspects, including their perception of the difficulty of the tasks, the helpfulness of the guidelines, the value they perceived in the platform, and their intention to use it themselves and suggest it to others.

The information obtained from the think-aloud methodology and the semi-structured interviews was transcribed. Performance scores conducted by two assessors independently were recorded and compared in Microsoft Excel. Any disagreements were resolved through discussions between the two assessors. Performance score and time to completion were analyzed by comparing with the expert users' performance using median values. The success level of each task was determined by calculating the completion rate of tasks with a performance score of 4 or above. The mean and standard deviation of satisfaction scores were calculated for each of the three domains—learnability, helpfulness, and affect. For negatively worded items, rating scores were reversed prior to analysis to ensure consistent interpretation, with higher scores uniformly indicating greater satisfaction.

Results

Participant's characteristics

Among the total of 5 participants, 3 individuals (60.0%) were males. All testers were in their early 20s, while the expert's age was in the late 20s. All participants reported that they had experience with the Botnoi platform and had created 3-10 chatbots. Only one individual was not familiar with other platforms (Table 2).

Table 2 Characteristics of Wowbot platform participants

Characteristics	Expert	Participants				
		#1	#2	#3	#4	#5
Gender	Female	Female	Male	Male	Male	Female
Age (years)	29	22	21	22	21	24
Experience in Botnoi platform	Yes	Yes	Yes	Yes	Yes	Yes
Experience in other platforms	Yes	No	Yes	Yes	Yes	Yes
No. of chatbots ever created	3	5	10	7	10	10

#=participant

Usability of the Wowbot platform

The subsequent section presented various measures encompassing the three fundamental usability aspects: effectiveness, efficiency, and satisfaction.

Effectiveness

Performance scores and completion rate

The performance score encompassed the simultaneous assessment of both success and accuracy. According to Table 3, the expert successfully completed most tasks with a score of 5, except for three activities that involved using APIs (T3.1, T3.2, and T3.3), which can be completed within two attempts (score=4). All testers completed the following tasks: composing text (T1.1) and images (T1.2) with a perfect score of 5 (completion rate=100%). Most participants (80%) completed four tasks comprising creating dialogue using VDO (T1.3), creating parameters (T3.2), using agent mode (T5), and training the bot using NLP (T6) with a score ≥4. Regarding push messages (T2), comparing parameters (T3.3), and creating photo frames (T3.1), about 40.0–60.0% of participants completed the tasks but required three attempts or more.

Incomplete task incidents

Through direct observation and validation by two observers, it was determined that one participant (Participant #5) was unable to accomplish three tasks (T2, push message; T3.1, frames; and T3.3, compare parameters). The action failures were due to the inability of the participant to follow the guidelines.

Efficiency

Comparing time completion between the standard and participants

As indicated in Table 4, the participants took nearly one hour longer (median=188 minutes) than the expert to complete the tasks (130 minutes). However, two participants could finish the activities at approximately the same time as

Table 3 Performance score and completion rate of the participants compared to the expert

Task	Expert	Participants						Completion rate with a score ≥4 (%)
		#1	#2	#3	#4	#5	Median (range)	
T1 Dialogue								
T1.1 Text	5	5	5	5	5	5	5 (5–5)	100
T1.2 Images	5	5	5	5	5	5	5 (5–5)	100
T1.3 VDO	5	5	4	5	5	2	5 (2–5)	80
T1.4 Decision dialogue	5	5	5	5	4	5	5 (4–5)	100
T2 Push message	5	5	3	3	4	0	3 (0–5)	40
T3 APIs								
T3.1 Frame	4	5	2	1	1	0	1 (0–5)	20
T3.2 Parameters	4	5	5	1	4	4	4 (1–5)	80
T3.3 Compare Parameters	4	4	4	1	3	0	3 (0–4)	40
T4 Export data	5	5	5	5	5	4	5 (4–5)	100
T5 Agent mode	5	4	5	2	4	5	4 (2–5)	80
T6 NLP	5	2	4	5	5	5	5 (2–5)	80
Time (min)	130	120	121	191	220	188	188 (120–220)	74.5

#=participant, T=task, APIs=application programming interfaces, NLP=natural language processing, VDO=video

the expert. It is worth noting that most participants spent their time on tasks involving API functions (T3.1 and T3.2). All participants spent less time on tasks that utilized the Botnoi function and new functions such as agent mode and data exporting.

Satisfaction

To assess the cognitive load experienced by the participants during the activity, we requested both the expert and participants (N=6) to evaluate the level of difficulty using the statements provided in self-reported questionnaires on a 5-point scale, as shown in Table 4.

Learnability and perceived task difficulty

Most participants agreed that the platform's functionalities were straightforward to use. After reversing scores for negatively worded items, the mean satisfaction scores for each domain ranged from 3.79 to 4.44. The lowest mean score was observed in the helpfulness domain. Participants' feedback during the exit interviews supported their satisfaction ratings. All participants agreed that tasks involving the use of Botnoi platform features—including creating dialogues by incorporating various items and employing NLP—were effortless and did not require assistance or reference to a manual. Notably, participant #3 indicated that the difficulty level was comparable to that of other chatbots used on different platforms. Another participant, who recorded the shortest completion time across all tasks, remarked that any challenges encountered were attributable to their own abilities rather than to limitations of the platform itself.

…Well, all assignments are pretty easy. The reason I had trouble making parameters was that I can't remember. It was clear to me when I went back and looked at the guide, so I did it. [Participant #1]

Regarding the new features built into the Wowbot platform, all participants agreed that the condition broadcast and export data functions were simple to operate. Most participants reported that the tasks involving the APIs were the most challenging. They mentioned unfamiliarity with the features, complex steps, and insufficient guidance in the handbook as the reason for their difficulty.

Table 4 Learnability, helpfulness, and affect rating among the participants (N= 6)

Items	Mean (S.D.)
Learnability	4.33 (0.83)
"It takes too long to learn the platform function*"	4.67 (0.47)
"Using this framework, tasks can be executed in a straightforward manner"	4.50 (0.76)
"It is easy to make the platform do exactly what you want"	4.33 (0.75)
"It is relatively effortless to move from one part of a task to another"	4.00 (1.15)
"I often need to seek help most times when I use this platform*"	4.17 (0.69)
Helpfulness	3.79 (1.12)
"The instructions and prompts are helpful"	4.33 (0.75)
"I find that the help information given by this platform is not very helpful*"	4.00 (0.82)
"Error messages are not adequate"	3.00 (1.29)
"Across the system, either the amount or quality of the help information varies*"	3.83 (1.07)
Affect	4.44 (0.60)
"I enjoy the time I spend using this platform"	4.17 (0.69)
"Working with this platform is satisfying"	4.67 (0.47)
"I would not like to use this platform on a daily basis*"	4.50 (0.50)

*The rating scores were reversed for negative questions
S.D.=standard deviation

…It was mostly easy, except for making the photo frame; I know what I have to do. I looked at the handbook, but I was confused when I had to copy the API syntax. It is quite complicated. Also, I did not use the API as usual. [Participant #2]

Helpfulness

Although the participants gave positive remarks on the helpfulness of the manuals for this platform, they also stated that the amount and quality of instruction varied (Median score=4). We asked the participants to provide feedback supporting their rating during the exit interview. Most participants mentioned that the handbook should provide more details to clarify the ambiguity of certain functions, such as conditional broadcast and API.

…I need more information about the instructions. For example, on page 59, it wasn't clear to me what "Last click" referred to; I just found out during a test that "Last click" means "Last click" button when creating the dialogue. [Participant #1]

…On page 40, it would be better if you added a note about how to replace the variable for link, image, and frame. [Participant #2]

…To me, manual details are not enough, like a carousel. I want you to show me how to insert objects. A demonstration video clip would be helpful. Moreover, comparing parameter values via API is quite difficult; I want to see all the steps and a screen capture of the result in the video. [Participant #3]

…You should add information about testing or broadcasting that repetition is required. [Participant #4]

Affect

Most participants expressed positive ratings toward using the Wowbot platform, with the mean scores obtained by reversing the scores for negative items, ranging from 4.17 to 4.67. We also asked all participants their opinions about the potential impact of using this platform and applications for their work. All participants agreed that this platform offers advantageous features for healthcare applications and expressed their intention to suggest it to other developers.

..."To me, photo frame and score comparison are the features that can be applied with the chatbot to make the conversion more interesting and engaging. [Expert]

…It's an easy-to-use platform. In terms of the developers, I like the export function and broadcast time because these can be used as a warning application, like taking drugs. [Participant #1]

… I'd like to add a push message feature to the chatbot app for Covid users so that we can ask questions and gather information every day. And with the export function, we can quickly look at the data, such as body temperature. [Participant #2]

Discussion

According to ISO standard 9241-11[13], the components of usability testing should encompass effectiveness, efficiency, and satisfaction. In this study, effectiveness was tested by assessing the success and accuracy of specified tasks, namely the "performance score", while efficiency was evaluated by considering the ability to achieve success within a comparable timeframe to that of expert users. Observations were conducted to assess the methods employed, errors encountered or repeated, and participant feedback, which was expressed through verbal and non-verbal cues such as thinking aloud. In evaluating satisfaction, the self-administered questionnaire modified from SUMI[12] was utilized for responses and supplemented by in-depth interviews. This comprehensive approach provided various perspectives for developing and enhancing the chatbot platform before its widespread implementation, ensuring well-rounded insights and integrated data.

When assessing the usage of newly developed technological products, it is advisable to study the product among experienced individuals, rather than those lacking experience in the field. This is because they can provide valuable perspectives that can contribute to product development without being constrained by the basic ability of testers[14]. In this study, testers were designated as individuals who had previously used the Botnoi platform. This approach facilitates a clearer assessment of abilities, which are directly linked to the usage of both the Botnoi platform and the Wowbot platform, with a focus on the enhanced features of the Wowbot platform.

To assess the feasibility of wider adoption and ascertain the level of difficulty, we integrated the development and evaluation processes to eliminate the need for several study groups. Additionally, it was proposed that a minimum of 3–5 target groups could provide sufficient data for a preliminary evaluation of a Web-based application[15]. This is because a sufficient number of major problems can be identified from system usage, and increasing the number of testers does not significantly impact problem identification. Therefore, a total of 5 participants and 1 expert were utilized in this study, with specific problem statements and target groups identified. Comprehensive evaluations were conducted in line with predefined concepts, including comparisons with the experienced expert who closely resembled the testers with prior experience in using the Wowbot platform.

In this study, the think-aloud technique was employed, which is a method commonly utilized in various research endeavors, including studies related to information systems[16,17]. The fundamental principle of this ideation method is to encourage participants to articulate their thought processes while working with the platform. This aids researchers in gaining deep insights into participants' perception processes, including their needs, thought processes, decision-making methods in platform usage, and the sources of usage-related problems, as well as pathways toward their resolution[17]. In this study, before testing, the participants were instructed to speak freely about their thoughts or feelings during the testing process at all times, regardless of whether it was positive or negative. Researchers encouraged participants to reflect on their feelings if they remained silent or did not express their thoughts for an extended period. In addition, researchers observed participants' behaviors, facial expressions, gestures, and emotional states while testing. This method can generate a wealth of rich and multidimensional data[15].

An expert spent a total of 130 minutes testing all the tasks. Out of 5 testers, 2 completed the tasks faster than the expert by approximately 10 minutes. The three remaining testers spent between 188 and 220 minutes. When comparing evaluation results using median values for each task (T1.1, T1.2, T1.3, T1.4, and T4), both experts and testers achieved equal scores of 5. Task T3.2 also yielded equal scores of 4. However, for other tasks (T2, T3.1, T3.3, T5, and T6), testers scored slightly lower than the experts. This indicated that tasks such as creating dialogues, inserting images, videos, and text – fundamental tasks used in general chatbot creation in the Botnoi platform – are typically straightforward. Additionally, functions like NLP and live chat, which are features of the Botnoi platform, are also easily executable even by some participants who are unfamiliar with them. Similarly, new tasks in the Wowbot platform, such as data export, broadcasts, and score comparison, may need extra effort, but they are not excessively complicated to accomplish. Nevertheless, the T3.1 task, which involves constructing a framework using API commands, received the lowest median score of 1 point. Most testers completed this task in at least four attempts, and one tester was unable to complete it. In general, there was a correlation between performance scores

and behaviors observed in the testers who experienced increased stress. Consulting the manual was common when using new functions, and this finding aligned with responses to the self-administered questionnaire assessment, which generally indicated a very high level of acceptance regarding the ease of use and helpfulness of the instructions.

Based on the exit interviews, ease of use was influenced by prior experience and clear instructions provided in the manual. In this study, testers demonstrated proficiency in using features derived from the Botnoi platform (i.e., live chat dialogues and NLP). Additionally, the participants performed well in a newly developed function that was closely linked to the Botnoi platform, which involved exporting data. Their prior expertise contributed to their success in this task. However, the complex new functions required specific configuration settings and API connections, such as broadcast/push messaging and creating photo frames, resulting in lower performance scores for the testers compared to the expert. During the interviews, the participants provided suggestions for improving the clarity of the manual, highlighting potential pitfalls along with examples. Therefore, future improvements should focus on both the manual and simplifying the usage of these functions, such as incorporating features like "ONE CLICK" or providing step-by-step guidance with examples. Feedback and data exchange from user studies will lead to targeted development[18].

In terms of the advantages of the new functions of the Wowbot platform, it was found that data from the interviews and self-administered questionnaires were consistent, showing a high acceptance. This indicated that the system was interesting, responsive to needs, and recommendable for further use. Additionally, from the interviews, the majority of the research participants agreed that the Wowbot platform was beneficial and had logical functionality, enabling continuous activities, generating interest, increasing engagement, fostering participation, and utilizing data for analysis and performance monitoring in various fields, such as research, education, and business. This is consistent with the findings of numerous studies indicating that push notifications increased application user engagement[3-5,19,20]. Regarding the scoping review for technical aspects in developing chatbots for medication data, the developed chatbots maintained three distinct types of database management: a medical knowledge base comprising a repository of medical facts, a user information database containing demographic and preference information, and a dialogue script database encompassing all the conceivable entries of conversational text responding to users[21]. The focus of our investigation was the final one: to apply statistical tools to analyze user feedback.

Our comparison score for APIs demonstrates potential in healthcare and education by enhancing engagement through personalized metric tracking. It facilitates data-driven insights and adaptive learning, exemplified by monitoring dental plaque levels, knowledge acquisition, and satisfaction longitudinally. This technology augments digital health interventions by integrating personalized monitoring, motivation, and analytics in an accessible interface. Additionally, photo frames were noted to boost user motivation and engagement. Notably, improving user satisfaction and engagement is crucial for overall usability and intervention efficacy[22].

The Wowbot platform is a plugin that enables push notifications and data reporting, extending the functionality of existing software applications without modifying their core code or architecture. This study demonstrates its utility across various application fields. The strength of this study was the utilization of multidirectional study approaches that can enhance the likelihood of acquiring reliable data. These findings can be utilized further for platform development purposes. However, this study has several

limitations. First, the usability evaluation was conducted with chatbot developers, who are the primary intended users of the Wowbot platform. Developers were chosen as the initial evaluators because the Wowbot platform is primarily designed for those building chatbots, and their feedback was most relevant for assessing system-level functions. This choice limits generalizability to other groups, such as general dentists, other healthcare providers, and patients whose perspectives and usability experiences were not assessed. However, evidence from subsequent research has provided reassurance in this regard: chatbots developed using the Wowbot platform were tested with caregivers in a clinical trial, where high satisfaction was reported across multiple domains, including overall satisfaction, user experiences, AI perspective, usefulness, conversational quality, and healthcare quality perspective, etc. (mean scores ranging from 3.6 to 4.2 out of 5)[11]. These findings suggest that the platform can produce usable and acceptable chatbots for end-users (caregivers). Second, although the modified SUMI questionnaire underwent expert review to establish content validity, we did not reassess internal consistency reliability within our sample. Nonetheless, the SUMI instrument has been extensively validated in prior studies, demonstrating Cronbach's alpha coefficients exceeding 0.80 across its scales, which supports confidence in its use as the basis for our adapted measure. Lastly, the diverse skill levels of testers may have influenced task performance, independent of the system's usability. To comprehensively evaluate overall usability and applicability, future research should incorporate larger and more diverse groups, including healthcare providers and end-user perspectives.

Conclusion

Chatbot developers found that the Wowbot platform, including its comparison score and photo frame APIs, showed potential usefulness for healthcare applications, particularly in supporting user engagement and motivation. However, task completion challenges and the small participant group suggest that these findings represent early-stage insights. The main recommendation is to improve the guidebook to enhance the clarity and efficiency of platform use.

Acknowledgement

The authors would like to express their sincere gratitude to the Botnoi team for their outstanding collaboration in developing the Wowbot platform, which was instrumental to this project. We also extend our appreciation to the staff of the Rural Oral Health Center (ROHC), Faculty of Dentistry, Prince of Songkla University, and to the participants for their enthusiastic and valuable contributions to this study.

Funding sources

This research was supported by the Health Systems Research Institute (HSRI), Thailand (DEN6405172S).

Conflict of interest

None declared.

References

1. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023;25:e40789.
2. Pithpornchaiyakul S, Naorungroj S, Pupong K, Hunsrisakhun J. Using a chatbot as an alternative approach for in-person toothbrushing training during the covid-19 pandemic: comparative study. J Med Internet Res 2022;24:e39218.
3. Sakib MN, Butt ZA, Morita PP, Oremus M, Fong GT, Hall PA. Considerations for an individual-level population notification system for pandemic response: a review and prototype. J Med Internet Res 2020;22:e19930.
4. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and

e-health: a scoping review. Telemed J E Health 2022;28:768-80.
5. Kornfield R, Mohr DC, Ranney R, Lattie EG, Meyerhoff J, Williams JJ, et al. Involving crowdworkers with lived experience in content-development for push-based digital mental health tools: lessons learned from crowdsourcing mental health messages. Proc ACM Hum Comput Interact 2022;6(Cscw1).
6. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020;22:e22845.
7. One-time and recurring notifications 2024 [homepage on the Internet]. San Francisco: Chatfuel; 2024 [cited 2024 May 21]. Available from: https://docs.chatfuel.com/en/articles/3705639-one-time-and-recurring-notifications
8. Kingchang T, Chatwattana P, Wannapiroom P. Artificial intelligence chatbot platform: ai chatbot platform for educational recommendations in higher education. IJIET 2024;14:34-41.
9. Botnoi Group [homepage on the Internet]. [cited 2024 May 21]. Available from: https://botnoigroup.com/
10. Nielsen J. Why you only need to test with 5 users [homepage on the Internet]. 2000 [cited 2024 May 21]. Available from: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/
11. Hunsrisakhun J, Naorungroj S, Tangkuptanon W, Wattanasit P, Pupong K, Pithpornchaiyakul S. Impact of oral health chatbot with and without toothbrushing training on childhood caries. Int Dent J 2025;75:1348-59.
12. Kirakowski J. SUMI [homepage on the Internet]. [cited 2025 Aug 20]. Available from: https://sumi.uxp.ie/index.html
13. International Organization for Standardization (ISO). Ergonomics of human-system interaction—Part 11: Usability: definitions and concepts. ISO 9241-11:2018 [homepage on Internet]. Geneva: ISO; 2018 [cited 2023 Nov 11]. Available from: https://www.iso.org/standard/63500.html
14. Turner CW, Lewis JR, Nielsen J. Determining usability test sample size. In: Karwowski W, editor. International Encyclopedia of Ergonomics and Human Factors. 2nd ed. Boca Raton, FL: CRC Press; 2006;p.3076-80.
15. Hwang W, Salvendy G. Integration of usability evaluation studies via a novel meta-analytic approach: what are significant attributes for effective evaluation? Int J Hum Comput Interaction 2009;25:282-306.
16. Chen W, Lin T, Li C, Yuan P. Automated comprehensive evaluation approach for user interface satisfaction based on concurrent think-aloud method. Uni Access Inf Soc 2018;17.
17. Jaspers M, Steen T, Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004;73:781-95.
18. Dwivedi Y, Rana N, Jeyaraj A, Clement M, Williams M. Re-examining the unified theory of acceptance and use of technology (utaut): towards a revised theoretical model. Inf Syst Front 2019;21:1-16.
19. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016;18:e6.
20. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22:e17572.
21. Safi Z, Abd-Alrazaq A, Khalifa M, Househ M. Technical aspects of developing chatbots for medical applications: scoping review. J Med Internet Res 2020;22:e19127.
22. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021;7:e27850.