LONG PAPER

Improving the user experience of a gambling support and education website using a chatbot

S. S. Merkouris[1] · G. Loram[1] · M. Abdelrazek[2] · S. N. Rodda[1,3] · A. Ibrahim[2] · A. Bonti[2] · N. A. Dowling[1,4]

Abstract

Gambling help websites are an important source of information for gamblers and affected others, but they contain large amounts of information, making it difficult to navigate and access required resources. Usability of such websites can be improved through the implementation of chatbots, which are programmed to respond to user requests in natural language. This study aimed to evaluate the impact of a text-based chatbot on the usability, user satisfaction, and experience of the Australian New South Wales GambleAware website; and explore the usability, user satisfaction, and experience of the chatbot itself. A convenience sample of Australian residents were allocated into website with chatbot access ($n = 32$) or website only ($n = 28$). Compared with the website-only group, the website with chatbot access group reported significantly greater ratings of the website's usability and aspects of user satisfaction, but not user experience, and resulted in higher ease of task completion. Compared with available norms, the chatbot was highly rated on usability, usefulness, information quality, and credibility, with the layout, readability of responses, and look and feel identified as areas for improvement. The usability of the NSW GambleAware website could be enhanced by the integration of a text-based chatbot, with potential applications for similar websites.

Keywords Gambling · Chatbot · Usability · Website · User satisfaction · User experience

1 Introduction

Problem gambling is defined as gambling behaviour that leads to adverse consequences for individuals, families, and communities [1]. In Australia, past year problem gambling prevalence estimates range from 0.4 to 0.6%, with 1.9 to 3.7% and 3.0 to 7.7% displaying moderate-risk and low-risk gambling, respectively [2, 3]. Gambling-related harms include financial hardship, disruption, conflict or breakdown of relationships, emotional or psychological distress, decrements to health, cultural harm, diminished performance at work or study, and criminal activity [4]. Gambling-related burden of harm is estimated to be similar to major depressive disorder and alcohol misuse and dependence [5].

Research has shown that face-to-face delivered psychological treatment is the most efficacious intervention for gambling-related problems [6–9]. It has been estimated, however, that only 8 to 16% of individuals with gambling-related problems access face-to-face services [10]. These low rates of help-seeking have been attributed to numerous personal and resource barriers, including shame, stigma, low awareness of treatment options, cost and time commitment, and denial or minimisation of the problem, or not realising that a problem exists [11, 12], as well as a preference to self-manage [13].

Due to these barriers, people seeking help for their own or someone else's gambling-related problems are more likely to pursue self-directed help-seeking options rather than distance-based (e.g. gambling helplines or online counselling) or face-to-face options (e.g. psychological therapies) [14–17]. These low-intensity options tend to be the first

✉ S. S. Merkouris
stephanie.merkouris@deakin.edu.au

1   School of Psychology, Faculty of Health, Deakin University, Geelong, VIC 3220, Australia

2   School of Information Technology, Faculty of Science, Engineering and Built Environment, Deakin University, Geelong, VIC 3220, Australia

3   Psychology and Neuroscience, Faculty of Health and Environmental Sciences, Auckland University of Technology, Auckland, New Zealand

4   Melbourne Graduate School of Education, University of Melbourne, Parkville, Australia

help-seeking options accessed by many gamblers and their affected others before they access professional help, due to their potential anonymity, discretion, and ease of access [15, 17–19]. One of the most commonly employed self-directed help-seeking options is reading information on gambling help websites, with recent studies indicating that 70–80% of online help-seeking gamblers and their affected others engage in this form of self-directed help-seeking [14, 16, 20]. These findings highlight the fact that gambling help websites are an important initial source of information when a problem is first detected. Such websites are therefore a key source for the provision of self-help information and the facilitation of professional support by linking individuals into the service system.

In Australia, gambling resources are provided by state, territory, and federal government departments. The state government of Australia's most populous state, New South Wales (NSW), funds and administers the "NSW Gambling Help" website (https://www.gamblinghelp.nsw.gov.au/). This website provides a number of resources that can assist with: (1) problem identification (e.g. the "Gambling Quiz", which provides an assessment of gambling symptom severity); (2) identifying self-help strategies via the provision of "responsible gambling" tips (e.g. limit-setting); and (3) facilitating further support by providing contact information for various face-to-face and distance-based support services. While the NSW Gambling Help website necessarily provides a comprehensive set of resources to meet the large and diverse range of user needs, barriers such as minimal human presence, static resources, and a large body of information may impede usability.

Website "usability" refers to the ease with which a user is able to navigate an information system or website without formal training [21]. Research has found that greater website usability and reduced website complexity are associated with greater engagement, which in turn impacts on behaviour change outcomes, such as reduction of risky or harmful behaviours (e.g. addictive behaviours; [22]). Conversely, increased complexity can negatively impact engagement. Lenert, Muñoz [23] found that 75% of participants who were given access to a smoking cessation intervention via a website ceased engagement prematurely, with participants indicating that the complexity and navigation difficulties of the website contributed to these low engagement rates.

One way to increase the usability of such websites is with the use of text-based Conversational Agents (i.e. chatbots; [24]), which are software programs designed to interpret and respond to user statements or questions in natural language [25]. Chatbots, which act as an intermediary between the website user and website content, have been used in a number of different contexts, including education, marketing, and customer service [26, 27]. There is evidence that chatbots can increase a sense of social presence and a more

positive and engaging user experience due to their ease of use and provision of easy and pleasant conversation [27, 28]. Chatbots are also advantageous as they are typically faster than other forms of communication (e.g. email), can provide immediate and around-the-clock support, and are able to identify issues and rollout solutions quickly, thereby guaranteeing an immediate response to the various barriers users may face when interacting with websites [25, 27, 29]. Moreover, chatbots are a cost-effective approach, with low ongoing operations and maintenance costs [30, 31].

Despite their cost effectiveness and demonstrated advantages in other fields, few studies have evaluated the use of chatbots in the mental health field. Those that are available, however, display promising results. Specifically, chatbots have been shown to be useful in administering screening tools; for example, Lucas, Gratch [32] demonstrated that a chatbot was effective in eliciting post-traumatic stress disorder (PTSD) symptom information from participants, with the experience of interacting with a non-human (i.e. chatbot) resulting in lower fear of self-disclosure and subsequently higher PTSD symptom disclosure. Additionally, research has shown that psychological interventions (e.g. cognitive-behavioural therapy [CBT], psychoeducation) delivered via chatbots have been effective in increasing psychological well-being [24], as well as reducing stress [24], depressive symptomatology [33–35] and anxiety symptomatology [33, 34]. Importantly, these studies demonstrated relatively high rates of engagement and user satisfaction with the chatbots, indicating that chatbots may be an acceptable form of assessing, treating, and provision of psychoeducation for mental health issues [34–36].

Despite a growing literature evaluating the use of chatbots in the broader mental health field, to date, only one study has examined the use of a chatbot in the gambling field [37]. So, Furukawa [37] used a chatbot to deliver personalised and normative feedback, self-monitoring, and trigger management, with 197 participants randomised to the intervention or a bi-monthly assessment-only control group. So, Furukawa [37] reported a high rate of daily interaction with the chatbot (average of 22/27 days it was offered) and 93% retention rate at 28-day follow-up evaluation. The intervention demonstrated a small but significant reduction in gambling symptom severity but no change in other gambling-related variables (e.g. expenditure). The authors concluded that chatbots may be helpful for those with less severe gambling problems and those who are currently reluctant to access other treatment options.

1.1 Rationale and aims

Taken together, while evidence suggests that text-based chatbots may be an effective mode of delivering screening tools and treatment-related information across the broader

mental health field, there is a paucity of research exploring the utility of chatbots within the gambling field. While the study conducted by So, Furukawa [37] provides valuable information regarding engagement with a chatbot to deliver personalised feedback and CBT messaging, further research into the potential for chatbots to improve the user experience of gambling-related websites is required. Sourcing information on gambling help websites is one of the most commonly employed options to support self-management and is often an entry point for further help-seeking [14–17]. A chatbot implemented into existing gambling help websites has the potential to improve the usability of such websites and enable users to more easily engage with the website content. The current study therefore proposed to evaluate the augmentation of the existing content on the NSW Gambling Help website, which has now been rebranded as the GambleAware website (henceforth, referred to as the NSW GambleAware website; https://www.gambleaware.nsw.gov.au/), using a text-based chatbot, with a view to facilitating website engagement.

Specifically, this study aimed to examine the usability, user satisfaction, and user experiences of the website with and without access to the chatbot. It was hypothesised that, compared to participants who were not provided access to the chatbot (i.e. website only), participants who were provided access to the website and chatbot would report higher website usability, user satisfaction (system usefulness, information quality, and interface quality) and user experience (usability, credibility, loyalty, and appearance), and would find specific tasks easier to complete. A secondary aim was to explore the usability, user satisfaction, and user experience of the chatbot itself.

2 Methods

2.1 Study design

This usability study was a two-arm parallel group quasi-randomised trial designed to evaluate the usability, user experience, and user satisfaction of the NSW GambleAware website with and without access to a text-based chatbot.

2.2 Participants

A convenience sample of 60 participants (65% females), aged from 18 to 85 years ($M = 30.77$, $SD = 13.34$), were recruited from the Australian community. Most participants (80%) used the internet for 22 h or more on a weekly basis for work, personal use, education, or research, and on average had gambled 0.89 ($SD = 2.26$) days in the past month (see Table 1). Individuals were eligible if they were Australian residents; aged 18 years or over; had access to

a computer; and were fluent in English. To participate in this study, individuals were not required to have participated in any gambling activity or have previously used the NSW GambleAware website. Based on a sample size of 26 people per group and 80% power, an effect size of 0.70 would be detectable on the primary outcome measure (System Usability Scale SUS; [38]).

2.3 Chatbot development

The chatbot, named Lilibot, was developed using the IBM Watson platform, which provides advanced natural language processing and machine learning capabilities. To develop the content within Lilibot, the NSW GambleAware website was manually reviewed. All relevant information and resources were extracted and collated via a number of steps. First, the sitemap (all the links and pages these referred to) and all the website pages were reviewed. From this collated information, a list of "intents" (i.e. user goals in terms of what the user might ask about) was developed, along with a list of variations of the same intents (i.e. different ways a user might ask a question). This corpus of intents, questions, and variations was fed into IBM Watson Assistant to train Question-to-Intent Machine Learning Model. This process does not require extensive data because it employs transfer learning (i.e. IBM Watson Assistant is already trained on massive corpus for general Natural Language Processing tasks). IBM Watson Assistant therefore continued learning on the corpus of intents developed for Lilibot to map user questions to one of Lilibot's intents.

For example, a user with the "intent" of accessing "gambling counselling", might ask for help in different ways, such as "Help near me", "I need help" or "I want to talk to a counsellor". Due to the overlap between user queries for similar intents, an intermediate level of intents was developed to help direct the user to the desired information. For example, if a user requested "help", Lilibot would ask whether the help was for them or for someone else (e.g. family member). If the user was requesting help for themselves, then Lilibot asked further questions (e.g. whether they required help in a language other than English) to enable the user to obtain the relevant information as quickly as possible. This process was iterative, with multiple members of the research team testing and contributing to Lilibot on a number of occasions.

The next task in the chatbot development process was the response generation. In this process, appropriate corresponding responses to these user intents (outcome of Question-to-Intent mapping process explained above) were generated. This could take various forms: (1) a direct response to user question (e.g. if the user asked "What is the average expenditure on gambling in Australia?", the bot provides the response extracted from the website); (2) a set of possible options that the user could choose from (e.g. I need help

Table 1 Sample characteristics by group[a]

Demographic and background characteristics	Website with Lili-bot access (n = 32)	Website only (n = 28)	Total (n = 60)
Age (M, SD)	25.15 (14.23)	31.21 (12.50)	30.77 (13.34)
Sex (male) (n, %)	10 (31.25)	11 (39.29)	21 (35.00)
English primary language (n, %)	29 (90.63)	25 (89.29)	54 (90.00)
Born in Australia (n, %)	26 (81.25)	22 (78.57)	48 (80.00)
Employment (n, %)			
Full-time work	8 (25.00)	7 (25.00)	15 (25.00)
Part-time/casual work	14 (43.75)	8 (28.57)	22 (36.67)
Unemployed	1 (3.13)	1 (3.57)	2 (3.33)
Full-time student	8 (25.00)	11 (39.29)	19 (31.67)
Retired	1 (3.13)	1 (3.57)	2 (1.67)
Education (n, %)			
Less than high school	0 (0.0)	1 (3.57)	1 (1.67)
High school graduate	6 (18.75)	4 (14.29)	10 (16.67)
Technical school diploma	1 (3.13)	2 (7.14)	3 (5.00)
Bachelor degree	19 (59.38)	17 (60.71)	36 (60.00)
Masters degree	6 (18.75)	3 (10.71)	9 (15.00)
Doctorate	0 (0.0)	1 (3.57)	1 (1.67)
Weekly internet use (n, %)			
1–7 h	3 (9.38)	1 (3.57)	4 (6.67)
8–14 h	1 (3.13)	2 (7.14)	3 (5.00)
15–21 h	3 (9.38)	2 (7.14)	5 (8.33)
22–28 h	4 (12.50)	6 (21.43)	10 (16.67)
29–35 h	6 (18.75)	6 (21.43)	12 (20.00)
36–42 h	6 (18.75)	4 (14.29)	10 (16.67)
42 + hours	9 (28.13)	7 (25.00)	16 (26.67)
Past-month gambling frequency (M, SD)	0.59 (1.43)	1.21 (2.94)	0.88 (2.26)

[a]There were no significant differences between groups on any of these variables

for me or family member); and (3) an action the user could take (e.g. call hotline, or another follow-up question to narrow down user intent). Examples of the chatbot interactions (questions and responses) are provided in Fig. 1.

2.4 Measures

A questionnaire delivered via the online Qualtrics platform comprised measures evaluating socio-demographic and background characteristics (age, sex, main language spoken at home, country of birth, employment status, highest level of education, weekly internet use, and past-month gambling frequency), as well as the primary (system usability) and secondary outcomes (user satisfaction, user experience and ease of completing specific tasks on the NSW GambleAware website). Participants who were given access to Lilibot were also required to evaluate the system usability, user satisfaction, and user experience specifically associated with the use of Lilibot (using measures that were modified to evaluate Lilibot specifically), and were asked to rank the usability of

various aspects of Lilibot and complete open-ended items about Lilibot. The following measures were selected in order to capture different facets of the overall user experience.

2.4.1 System usability

System usability was assessed using the 10-item System Usability Scale (SUS; [39]). While the SUS consists of eight items that assess usability (ease of use) and two items that assess learnability (ease of learning), evidence suggests that the SUS is a unidimensional measure of subjective usability [39, 40]. Responses are given on a 5-point Likert scale, with response options ranging from 1 (strongly disagree) to 5 (strongly agree). Total scores are calculated by summing items, which are then weighted to range from 0 to 10. Total scores range from 0 to 100, with higher scores indicating greater System Usability. Grading scales developed for the SUS indicate that a score of 68 or more is considered to reflect above average usability [41]. The SUS has demonstrated excellent internal consistency in previous research ($\alpha = 0.91$; [42]).
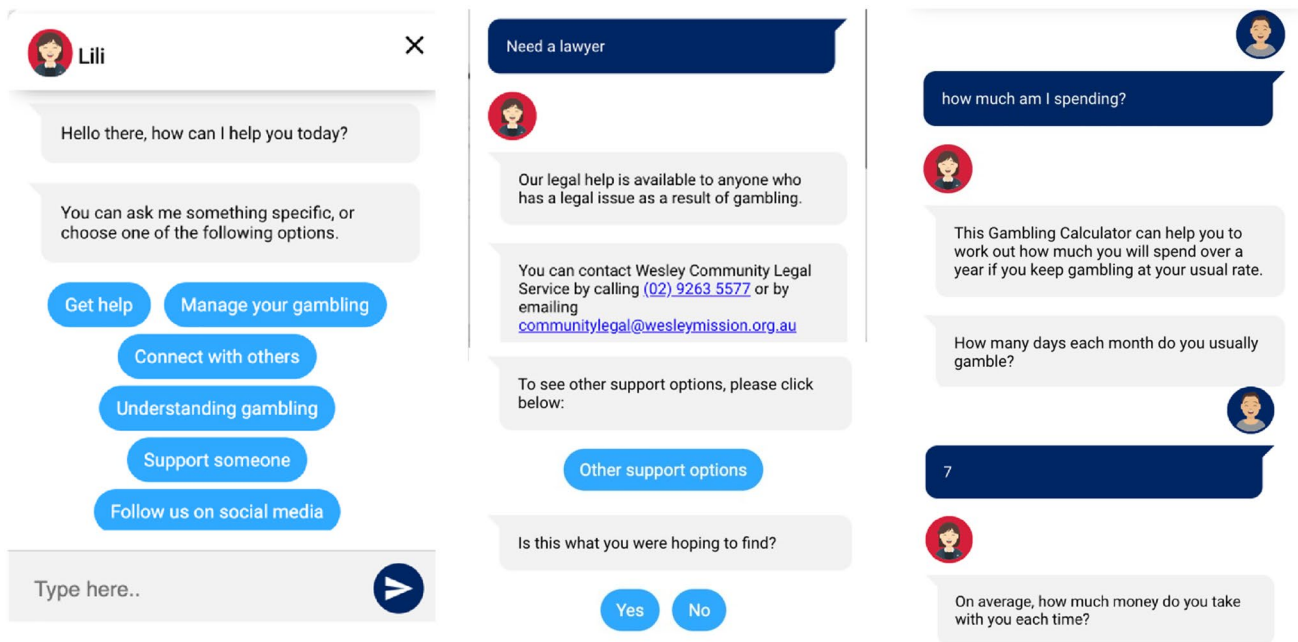
Fig. 1 Examples of Lilibot functionality

2.4.2 User satisfaction

User satisfaction was measured via the 16-item Post-Study System Usability Questionnaire (PSSUQ; [43]). The PSSUQ consists of three subscales: System Usefulness (six items: ease and simplicity of use, and potential effectiveness of the website in increasing productivity due to its ease of use), Information Quality (six items: ease of finding information, ease of dealing with potential errors and potential effectiveness of the information provided in assisting with task completion), and Interface Quality (four items: user interface likeability and satisfaction). Items are scored on a 7-point Likert scale, with response options ranging from 1 (strongly agree) to 7 (strongly disagree). Subscale scores are calculated by averaging the scores on the relevant items, with lower scores indicating better performance and user satisfaction. Norms have been identified to assist in the interpretation of PSSUQ subscale scores: System Usefulness ($M = 2.80$), Information Quality ($M = 3.02$), and Interface Quality ($M = 2.49$; [44]. The PSSUQ subscales have demonstrated high internal consistency in past research ($\alpha = 0.83$–0.96; [44]).

2.4.3 User experience

User experience was assessed using the 8-item Standardised User Percentile Rank Questionnaire (SUPR-Q; [45]). The SUPR-Q consists of four subscales, each consisting of two items: Usability (ease of use and navigation), Credibility (trust and value), Loyalty (would recommend to others and

re-visit in the future), and Appearance (clean, simple, and attractive user interface). Most items are scored on a 5-point Likert scale, with response options ranging from 1 (strongly disagree) to 5 (strongly agree). One item ("Would you recommend this website to a friend or colleague?"), however, is rated on an 11-point Likert scale, with response options ranging from 0 (not at all likely) to 10 (extremely likely). Higher scores on the SUPR-Q are indicative of more positive user experiences. Norms have been identified to assist with the interpretation of the SUPR-Q subscales, with scores above these means indicative of greater user experience relative to other websites: Usability $= 4.06$ (SD $= 0.29$), Credibility $= 3.80$ (SD $= 0.52$), Loyalty $= 3.91$ (SD $= 0.46$), and Appearance $= 3.88$ (SD $= 0.25$; [45]). These subscales have also demonstrated acceptable internal consistency in previous research ($\alpha = 0.64$–0.88; [45]).

2.4.4 Ease of Task Completion

The Single Ease Question (SEQ; [46, 47]) was employed to evaluate the ease of completing five specific tasks, on a 7-point scale ranging from 1 (very difficult) to 7 (very easy). The tasks required participants to find information on the NSW GambleAware website, including: (1) one of the questions asked in the Gambling Quiz; (2) one of the questions asked in the Gambling Calculator; (3) one of the responsible gambling tips; (4) the phone number to access legal help; and (5) how to share a personal experience of gambling on the website. Responses to the SEQ across the five tasks were averaged to derive a total Ease of Task Completion score.

2.4.5 Usability rankings

Participants in the website with Lilibot access group were asked to "rank" the usability of various aspects of Lilibot, from one to three, in relation to what it did well and how it could be improved. The seven usability aspects were: ease of navigation, accuracy of information provided, relevance of information provided, intuitive design, layout of Lilibot, readability, and links to external resources.

2.4.6 Free-text items

Participants in the website with Lilibot access group were asked four open-ended items, including "What did you like about the Chatbot?", "How do you think the chatbot could be improved?", "Currently, the Chatbot has been designed to improve access to content that is already on the NSW GambleAware website. Do you see any other ways that this Chatbot could be used within the NSW GambleAware website? For example, gambling awareness and education, treatment, prevention, etc.?", and "Please provide any other feedback". The first two of these questions required participants to provide a response before continuing.

2.5 Procedure

Ethical approval was obtained from the Deakin University Human Ethics Advisory Group (Ethics ID: SEBE-2020-12). Participants were recruited via convenience and snowball sampling, including social media platforms such as Facebook and LinkedIn and word-of-mouth. Following online consent, participants were automatically allocated into either the website with Lilibot access group or website-only group, based on date of study entry. Allocation was conducted by building a randomiser with evenly presented elements into the Qualtrics survey flow to ensure group sizes were approximately equivalent.

Once allocated, participants commenced the online questionnaire. All participants were asked to complete the five tasks involving finding information contained within the NSW GambleAware website. Participants allocated to the website with Lilibot access group were instructed to use Lilibot to complete the tasks, whereas participants in the website-only group were instructed to use the website. All participants had the task instructions presented to them within the online questionnaire and were provided with a link to the website either with or without Lilibot access. After completing each task, participants were instructed to return to the questionnaire in order to complete the SEQ. After completing all tasks, participants then completed the remainder of the questionnaire. The questionnaire took on average 15.09 (SD = 6.00) minutes for the website with Lilibot access group to complete, and 18.52 (SD = 15.48) minutes for the

website-only group to complete. Data were collected from June 2020 to August 2020. Participants received a \$15AUD Target e-gift card as remuneration for their time and effort.

2.6 Data analysis

Statistical analyses were conducted in STATA v.16 [48]. Due to the use of forced responses, there were no missing data, with the exception of some of the open-ended items. Of the 201 individuals who clicked on the link to the online questionnaire, 150 participants completed the survey. It was determined that 90 of these survey completions were either duplicate responses by individuals or fraudulent bot responses. This conclusion was based on a number of factors: incoherent qualitative responses, large numbers of responses submitted in a short time period, and low scores on reCaptcha measures embedded within the Qualtrics platform (with lower scores indicating bot responses). Of the 51 individuals that commenced but did not complete the questionnaire, 22 were allocated to the website-only group and 29 to the website with chatbot access group. Data analysis was therefore based on a sample size of 60 participants.

Group allocation was regressed onto SUS (system usability), PSSUQ (user satisfaction) subscales, SUPR-Q (user experience) subscales, and the combined Ease of Task Completion (SEQ) scores, in a series of univariate linear regressions. With the exception of SUS System Usability, PSSUQ Information Quality subscale, SUPR-Q Usability subscale, and SUPR-Q Appearance subscale, the data were non-normally distributed. Where data were non-normally distributed, robust estimators were employed. One-sample t-tests were conducted to compare the means of each group (website with and without Lilibot access) on measures with available norms (i.e. PSSUQ User Satisfaction and SUPR-Q User Experience subscales; [44, 45]). Thematic content analysis was used to analyse the data from each free-text item [49]. These analyses were conducted at a semantic level in which the focus is on what each participant said rather than any latent meaning.

3 Results

3.1 Impact of Lilibot on website usability

Table 2 displays the descriptive statistics for the variables of interest, broken down by group. All scales and subscales displayed good internal consistencies ($\alpha = 0.77$–0.93).

The results of the linear regressions (Table 2) indicated that access to Lilibot positively predicted SUS System Usability (SUS; $\beta = 0.28$) and Ease of Task Completion

Table 2 Linear regressions predicting website usability

Variable	Cut-off/ norm	Website only ($n=28$) (M, SD)	Website with Lilibot ($n=32$) (M, SD)	Beta	95% CI of Beta	β	p value	Internal consistency (α)
System usability (SUS)								
Usability	68	62.68 (19.08)	73.59 (19.42)	10.92	0.94, 20.89	0.28	0.033	0.90
User satisfaction (PSSUQ)[a]								
System Usefulness	2.80	2.88 (1.31)	2.08* (1.12)	− 0.80	− 1.43, -0.16	− 0.32	0.015	0.93
Information Quality	3.02	3.02 (1.04)	2.30*** (0.98)	− 0.72	− 1.24, − 0.20	− 0.34	0.008	0.89
Interface Quality	2.49	2.78 (1.32)	2.13 (1.13)	− 0.64	− 1.29, − 0.00	− 0.26	0.049	0.90
User experience (SUPR-Q)								
Usability	4.06	3.64* (0.89)	4.06 (0.93)	0.42	− 0.05, 0.89	0.23	0.081	0.92
Credibility	3.80	4.09 (0.94)	4.11 (0.95)	0.02	− 0.47, 0.51	0.01	0.935	0.90
Loyalty	3.91	3.97 (0.99)	4.29* (0.94)	0.32	− 0.19, 0.82	0.16	0.213	0.77
Appearance	3.88	3.79 (0.91)	3.92 (0.94)	0.14	− 0.34, 0.62	0.07	0.572	0.86
Ease of task completion (SEQ)								
Ease of task completion	–	4.84 (1.28)	5.66 (1.35)	0.81	0.13, 1.49	0.30	0.020	–

Results of one-sample t-tests comparing study means to available norms

PSSUQ Post-Study System Usability Questionnaire, *SEQ* Single Ease Question, *SUPR-Q* Standardised User Percentile Rank Questionnaire, *SUS* System Usability Scale

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

[a]Lower PSSUQ User Satisfaction scores represent better user satisfaction

(SEQ; $\beta = 0.30$), and negatively predicted all PSSUQ User Satisfaction subscales: System Usefulness ($\beta = − 0.32$), Information Quality ($\beta = − 0.34$) and Interface Quality ($\beta = − 0.26$). These results suggest that participants allocated to the website with Lilibot access group were more likely to experience greater Information Quality, System Usefulness, Ease in Completing Tasks, System Usability, and Interface Quality associated with the NSW GambleAware website than the website-only group. In contrast, access to Lilibot did not predict any of the SUPR-Q User Experience subscales.

The results of the one-sample t-tests (Table 2) further suggest that the addition of Lilibot improved certain aspects of the website's system usability, user satisfaction, and user experience to above average levels, when compared to available cut-offs and other website norms [41, 44, 45]. With the exception of SUS System Usability and SUPR-Q Usability, which had slightly lower ratings than available cut-offs and norms, participants in the website-only group rated the NSW GambleAware website as a well-performing website, comparable to other available websites. In contrast, ratings from participants with access to Lilibot were above the average cut-off on SUS (system usability) scores, and scored significantly better than the norms for the PSSUQ System Usefulness subscale, PSSUQ Information Quality subscale, and SUPR-Q Loyalty subscale, indicating that the addition of Lilibot increased the ratings of these aspects of the website to above average levels.

Table 3 Descriptive statistics of website with Lilibot access group

Variable	Cut-off/norm	Website with Lilibot access ($n=32$) (M, SD)	Internal consistency (α)
System usability (SUS)			
Usability	68	71.64 (18.20)	0.87
User satisfaction (PSSUQ)			
System Usefulness	2.80	2.13* (1.15)	0.92
Information Quality	3.02	2.36** (0.99)	0.83
Interface Quality	2.49	2.35 (1.18)	0.89
User experience (SUPR-Q)			
Usability	4.06	4.02 (0.96)	0.83
Credibility	3.80	4.08* (0.86)	0.97
Loyalty	3.91	3.96 (1.13)	0.80
Appearance	3.88	3.59 (1.17)	0.88

Results of one-sample t-tests comparing study means to available norms.

PSSUQ Post-Study System Usability Questionnaire, *SEQ* Single Ease Question, *SUPR-Q* Standardised User Percentile Rank Questionnaire, *SUS* System Usability Scale

Lower PSSUQ User Satisfaction scores represent better user satisfaction

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 4 Ranking of Lilibot features

Lilibot features	Top three Lilibot features ($n=32$) (n, %)	Top three Lilibot improvements ($n=32$) (n, %)
Ease of navigation	19 (59.38)	11 (34.38)
Accuracy of information provided	17 (53.13)	7 (21.88)
Relevance of information provided	19 (59.38)	13 (40.63)
Intuitive design	12 (37.50)	15 (46.88)
Layout of Lilibot	9 (28.13	19 (59.38)
Readability	9 (28.13)	18 (56.25)
Links to external resources	11 (34.38)	13 (40.63)

3.2 Lilibot usability

Table 3 presents the internal consistencies and descriptive statistics on the modified measures (i.e. system usability, user satisfaction, and user experience of Lilibot) completed by participants allocated to the website with Lilibot access group. All scales and subscales displayed good internal consistencies ($\alpha = 0.80–0.97$).

With an average of 71.64 (SD = 18.20), the findings indicate that Lilibot was rated above average on its system usability (SUS; [41]). Moreover, the means of the PSSUQ System Usefulness and Information Quality subscales were significantly lower than the normative means available [44], while the mean of the SUPR-Q Credibility subscale was significantly higher than the normative means available [45], indicating that participants rated these aspects of Lilibot highly compared to other websites. In contrast, the means of the PSSUQ Interface Quality, as well as the SUPR-Q Usability, Loyalty, and Appearance subscales did not significantly differ from the available norms, indicating that they were consistent with the available norms of other websites [44, 45].

3.3 Usability rankings

Participants with Lilibot access ranked the top three features of Lilibot and the top three features on which Lilibot could be improved (Table 4). The top three endorsed features were relevance of information (59.4%), ease of navigation (59.4%), and accuracy of information (53.1%). The top three improvements to Lilibot included layout of Lilibot (59.4%), readability (56.3%), and intuitive design (46.9%).

3.4 Qualitative free-text responses

3.4.1 Lilibot strengths

Participants allocated to the Lilibot access group were asked to report on what they liked about Lilibot. Themes that arose from these responses included: (1) usability of Lilibot; (2)

functionality of Lilibot; (3) efficiency of Lilibot; and (4) aesthetics of Lilibot.

Fourteen participants commented on the usability of Lilibot, indicating that it was easy to use and enhanced website navigation.

"It was really easy to use and provided the information I requested" (Female, age 24).
"Gave answers quickly and saved searching for answers" (Female, age 23).

Eleven participants indicated satisfaction with Lilibot functionality in terms of responding appropriately to their queries.

"It provided the right information or asked the right questions to find what I needed" (Female, age 26).
"It was quite good at understanding what I was looking for" (Male, age 35).

Eight participants liked the efficiency of Lilibot, commenting that it was able to provide answers quickly, which saved them time.

"It quickly provided access to information I needed" (Male, age 32).
"It returned helpful results quickly" (Female, age 28).

Finally, four participants like the aesthetics of Lilibot. Specifically, these participants commented on the attractive design and the simplicity of the interface.

"Simple interface" (Male, age 33).
"It looked attractive" (Male, age 23).

3.4.2 Lilibot improvements

Participants were also asked how Lilibot could be improved, together with another opportunity for general feedback. Themes that arose from these responses related to the: (1) usability of Lilibot; (2) aesthetics of Lilibot; (3) technical difficulties experienced; and (4) interactivity of Lilibot.

The most commonly reported area for improvement related to the usability of Lilibot ($n = 12$). Participants thought that the usability of Lilibot could be improved by providing instructions or examples for how to use Lilibot, as well as by presenting fewer responses at once and removing unnecessary follow-up questions. Moreover, some participants commented on the difficulties in using and navigating Lilibot via a mobile device and thought that the usability of Lilibot was complicated at times by the need to scroll up to read responses.

"Provide instructions—i.e. that you can simply type in questions in a freeform manner" (Male, age 23).
"Too many responses pop up" (Female, age 25).
"These messages don't always all fit within the window so I had to scroll up to see the initial messages (only after I read what I could see and it didn't make sense... I scrolled up and realised I had missed some messages)" (Female, age 26).

Six participants indicated they thought the aesthetics of Lilibot could be improved. Specifically, these participants thought Lilibot could be presented more centrally on the website and that the layout and appearance of Lilibot could be improved, for example, through the use of larger icons.

"The overall layout of the bubbles needs a bit [of] reworking" (Male, age 22).

"The visuals of the chatbot are monotonous. If I'm not looking closely, I can't tell if it's responded to my latest message" (Male, age 35).

Six participants commented on a technical difficulty that they faced while using Lilibot. Specifically, participants commented on the inability to enter a new query during the gambling quiz or gambling calculator. These participants thought that Lilibot could be improved if they were able to stop the quiz or calculator at any time by entering a new query.

"I needed to provide an answer before searching the next question, so you are unable to make a new search without facing a barrier" (Female, age 26)
"I had to finish the whole gambling quiz/reload the page before I could ask another question" (Female, age 20).

Participants also reported that the interactivity of Lilibot could be improved ($n = 5$). These participants noted that Lilibot was not a real person and that it could be improved if it was better able to mimic human conversation.

"It could be more personified" (Female, age 27).
"...unless you made it mimic human conversation more" (Female, age 24).

Finally, two participants commented on the information presented by Lilibot. These comments related to the use of "responsible gambling" rhetoric and that the gambling quiz may not be applicable to all gambling forms (e.g. sports betting and online). These are not, however, direct areas for improvement, as Lilibot was designed to reflect the existing content on the NSW GambleAware website.

3.4.3 Augmentation of current website content

A separate item regarding other ways in which Lilibot could be used was presented to participants in the website with Lilibot access group. This question was optional, with 22 participants responding to this item. The majority of these responses, however, either provided no suggestions or provided ways in which the chatbot could be improved to further augment the current content of the NSW GambleAware website. Of the limited responses ($n = 3$) that directly related to this open-ended item, participants indicated that the chatbot could also be used to: (1) provide gambling education; (2) provide immediate access to counselling strategies or self-help skills rather than just referrals and links; and (3) quickly and efficiently provide information on frequently asked questions, for example, relating to effective treatments.

"Counselling strategies/self-help skills people can use or implement immediately" (Male, age 32)
"It could be a good tool for people to use when they need common information (e.g. education and treatment) but don't want to search through a whole page for the answer. For example, if someone want to know the most common treatment then they can ask that and the Chatbot will give the answer and a link to the page if they want to read further" (Female, age 22)

4 Discussion

The present study was the first to evaluate a text-based chatbot embedded within a gambling help website. Specifically, the study examined the usability, user satisfaction, and user experience of using the NSW GambleAware website with and without access to Lilibot, a text-based chatbot developed to augment existing content on the website. This study also explored the usability, user satisfaction, and user experiences of Lilibot itself.

4.1 Comparison of website usability with and without Lilibot

As hypothesised, access to Lilibot was positively associated with both system usability via the SUS, and ease of task completion via the SEQ, and negatively associated with user

satisfaction via the PSSUQ subscales of System Usefulness, Information Quality, and Interface Quality. Consistent with previous research [24, 27, 28], these findings indicate that participants with access to Lilibot reported greater usability (i.e. ease of use and learning), system usefulness (i.e. ease and simplicity of use and effectiveness in increasing productivity due to its ease of use), information quality (i.e. ease of finding information and dealing with potential errors and effectiveness of the information provided on the website in assisting with task completion), interface quality (i.e. likeability and satisfaction with the user interface) and ease of task completion, compared to participants who were only provided access to the website.

Contrary to expectations, however, Lilibot access was not significantly associated with any of the SUPR-Q user experience subscales of Usability (i.e. ease of use and navigation), Credibility (i.e. trust and value), Loyalty (i.e. would recommend to others and re-visit in the future) or Appearance (i.e. clean, simple and attractive user interface). These findings are inconsistent with previous research, which has shown that text-based chatbots can lead to a more positive and engaging user experience, due to their ease of use and provision of easy and pleasant conversation [27, 28]. These findings are also inconsistent with some of the other findings from the current study, in which usability assessed via the SUS and PSSUQ (i.e. System Usefulness) and Interface Quality assessed by the PSSUQ, which measure similar constructs to the SUPR-Q subscales of Usability and Appearance, respectively, were significantly associated with access to Lilibot. The difference in these usability findings may be attributed to lack of power, with the SUPR-Q usability subscale approaching significance ($p = 0.081$) and displaying a relatively high effect size ($\beta = 0.23$). These differences may also be attributed to the measures employed, with the SUPR-Q employing fewer items per subscale (2 items) than its SUS and PSSUQ counterparts. The higher number of items in the SUS and PSSUQ Interface Quality subscale may allow for a more comprehensive assessment of these constructs. For example, these longer instruments explored the effectiveness in increasing productivity due to ease of use and satisfaction with the user interface, and not just whether the website was easy to use and attractive.

The comparison of study means with available cut-offs and other website norms indicated that, with the exception of usability (via the SUS and SUPR-Q Usability subscale), the NSW GambleAware website alone was rated as comparable to other websites. These findings suggest that, when compared to other websites, the NSW GambleAware website alone is a relatively well-performing website, particularly in aspects of user experience and user satisfaction. The addition of Lilibot, however, improved some aspects of system usability, user experience, and user satisfaction to above average levels. Specifically, the comparison of study means

with available cut-offs and other website norms indicated that participants given access to the website with Lilibot rated the website's system usability above average (via the SUS) and reported significantly higher means on some aspects of user satisfaction (System Usefulness and Information Quality) and user experience (Loyalty). Taken together, these findings suggest that Lilibot improves the usability of the website in the following ways: (1) by making it easier to use, navigate and learn; (2) by providing users with a simple and easy-to-use website that can increase productivity due to its ease of use; (3) by presenting information in a way that is easier to find, assisting with task completion and easily dealing with any encountered errors; and (4) being re-visited and recommended to others.

In contrast, ratings on interface quality, user experience (SUPR-Q) Credibility and user experience (SUPR-Q) Appearance were comparable to other websites for both the website with and without Lilibot access groups. These findings suggest that the website itself, regardless of the addition of Lilibot, provides users with trustworthy and valuable information, as well as a likeable, clean, simple, and attractive user interface.

4.2 Usability of Lilibot

Taken together, the above findings suggest that the usability of the NSW GambleAware website can be improved via the addition of a text-based chatbot. The usability of Lilibot itself, however, was also evaluated in the current study. These findings indicated that when compared to other websites, the usability of Lilibot was above average (SUS). Moreover, Lilibot provided users with a simple and easy-to-use chatbot that could increase productivity due to its ease of use (PSSUQ System Usefulness), presented information in a way that was easy to find, could assist with task completion and could easily deal with any encountered errors (PSSUQ Information Quality) and presented trustworthy and valuable information (SUPR-Q Credibility). These findings are supported by the usability rankings of Lilibot, whereby participants rated the ease of navigation and accuracy and relevance of information as aspects in which Lilibot performed well (i.e. usability and information quality of Lilibot). The findings from the open-ended items tended to further validate these results, with the usability and appropriateness of information provided reported as some of the key strengths of Lilibot. Of note, however, there were some participants that commented that these aspects of Lilibot could still be improved upon, such as presentation of information and usability on mobile devices.

There were additional aspects of Lilibot that could be improved, with results indicating no differences between Lilibot and relevant norms on PSSUQ Interface Quality (i.e. a likeable and satisfactory user interface), SUPR-Q Usability

(i.e. an easy to use and navigate chatbot), SUPR-Q Loyalty (i.e. a chatbot that could be recommended to others or re-visited) or SUPR-Q Appearance (i.e. a clean, simple and attractive user interface). These findings are supported by the usability rankings of Lilibot, whereby participants rated the intuitive design, layout, and readability of Lilibot, which mostly relate to aesthetics and interface quality, as areas for improvement. The findings from the open-ended items tended to further validate these results, with the aesthetics and interactivity of Lilibot reported as some of the key areas for improvement.

4.3 Study implications

The findings from this usability study have important implications. Lilibot was positively evaluated by the participants and helped to increase the usability of the NSW GambleAware website, as well as certain aspects of user satisfaction and user experience. This highlights that a text-based chatbot could be integrated into the NSW GambleAware website as a way of enhancing the website's usability and providing end-users with a more effective and efficient means of obtaining relevant information and resources from the website. Moreover, the findings from this study suggest that Lilibot improved access to information and content within the NSW GambleAware website, suggesting that users may spend more time using the website as the addition of Lilibot made it more engaging and interactive. Prior to integration into the website, however, the usability of Lilibot could be enhanced by improving on the aesthetics and interactivity of Lilibot (e.g. inclusion of more personal elements such as the use of the person's name and an avatar for the chatbot), as well as the way information is presented within Lilibot (e.g. fewer follow-up questions). Finally, the findings of this study also suggest that a chatbot integrated into the NSW GambleAware website could do more than just enhance access to the current content on the website. Future versions of this chatbot could also provide further resources and assistance to end-users, including gambling education, access to counselling strategies or self-help skills, and answers to frequently asked questions, such as the most effective treatment options. Taken together, this low-cost option has the potential to be used to enhance usability across any gambling help website, and more broadly demonstrates that the usability of health service websites can be enhanced by the integration of a text-based chatbot.

4.4 Study limitations

The findings of the current study should be interpreted in the context of several limitations. The use of a convenience sample recruited from the Australian community limits the generalisability of the findings (e.g. the majority of participants were tertiary educated). Furthermore, while participants in this study were regular internet users, their gambling frequency was low, which also has implications for the generalisability of the findings to end-users (i.e. gamblers or affected others seeking help for gambling-related problems). Additionally, while the power calculation indicates that the sample size per group was sufficient to detect differences between the two groups on the primary outcome (i.e. usability), some of the findings suggest that there may have been a lack of statistical power. Taken together, future research with larger sample sizes evaluating the usability of this chatbot on gamblers and affected others seeking help for gambling-related problems is required. As there are no available validated measures developed to assess the specific usability, user satisfaction, or user experience of chatbots [50], the measures employed in the current study were adapted from measures evaluating these constructs for websites. Therefore, while the measures selected were psychometrically sound for the evaluation of the website (and displayed good internal consistency in this evaluation of Lilibot), the psychometric properties of these measures in relation to chatbots are unclear. Similarly, the available norms were based on websites and not chatbots, therefore the comparisons made to explore the usability of Lilibot in comparison with the available norms should be interpreted with caution. Finally, while difficult to compare to other studies given the prevalence of bot-generated data in academic research is currently unknown, over half of the participants who completed the survey were identified as duplicate or fraudulent bot responses [51]. Future research should consider how to protect online surveys from such fraudulent bot responses [51]).

5 Conclusion

Notwithstanding the limitations previously discussed, findings from this study indicate that the usability of the NSW GambleAware website could be enhanced by the integration of a text-based chatbot. Specifically, the addition of a chatbot could enhance the overall usability of the website and key aspects of user satisfaction (i.e. system usefulness, information quality). Lilibot itself, however, could also be improved to further enhance the usability of the website by improving its aesthetics and interactivity. Further research employing samples of end-users (i.e. gamblers and affected others) is required to explore the usability of the website when this chatbot is added.

Data availability Data can be made available upon request to the authors.

Code availability The code can be made available upon request to the authors.

Declarations

Conflicts of interest The authors have no conflicts of interest to declare in relation to this article. The 3-year declaration of interest statement of this research team is as follows: ND, SM, and SR have received funding from multiple sources, including government departments and the Victorian Responsible Gambling Foundation (through hypothecated taxes from gambling revenue). ND, SM, and SR have also received funding from the International Center for Responsible Gaming (ICRG), a charitable organisation, which derives its funding through contributions from the casino gaming industry, equipment manufacturers, vendors, Native American casinos, related organisations, and individuals; funding decisions are the responsibility of a scientific advisory board and neither donors nor the ICRG Governing Board has influence over funding decisions or publishing outputs. SM is the recipient of a New South Wales Office of Responsible Gambling Postdoctoral Fellowship. None of the authors have knowingly received research funding from the gambling, tobacco, or alcohol industries or any industry-sponsored organisation.

Ethics approval Ethical approval was obtained from the Deakin University Human Ethics Advisory Group (Ethics ID: SEBE-2020-12).

Consent to participate All participants provided online informed consent.

Consent for publication N/A.

References

1. Neal, P., Delfabbro, P., O'Neil, M.: Problem gambling and harm: toward a national definition. Office of Gaming and Racing, Melbourne (2005)
2. Dowling, N.A., et al.: National estimates of Australian gambling prevalence: findings from a dual-frame omnibus survey. Addiction **111**(3), 420–435 (2016)
3. Gainsbury, S.M., et al.: The prevalence and determinants of problem gambling in Australia: assessing the impact of interactive gambling and new technologies. Psychol. Addict. Behav. **28**(3), 769 (2014)
4. Langham, E., et al.: Understanding gambling related harm: a proposed definition, conceptual framework, and taxonomy of harms. BMC Public Health **16**, 80 (2016)
5. Browne, M., et al.: A population-level metric for gambling-related harm. Int. Gambl. Stud. **17**(2), 163–175 (2017)
6. Cowlishaw, S., et al.: Psychological therapies for pathological and problem gambling. Cochrane Database Syst. Rev. **11** (2012)
7. Gooding, P., Tarrier, N.: A systematic review and meta-analysis of cognitive-behavioural interventions to reduce problem gambling: Hedging our bets? Behav. Res. Ther. **47**(7), 592–607 (2009)
8. Thomas, S.A., et al.: Australian guideline for treatment of problem gambling: an abridged outline. Med. J. Aust. **195**(11), 664–665 (2011)
9. Yakovenko, I., et al.: The efficacy of motivational interviewing for disordered gambling: systematic review and meta-analysis. Addict. Behav. **43**(11), 72–82 (2015)
10. Productivity Commission: Gambling, Report no. 50. Canberra (2010)
11. Slutske, W.S., Blaszczynski, A., Martin, N.G.: Sex differences in the rates of recovery, treatment-seeking, and natural recovery in pathological gambling: results from an Australian community-based twin survey. Twin Res. Hum. Genet. **12**(5), 425–432 (2009)
12. Suurvali, H., et al.: Barriers to seeking help for gambling problems: a review of the empirical literature. J. Gambl. Stud. **25**(3), 407–424 (2009)
13. Hing, N., Nuske, E., Gainsbury, S.: Gamblers at-risk and their help-seeking behaviour. Gambling Research Australia, Melbourne (2012)
14. Dowling, N.A., et al.: Development and evaluation of an online gambling self-directed program: Effective integration into existing services. Victorian Responsible Gambling Foundation, Melbourne (2018)
15. Hing, N., et al.: All in the family: help-seeking by significant others of problem gamblers. Int. J. Ment. Heal. Addict. **11**(3), 396–408 (2013)
16. Rodda, S.N., Dowling, N.A., Lubman, D.I.: Gamblers seeking online help are active help-seekers: time to support autonomy and competence. Addict. Behav. **87**, 272–275 (2018)
17. Rodda, S.N., et al.: Reasons for using web-based counselling among family and friends impacted by problem gambling. Asian J. Gambl. Issues Public Health **3**(1), 1 (2013)
18. Cooper, G.: Exploring and understanding online assistance for problem gamblers: the pathways disclosure model ecommunity. Int. J. Ment. Health Addict. **1**(2), 32–38 (2004)
19. Rodda, S.N., Lubman, D.I.: Characteristics of gamblers using a national online counselling service for problem gambling. J. Gambl. Stud. **30**(2), 1–13 (2014)
20. Rodda, S.N., et al.: Treatment for family members of people experiencing gambling problems: family members want both gambler-focused and family-focused options. Int. J. Ment. Health Addict. **18**(5), 1–17 (2019)

21. Hirakis, E., Casey, L.M., Clough, B.A.: Investigating website usability: enhancing engagement of amphetamine users in online treatment. Int. J. Ment. Heal. Addict. **16**(2), 420–435 (2018)

22. Strecher, V., et al.: The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J. Med. Internet Res. **10**(5), e36 (2008)

23. Lenert, L., et al.: Design and pilot evaluation of an internet smoking cessation program. J. Am. Med. Inform. Assoc. **10**(1), 16–20 (2003)

24. Ly, K.-H., Ly, A.-M., Andersson, G.: A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv. **10**, 39–46 (2017)

25. McTear, M., Callejas, Z., Griol, D.: The Conversational Interface: Talking to Smart Devices. Springer International Publishing, Cham (2016)

26. Castiglione, A., et al.: In: Cyberspace Safety and Security: 10th International Symposium. Amalfi, Italy. Springer (2018)

27. Di Gaetano, S. and P. Diliberto.: Chatbots and conversational interfaces: three domains of use. In: Fifth International Workshop on Cultures of Participation in the Digital Age, Castiglione della Pescaia, Italy (2018)

28. Brandtzaeg, P.B., Følstad, A.: Why People Use Chatbots. In: Kompatsiaris, I., et al. (eds.) Internet Science. Springer International Publishing (2017)

29. Schuetzler, R.M., G.M. Grimes, and J.S. Giboney: An Investigation of Conversational Agent Relevance, Presence, and Engagement. In: AMCIS 2018 Proceedings (2018)

30. Følstad, A., et al. SIG: Chatbots for Social Good. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (2018)

31. Mittal, A., et al.: A comparative study of chatbots and humans. Situations **2**(2) (2016)

32. Lucas, G.M., et al.: It's only a computer: virtual humans increase willingness to disclose. Comput. Hum. Behav. **37**, 94–100 (2014)

33. Fitzpatrick, K.K., 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. JMIR ment. health **4**(2), e19 (2017)

34. Fulmer, R., et al.: Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR ment. health **5**(4), e64 (2018)

35. Inkster, B., Sarda, S., Subramanian, V.: An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth **6**(11), e12106 (2018)

36. Demirci, H.M.: User experience over time with conversational agents: case study of woebot on supporting subjective well-being. In: Graduate School of Natural and Applied Sciences. Middle East Technical University (2018)

37. So, R., et al.: Unguided chatbot-delivered cognitive behavioural intervention for problem gamblers through messaging app: a randomised controlled trial. J. gambl. stud. **36**(4), 1–17 (2020)

38. Vaziri, D.D., et al., Exploring user experience and technology acceptance for a fall prevention system: results from a randomized clinical trial and a living lab. European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity. **13**(6) (2016)

39. Brooke, J.: SUS: A 'Quick and Dirty' Usability Scale. Usability Evaluation in Industry. CRC Press (1996)

40. Lewis, J.R., Sauro, J.: Revisiting the Factor Structure of the System Usability Scale. J. Usability Stud. **12**(4) (2017)

41. Sauro, J.: A practical guide to the system usability scale: Background, benchmarks & best practices. CreateSpace Independent Publishing Platform (2011)

42. Lewis, J.R. and J. Sauro, (2009) The Factor Structure of the System Usability Scale. In: (Ed) M. Kurosu, Human Centered Design. 94–103. Springer. Berlin

43. Lewis, J.R.: Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In: Proceedings of the Human Factors Society Annual Meeting. Sage Publications Sage CA: Los Angeles, CA (1992)

44. Lewis, J.R.: Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int. J. Human-Comput. Interact. **14**(3–4), 463–488 (2002)

45. Sauro, J., SUPR-Q: A comprehensive measure of the quality of the website user experience. J. usability stud. **10**(2) (2015)

46. Sauro, J.: MeasuringU: 10 Things To Know About The Single Ease Question (SEQ). (2012)

47. Sauro, J., Dumas J.S.: *Comparison of three one-question, post-task usability questionnaires*. In: *Proceedings of the SIGCHI conference on human factors in computing systems* (2009)

48. StataCorp, *Stata Statistical Software: Release 16*. 2019, StataCorp LLC: College Station, TX.

49. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006)

50. Abd-Alrazaq, A., et al.: Technical metrics used to evaluate health care chatbots: scoping review. JMIR **22**(6), e18301 (2020)

51. Storozuk, A., Ashley, M., Delage, V., Maloney, E.A.: Got bots? Practical recommendations to protect online survey data from bot attacks. Quant. Methods Psychol. **16**(5), 472–481 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.