Patient Acceptance of Healthcare Digitalization Through an AI Chatbot: System Usability and Patient Satisfaction

M. M. R. A. P. B. Mapa[1], M. T. D. Lakshan[1], W. M. P. A. Gunasena[2], R. G. L. Dayawansha[2], T. G. S. Nuwan[2], V. Abeysuriya[3], T. N. H. Witharana[1], A. V. K. N. Dinithi[1], L. G. Chandrasena[1]

[1]Nawaloka Hospitals Research and Education Foundation, Colombo
[2]Department of IT, Nawaloka Hospitals PLC., Colombo
[3]Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Sri Lanka.
akilapbm@gmail.com

Abstract. AI chatbots in healthcare represent a significant advancement that enhances service delivery while reducing workload for healthcare professionals. Yet empirical research in low- and middle-income countries remains limited. This study evaluated patient satisfaction and usability of the NH Smart Health Assistant, a specialist referral chatbot developed inhouse and deployed at Nawaloka Hospitals, Sri Lanka. To evaluate patient satisfaction and usability of the chatbot and to explore the relationship between these two constructs, a cross-sectional survey of 400 outpatients assessed demographics, usability (UMUX-Lite), satisfaction, confidence in recommendations, and privacy concerns. Chi-square tests, correlations, and regressions identified predictors of satisfaction, reuse intent, and usability. The study revealed high overall satisfaction (85.5%) and strong intent to reuse (90.25%). Mean usability score was 82.0(±16.2). We identified data privacy(OR=3.14) and confidence in chatbot's recommendations (OR=1.96) as the strongest predictors of satisfaction. Intention to reuse was positively associated with higher education, prior chatbot use, ease of use, and data privacy (OR 4.1, 4.0, 7.4, 2.9). Viewing data privacy as secure boosted UMUX-Lite by ~18 points, while being educated up to A/L added 11 more (all $p < 0.05$). We demonstrated that healthcare chatbots achieve high user satisfaction. Successful integration depends on building user trust rather than demographics or prior experience of chatbots. Multilingualism, voice interaction support, and data transparency are recommended to sustain engagement when scaling AI nationally. Future research should broaden demographic sampling beyond urban populations to inform generalization strategies.

Keywords: Healthcare Chatbot, System Usability, Patient Satisfaction.

1 Introduction

Artificial Intelligence represents a transformative force in contemporary healthcare, fundamentally altering service delivery and patient interaction through AI-driven chatbots powered by Natural Language Processing and Machine Learning algorithms. These conversational agents function as automated digital assistants that engage with users, provide information, assist in preliminary symptom evaluation, and facilitate navigation through healthcare systems, including guidance towards appropriate specialist referrals. Healthcare chatbots primarily serve two main functions: facilitating remote health services including patient support, education, behavior modification, care coordination, and providing administrative support for healthcare providers [1, 2].

Given the rapid rise and integration of artificial intelligence in healthcare, evidence suggests these technologies contribute to improved patient experience metrics by reducing wait times for basic information, providing 24/7 availability, reducing hospital congestion, minimizing non-essential outpatient visits, reducing hospital readmissions, improving patient engagement, and streamlining pathways to specialized care [3]. By automating responses to common queries and handling large amounts of data, chatbots alleviate workload pressures on healthcare personnel, allowing them to dedicate more time to complex clinical tasks [3].

The successful integration of AI chatbots into routine healthcare practice depends critically on user acceptance, which is significantly influenced by perceived usability and patient satisfaction. Research emphasizes the importance of intuitive interaction design, perceived accuracy and reliability of information, user trust in the technology, adaptability for diverse demographics, and robust privacy and security assurances for sensitive health data [3–6]. Prior studies also emphasize that usability and satisfaction are moderately interrelated but context-dependent [4,5], warranting localized investigation. Studies demonstrate that response time, language complexity, conversational capabilities, emotional intelligence, persona presentation, and alignment with user experience principles all contribute significantly to user perception and engagement, with resistance to AI often stemming from perceived lack of empathy [3, 4, 7].

While these dimensions have been investigated in high-income countries, a notable research void exists concerning chatbot usability and patient satisfaction within low- and middle-income countries, including Sri Lanka. Within the Sri Lankan healthcare environment, characterized by human-centered healthcare delivery traditions, AI chatbot adoption remains in early stages with no published studies evaluating the integration, performance, and patient reception of AI-driven patient guidance chatbots within Sri Lankan hospitals. This absence of localized data creates a significant knowledge gap regarding patient perceptions, potential barriers, and overall congruence with existing healthcare seeking behaviors and expectations in Sri Lanka.

The recent introduction of the NH Smart Health Assistant chatbot at Nawaloka Hospitals represents the first large-scale implementation of its kind in a hospital in Sri Lanka, developed specifically to cater to perceived healthcare seeking behaviors of Sri Lankan patients. The technical foundation of this AI-powered chatbot consists of a Python backend using LangChain's RAG modules with a Chroma DB vector database storing medical guidelines and specialist profiles, generating embeddings from patient symptom descriptions to identify relevant clinical patterns aligned with specific medical specialties. Real-time communication is handled through a React frontend connected via WebSocket protocols, creating persistent bidirectional connections that enable instantaneous message exchange and natural multi-turn symptom clarification dialogues. The end-to-end pipeline processes patient inputs through the language model, queries the Chroma DB for relevant clinical guidelines, and accesses e-channeling links filtered by real-time availability through Flask WebSocket APIs managing asynchronous communication.

The NH Smart Health Assistant aims to streamline the specialist referral process, offering time and cost savings for patients, evidence-based guidance over word-of-mouth, and improved resource efficiency for healthcare institutions. However, adoption in Sri Lanka faces challenges such as varying digital literacy, preferences for face-to-face care, the need for multilingual support, and building trust in data privacy and algorithmic reliability [8]. Concerns about potential inaccuracies in symptom interpretation and referral misclassification highlight the need for ongoing validation and oversight [9].

This study addresses a critical gap in evidence around AI chatbot usability and patient satisfaction in Sri Lankan healthcare. The nation’s strong tradition of human-centered care, diverse digital literacy levels, and unique cultural expectations demand localized research rather than reliance on data from high-income settings. Given that user satisfaction with chatbots varies across contexts [6], this implementation offers a timely opportunity to generate locally relevant insights. Beyond academic relevance, the findings inform clinical practice, technology acceptance, and policymaking. Thus, potentially serving as a model for other low- and middle-income countries navigating similar digital health transformations.

The research question guiding this study was: To what extent do usability and satisfaction determine acceptance of a healthcare chatbot among Sri Lankan outpatients? Accordingly, the primary objective of this study was to assess both usability and satisfaction with the chatbot, while also investigating the correlation between these two constructs in the Sri Lankan healthcare context.

2 Methodology

A descriptive cross-sectional study was conducted at Nawaloka Hospitals, Colombo, from February 28 to March 30, 2025, targeting outpatients using the NH Smart Health Assistant.

The sample included 400 patients aged ≥18 who used the chatbot, selected via convenience sampling. Exclusion criteria covered non-users, unwilling participants, acutely ill patients, or those with cognitive or IT literacy limitations. Sample size was calculated using a 50% satisfaction rate compensating for the unknown prevalence, 5% margin of error, and 95% confidence level, yielding 384(rounded to 400) [10]. Convenience sampling was used, recruiting outpatients who interacted with the chatbot during the study period. The questionnaire was structured into five sections: demographics, chatbot usability (UMUX-Lite), satisfaction ratings, confidence in chatbot recommendations, and privacy concerns.

A questionnaire on Google Forms captured demographics, usability (UMUX-Lite), satisfaction, confidence, and privacy perceptions. UMUX-Lite was chosen for its brevity and reliability. Data collection spanned 8 weeks, with pilot testing on 10 patients. The questionnaire was adapted from validated usability and satisfaction instruments (UMUX-Lite, SUS-correlated scales) and refined through pilot testing for clarity and cultural appropriateness in the Sri Lankan context. Satisfaction was measured using a 5-point Likert scale covering courtesy of chatbot, confidence in recommendations, perceived privacy, and overall satisfaction.

Descriptive statistics summarized frequencies and means. Inferential analyses included chi-square tests, Spearman's correlations, and regression models (linear for usability, logistic for satisfaction) to identify predictors. For analysis, variables were dichotomized to simplify the logistic regression and avoid sparse data issues: education (Advanced Level or higher vs Ordinary Level or below), age (≥46 vs 18–45 years), prior chatbot use (ever vs never), and privacy perception ("secure" vs "not secure"). UMUX-Lite usability scores were also dichotomized using a threshold of 70, indicating 'good' usability [11]. Significance was set at $p<0.05$. Data were analyzed using IBM SPSS v27.

Ethical approval was obtained from Nawaloka Hospitals Research and Education Foundation Ethics Review Committee. Informed consent was secured, ensuring anonymity, confidentiality, and voluntary participation was enforced.

3 Results

3.1 Overview

The study sample consisted of a median age of 36.45 years (SD = 12.3), exhibiting a right-skewed age distribution (skewness = 0.46, $p < 0.001$), indicating a higher concentration of younger to middle-aged adults. The largest age subgroup was 36–45 years (32.5%), followed by 26–35 years (29.75%), while seniors aged 66 and above represented a small minority (1.75%). Gender distribution was nearly balanced, with females constituting 50.25% and males 47.5% of the sample; 2.25% preferred not to disclose their gender.

Educational attainment was diverse, with 45.25% having completed Advanced Level education, 31.5% Ordinary Level, 18.75% tertiary education, and 4.5% primary education or below. The education variable showed no significant skewness (skewness = -0.12, p = 0.305), reflecting a relatively even distribution across educational levels.

Overall, users rated the chatbot highly for its courtesy (87 %). Users were frequently confident in the recommendations made by the chatbot (73.5 %), and they mostly found data handling by the chatbot secure(67.6 %). Prior-use of chatbots clustered in the digitally advantaged: 68 % of 18–25-year-olds, 64 % of men and 73 % of tertiary graduates had used chatbots, versus 14 % of adults ≥66 years, 52 % of women and 33 % of respondents with only primary schooling (all χ^2 $p < 0.05$).

3.2 Assessment of Patient Satisfaction with The Chatbot

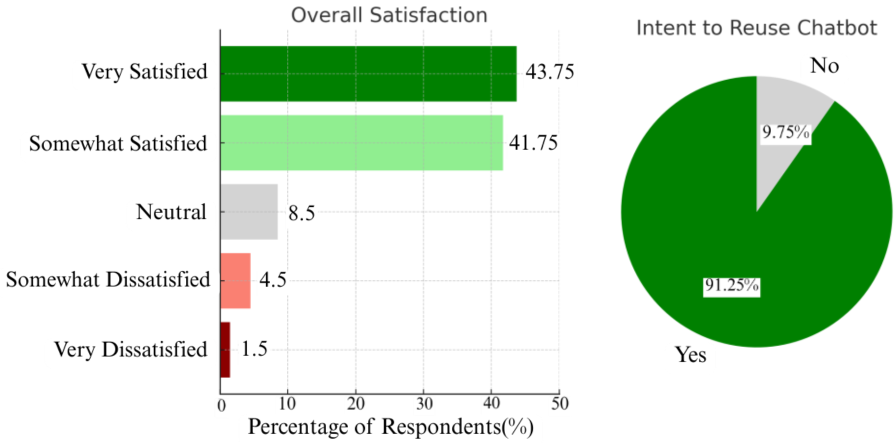

Fig. 1. Overall Satisfaction and the Intent to Re-use the Chatbot

Satisfaction was high: 85 % of respondents were “very” or “somewhat” satisfied (mean 4.22 ± 0.89 on a 5-point scale). Re-use intent was even stronger: 90 % said they would use the chatbot again.

Reuse intent was independently predicted by higher education (OR ≈ 4.1), prior chatbot experience (OR ≈ 4.0), perceived ease-of-use (OR ≈ 7.4), perceived data security (OR ≈ 2.9), and male gender (OR ≈ 2.8)

3.3 Assessment of The Usability of the Chatbot

Perceived usefulness was uniformly high (mean 4.36 ± 0.59; 88.8 % “agree/strongly agree”). Ratings did not vary by age or gender (both $\chi^2 < 0.40$, $p > 0.55$). A modest but significant educational gradient emerged where Ordinary Level(O/L) and Advanced Level(A/L) educated respondents gave the highest scores, whereas tertiary graduates found it slightly less useful ($\chi^2 = 7.26$, $p = 0.027$; $V = 0.13$). First-time users endorsed usefulness marginally more than prior users (97 % vs 92 %; $\chi^2 = 4.28$, $p = 0.038$).

Ease-of-use ratings were uniformly high (mean 4.33 ± 0.75; 92.3 % “agree/strongly agree”) and showed no significant differences by age, gender, education, or prior chatbot experience (all $\chi^2 \leq 1.30$, $p \geq 0.51$).

Mean UMUX-Lite usability was high (82 ± 16) and skewed toward the upper range. Scores differed modestly by subgroup: the highest means occurred in the youngest (18–25 years, 87) and older-senior brackets (56–65 years, 85 ≥ and 66 y, 86), while mid-age adults (46–55 y) scored lowest (78). A/L-educated respondents out-performed those with only primary schooling (83 vs 70). Overall, usability perceptions were favourable across the cohort, with only mild dips in middle-aged and primary educated users.

3.4 Assessment of the Factors Influencing Usability and Satisfaction, and the Correlation between Usability and Satisfaction.

Usability. A multiple linear regression model was developed based on how the results of chi square testing demonstrated the variation of usability. This was targeted to predict UMUX-Lite scores based on demographic factors (age, gender, education, prior chatbot use) and key user experience metrics (confidence in chatbot recommendations, perceived courteousness, and perceived privacy safety).

Reference categories: “primary or below” education and “neutral” for privacy. Model fit: $R = 0.492$, $R^2 = 0.242$, Adj $R^2 = 0.222$; SEE = 14.66; $F(10, 389) = 8.1$, $p < 0.001$. Multicollinearity: VIFs 1.1–6.1 and condition indices < 15 indicate no problematic collinearity.

Table 1. Predictors of Usability (UMUX-Lite score)

Predictor	B ± SE	β	t	p
Up-to GCE O/L	+10.9 ± 3.7	.31	2.92	0.004
Up-to GCE A/L	+11.0 ± 3.6	.33	3.02	0.003
Higher education	+7.3 ± 3.9	.17	1.89	0.060
Privacy "neutral/secure" ($safety_2$)	+9.24 ± 3.0	.16	3.07	0.002
Privacy "mostly secure" ($safety_3$)	+17.6 ± 2.2	.47	8.12	< 0.001
Privacy "very secure" ($safety_4$)	+17.5 ± 2.0	.52	8.76	< 0.001
Courtesy perceived (Yes)	+4.23 ± 2.22	.09	1.91	0.057

Multiple linear regression showed that perceived data security and education were the only independent predictors of usability ($F = 8.1$, $p < 0.001$; adj $R^2 = 0.22$). Compared with users who felt the chatbot was "not secure," those who rated data handling mostly/very secure scored ≈ +18 UMUX points ($p < 0.001$), while a "somewhat secure" rating added ≈ +9 points ($p = 0.002$). Completing O/L or A/L increased scores by ≈ +11 points each ($p < 0.01$); tertiary education (+7 points, $p = 0.06$) and courtesy of the chatbot language (+4 points, $p \approx 0.06$) offered a smaller, marginally significant gain. Gender and confidence in recommendations (not in table) were not significant once privacy and education were controlled.

Satisfaction. Similarly, for overall satisfaction, a multivariable logistic model was fitted with the binary outcome Satisfied ("very" or "somewhat" satisfied = 1; all other responses = 0). Predictors were chosen from the significant or theoretically relevant factors identified in the χ^2 screening: perceived data safety, confidence in recommendations, courtesy, education (A/L or higher vs O/L or below), prior-use (ever vs never), gender (male vs other) and age (≥ 46 years vs 18–45 years). To avoid sparse cells, data privacy was collapsed to secure (mostly/very secure) versus not-secure, yielding sufficient counts in every cell.

Table 2. Predictors of overall satisfaction.

Predictor	Adjusted OR (95 % CI)	Wald χ^2	p
Privacy secure	3.14 (1.85 – 5.33)	22.1	< 0.001
Confidence in recommendations	1.96 (1.11 – 3.46)	5.9	0.015
Courtesy (Yes)	1.52 (0.87 – 2.63)	2.2	0.14
Education ≥ A/L	1.31 (0.75 – 2.29)	0.9	0.33
Prior chatbot use (Ever)	1.27 (0.67 – 2.42)	0.6	0.44
Male gender	1.19 (0.66 – 2.15)	0.4	0.52
Age ≥ 46 y	0.91 (0.48 – 1.73)	0.1	0.80

In the multivariable model (LR $\chi^2 = 38.6$, $p < 0.001$; Nagelkerke $R^2 = 0.19$), only two factors independently predicted satisfaction: perceiving the chatbot's data privacy as secure (aOR = 3.1, 95 % CI 1.9–5.3) and having confidence in its recommendations (aOR = 2.0, 95 % CI 1.1–3.5). Other factors were non-significant (all $p \geq 0.14$).

The Relationship Between the Usability(UMUX-lite) and Perceived Satisfaction was explored using Spearman's rank correlation coefficient. The analysis revealed a moderate positive correlation between UMUX-Lite scores and satisfaction ratings (Spearman's $\rho = 0.446$, $p < 0.001$).

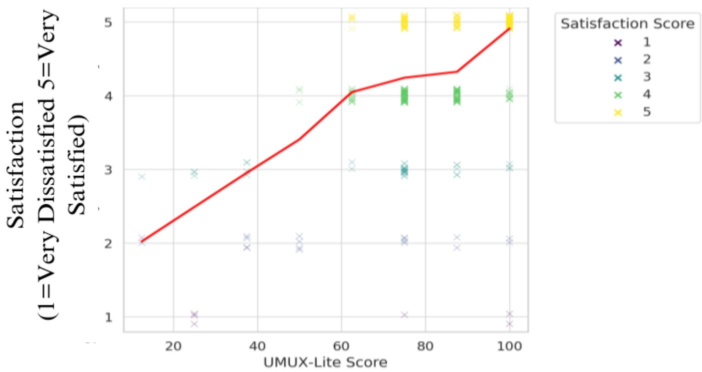

Fig. 2. The Relationship Between the Usability(UMUX-lite) and Perceived Satisfaction

3.5 Potential Areas for Improvement in Chatbot Functionality

Open-ended comments were overwhelmingly favourable. Most respondents used positive terms such as "good," "best," or "excellent," and several explicitly praised the overall "experience". Users asked for Sinhala and Tamil language options(n=17) and some preferred voice interaction over typing(n=11), noting that speech input would help visually-impaired or less tech-savvy patients. Only scattered remarks referred to minor "debug" or service tweaks; no systematic complaints about speed, navigation, or diagnostic accuracy emerged.

4 Discussion

This pioneering cross-sectional evaluation demonstrates that an AI-driven “NH Smart Health Assistant©” can achieve very high acceptance in an urban, middle-income setting. Importantly, this study’s findings directly align with our stated objective: usability and satisfaction were positively correlated, yet distinct constructs. This indicates that enhancing usability alone may not guarantee higher satisfaction unless trust and privacy concerns are addressed simultaneously. Younger, male, and more educated individuals represent early adopters of AI chatbots. These insights highlight the need for targeted strategies to enhance digital inclusion and trust among older, female, and less educated populations to optimize equitable chatbot utilization. More than four-fifths of attendees reported good to excellent usability (mean UMUX-Lite 82/100) with high usefulness (88.8 %) and ease of use (92.3%). Users were satisfied (86 %), and 90% expressed a clear intention to reuse the service. Regression demonstrated perceived data security (+18 UMUX points) and level of education (+11 points) were the most influential factors of high usability. Perceived data security (aOR $\approx$ 3.1) and confidence in chatbot recommendations (aOR $\approx$ 2) increased the odds of overall satisfaction, while an “easy to use” rating remained the strongest independent predictor of reuse intent (aOR $\approx$ 7.4). A moderate positive correlation ($\rho = 0.446$) between UMUX-Lite scores and satisfaction ratings were observed as expected. Users expressed the desire for Multilingualism and voice interaction support in the AI chatbot as potential improvements.

Urban Sri Lankans value fast, low-friction access to specialist advice. The chatbot facilitated this expectation and it was reflected in the exceptionally high reuse intent.

Strengths of the study were, deployment in routine clinical workflow; a sample size (n = 400) adequate for multivariate analyses; and use of UMUX-Lite, which correlates strongly with the System Usability Scale and minimises respondent burden. Although previous studies often report varying correlation between usability and satisfaction, our study found a moderate positive correlation. This may reflect contextual differences in Sri Lanka, where patients value not only ease-of-use but also trust in recommendations and data privacy as separate determinants of satisfaction. Hence, even if usability is high, satisfaction may not peak unless these other factors are addressed.

Limitations temper generalisability. Participants were predominantly young, highly educated, and digitally literate. Rural computer-literacy rates in Sri Lanka remain below 40 %, potentially limiting external validity. Convenience sampling within a single private hospital introduces selection bias, and the two-item UMUX-Lite, while efficient, is not yet fully validated for clinical chatbots. Finally, these outcomes were self-reported; objective measures of clinical impact were beyond scope.

5 Conclusions and Recommendations

This study is the first to evaluate an AI-enabled specialist-referral chatbot in Sri Lanka's healthcare setting. The usability was high (mean UMUX-Lite 82/100), 86 % of users were satisfied, and 90 % intended to reuse the service. Based on our results, future healthcare digitalization through chatbots should adopt data privacy, accuracy, and easy-to use- interfaces as their flagbearers while catering to the educational level of the target population rather than other demographics such as age, gender or prior chatbot use. Moving forward, approaches such as multilingualism, voice interaction support, and data transparency are recommended to sustain engagement when scaling AI nationally.

Our conclusion is consistent with the research objective, demonstrating that usability and satisfaction, while related, are independently influenced by data privacy and trust in chatbot recommendations." Four priorities are recommended as future research interests. First, a controlled comparison between the chatbot guided and conventional referral pathways should be undertaken to quantify effects on waiting time, costs, and downstream clinical outcomes. Second, an accuracy-validation loop is needed: each automated referral ought to be audited against the specialist's final diagnosis, with structured feedback from both patient and clinician, enabling iterative optimisation of the algorithm. Third, external validation across rural and public-sector settings is essential to determine whether the high usability observed in this urban, digitally literate cohort can be reproduced in populations with lower computer-literacy and different care-seeking patterns. Finally, the platform should be expanded to include Sinhala and Tamil interfaces together with optional voice input and output, thereby improving accessibility for older adults, users with limited literacy, and those with visual impairment. These steps will strengthen the evidence base, enhance equity, and guide responsible, scalable integration of conversational AI into Sri Lanka's healthcare system.

References

1. B. Luo, R. Y. K. Lau, C. Li, Y. W. Si, A critical review of state-of-the-art chatbot designs and applications. WIREs Data Min. Knowl. Discov. 12, e1434 (2022). https://doi.org/10.1002/widm.1434.
2. M. Clark, S. Bailey, Chatbots in Health Care: Connecting Patients to Information: Emerging Health Technologies. Canadian Agency for Drugs and Technologies in Health, Ottawa (ON) (2024).
3. J. N. K. Wah, Revolutionizing e-health: the transformative role of AI-powered hybrid chatbots in healthcare solutions. Front. Public Health. 13, 1530799 (2025). https://doi.org/10.3389/fpubh.2025.1530799.

4. J. J. H. Liou, T. T. Vo, Exploring the Relationships among Factors Influencing Healthcare Chatbot Adoption. Sustainability. 16, 5050 (2024). https://doi.org/10.3390/su16125050.
5. R. H. Bokhari, The relationship between system usage and user satisfaction: a meta-analysis. Journal of Enterprise Information Management 18(2), pp. 211–234 (2005). https://doi.org/10.1108/17410390510579927
6. C. C. Ekechi, E. G. Chukwurah, L. D. Oyeniyi, C. D. Okeke, AI-Infused Chatbots For Customer Support: A Cross-Country Evaluation of User Satisfaction in The USA AND THE UK. Int. J. Manag. Entrep. Res. 6, pp. 1259–1272 (2024). https://doi.org/10.51594/ijmer.v6i4.1057.
7. J. Biro, C. Linder, D. Neyens, The Effects of a Health Care Chatbot's Complexity and Persona on User Trust, Perceived Usability, and Effectiveness: Mixed Methods Study. JMIR Hum. Factors. 10, e41017 (2023). https://doi.org/10.2196/41017.
8. Y. Hua, W. Xia, D. W. Bates, G. L. Hartstein, H. T. Kim, M. L. Li, B. W. Nelson, C. Stromeyer, D. King, J. Suh, L. Zhou, J. Torous, Standardizing and Scaffolding Healthcare AI-Chatbot Evaluation, https://www.medrxiv.org/content/10.1101/2024.07.21.24310774v3, (2024). https://doi.org/10.1101/2024.07.21.24310774.
9. F. Reis, C. Lenz, Performance of Artificial Intelligence (AI)-Powered Chatbots in the Assessment of Medical Case Reports: Qualitative Insights From Simulated Scenarios. Cureus. 16, e53899. https://doi.org/10.7759/cureus.53899.
10. P. A. Lachenbruch, S. K. Lwanga, S. Lemeshow, Sample Size Determination in Health Studies: A Practical Manual. J. Am. Stat. Assoc. 86, 1149 (1991). https://doi.org/10.2307/2290547.
11. J. R. Lewis, B. S. Utesch, D. E. Maher, Investigating the Correspondence Between UMUX-LITE and SUS Scores. In: Marcus, A. (ed.) Design, User Experience, and Usability: Design Discourse. pp. 204–211. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-20886-2_20