

A Review of Recent Advancements in Healthcare Chatbots

Vibhav V. Sinai Pissurlenkar,^{1,*} Basabdatta Sen Bhattacharya² and Baskar Sundarrajan³

¹ Infuse Consultancy Ltd., Panaji, Goa, 403001, India

² BITS Pilani, Goa Campus, Vasco, Goa, 403726, India

³ Goa Business School, Goa University, Panaji, Goa, 403206, India

*Corresponding Author

Vibhav V. Sinai Pissurlenkar

✉ vpissurlenkar23@gmail.com

Received: 20 June 2025

Revised: 17 August 2025

Accepted: 01 September 2025

Published Online: 04 September 2025.

Citation

V. V. S. Pissurlenkar, B. S. Bhattacharya, B. Sundarrajan, A review of recent advancements in healthcare chatbots, *Journal of Smart Sensors and Computing*, 2025, 1(2), 25208, <https://doi.org/10.64189/ssc.25208>.

Open Access

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/), which permits the non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as appropriate credit to the original author(s) and the source is given by providing a link to the Creative Commons License and changes need to be indicated if there are any. The images or other third-party material in this article are included in the article's Creative Commons License, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons License and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this License, visit: <https://creativecommons.org/licenses/by-nc/4.0/>

© The Author(s) 2025

Abstract

The quick development of healthcare chatbot technologies, especially since the Coronavirus disease 2019 (COVID-19) pandemic broke out, has been a critical transformation of how healthcare services are presented and consumed. This is significantly fueled by the development of Natural Language Processing (NLP) and the emergence of Large Language Models (LLMs) such as ChatGPT, which have made more advanced, precise, and human-friendly conversational interfaces possible. Between 2023 and mid-2024, the healthcare field experienced an increase in chatbot usage across a range of areas, particularly in mental and overall healthcare. Chatbots have become great assets for simplifying administrative work, delivering preliminary consultations, helping with symptom checkers, delivering mental health services, and supporting education for patients. In mental health, for instance, chatbots can provide a degree of accessibility and anonymity that is highly effective in breaking down stigma, getting mental health care closer to people. In general healthcare, too, these systems have played a key role in easing the workload on healthcare professionals, triaging patients, making health information available 24/7, and even aiding chronic disease management through tailored health advice. Even with their increasing promise, however, there are major hurdles to successfully integrating chatbots into healthcare systems. Some of these include data privacy and security concerns, the risk of wrong diagnosis or recommendation, and the limited capacity of chatbots to address complicated or sensitive patient needs that demand human judgment. Moreover, there are governance and ethical concerns with the application of Artificial Intelligence (AI) in health care, especially regarding achieving fairness, transparency, and accountability. As sophisticated as these systems become, their widespread adaptation and harmonious integration into health care processes is a complicated and continuous process, with user trust and system interoperability shortfalls continuing to present hurdles. Yet, with ongoing improvements in AI and NLP, chatbots have massive potential to revolutionize the delivery of healthcare by offering more personalized, effective, and accessible care.

Keywords: Artificial intelligence in healthcare; Healthcare automation; Chatbots; Conversational agents; Large Language Models.

1. Introduction

In the post-COVID pandemic era, the potential of Artificial Intelligence (AI) in various domains of daily life has become increasingly evident. In particular, the healthcare sector has felt significant strain on its workforce during the pandemic, highlighting the urgent need for AI integration within this narrow domain. Chatbot, also known as conversational agent, have demonstrated considerable promise, particularly due to their features such as ease of access and convenience. These tools allow individuals to engage remotely at their own pace, enabling them to seek assistance for various purposes, including education, therapeutic support, diagnostics, and treatments in both mental and general healthcare. As healthcare systems continue to adapt to the evolving landscape, a chatbot can help bridge gaps in service delivery, enhance patient engagement, and alleviate some of the burdens on healthcare professionals.

Chatbots, including large language models (LLMs), have swiftly integrated into our daily routines. Various organizations around the world, including governments, have recognized their potential and necessity. For instance, some studies, such as those cited in U.S. Government Accountability Office [1], conducted background research during the early stages of the COVID-19 pandemic. More recently, certain governments, such as the UK, have established regulations for the use of AI in this specialized domain.[2] Many individuals have also acknowledged the capabilities of these technologies; for example, a Forbes article by Haseltine[3] tested the accuracy of responses from ChatGPT versions 3.5 and 4 on general medical questions.

Furthermore, an article on Medium by Katendejericho[4] discusses the use of chatbots in mental health, highlighting their advantages, adverse effects, and ethical considerations. The abundance of such materials available online and increase research in the bespoke domain indicates a growing awareness of the potential benefits and challenges associated with AI in healthcare. As these technologies continue to evolve, it is imperative to critically assess their effectiveness, safety, and ethical implications. The abundance of such materials available online and the increase in research in this domain indicate a growing awareness of the potential benefits and challenges associated with AI in healthcare, as illustrated in Fig. 1.

In this review, we have examined various chatbots in the healthcare domain published in the year 2023 to mid-2024. These chatbots utilize AI technology in various forms, including primitive machine learning algorithms such as (decision tree) and more advanced LLMs. Additionally, we will explore the author's perspectives on LLMs within this context based on some of the research articles published during the specified time frame. This article overall reviews the advancements in chatbots and

the role of LLM. With these findings, we aim to provide a comprehensive overview of the latest advancements in AI-powered chatbots in healthcare and their potential to shape future practices.

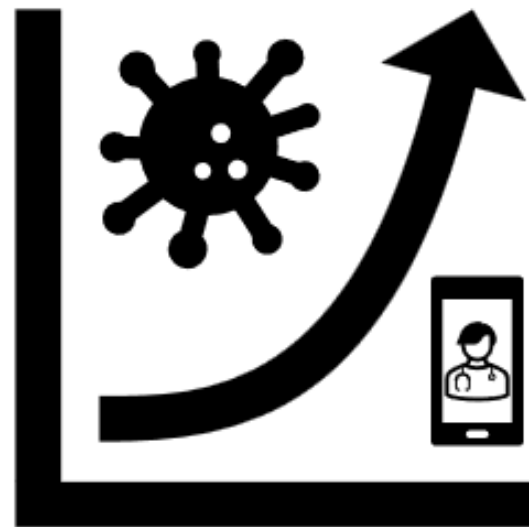


Fig. 1: Rising expectations for tech-driven healthcare solutions post-pandemic, reflected in increased research.

2. Literature review

Chatbot development has evolved significantly since the creation of ELIZA in 1966, one of the first natural language processing (NLP) systems, which employed basic pattern matching techniques to simulate conversation. ELIZA's "DOCTOR" script mimicked a Rogerian psychotherapist, offering a glimpse into the potential of human-computer interaction, as discussed by Weizenbaum.[5] ELIZA, despite its simplicity, laid the foundation for future developments in chatbot technology. Later, PARRY (1972), developed by Colby,[6] advanced the field by simulating the thought patterns of a paranoid schizophrenic. PARRY was an important step forward, demonstrating that chatbots could simulate not only structured conversations but also complex psychological profiles. The 1990s marked a shift toward more sophisticated rule-based systems, such as ALICE (1995), created by Wallace.[7] ALICE utilized an extensive set of predefined rules to engage in meaningful conversation with users, and it earned recognition through its success in the annual Loebner Prize competition. The system relied on pattern matching and heuristic techniques to facilitate conversations, providing an early look into how chatbots could be used for more diverse purposes beyond simple conversation simulation. The development of machine learning and statistical models in the 2000s brought about major breakthroughs. Chatbots like Apple's Siri (2011) harnessed voice recognition and natural language

understanding, making it easier for users to interact with technology through simple voice commands. These early machine learning-based chatbots marked a significant leap in terms of practical application and user engagement. The most recent advancements, such as OpenAI's GPT models (2018 onwards),^[8] have further transformed the landscape by leveraging deep learning and vast datasets. These models have dramatically improved chatbot fluency, context understanding, and the overall quality of interactions, making them more capable of handling diverse and complex tasks.

2.1 Chatbots in healthcare- A growing trend

The integration of chatbots in healthcare has garnered significant attention, largely driven by the increasing prevalence of mental health conditions and a shortage of healthcare professionals, particularly in under-resourced areas. According to Ahmed *et al.*,^[9] the application of chatbots in mental health support has proven to be valuable, particularly in managing conditions like anxiety and depression. Their scoping review, based on the PRISMA methodology, analysed 42 studies and identified several chatbots, including Woebot and Wysa, which leverage Cognitive Behavioral Therapy (CBT) techniques to assist patients in managing their mental health. Despite their growing use, the authors highlight a gap in large-scale usage data and clinical trials, suggesting the need for more comprehensive evaluations and research to better understand their effectiveness in real-world scenarios. Additionally, there is a notable lack of chatbot interventions aimed at general health care. A review by Phiri *et al.*^[10] examined the use of chatbots in healthcare across Africa from 2017 to 2022, identifying 12 relevant studies. The review notes that chatbots in Africa have primarily focused on educational purposes, particularly on social media platforms like Facebook and WhatsApp. These chatbots educate users on essential health topics such as vaccinations, contraception, and HIV testing. Notably, chatbots targeting HIV prevention and testing have been particularly impactful in reaching younger demographics, demonstrating the potential for chatbots to address critical health issues in resource-limited settings. However, the review also highlighted challenges such as technical barriers, low internet penetration, and trust issues among users, which significantly influence the adoption of health-related chatbots in the region. The findings also reveal a gap in research on user experience with health chatbots, particularly in rural areas where access to healthcare services is limited. While studies suggest that digital technologies could increase healthcare access in these regions, they also point to the absence of effective evaluation frameworks. As van Heerden *et al.*^[11] discussed, chatbots like Amanda Selfe and Nolwazi-designed for HIV prevention and self-testing-show promise in terms of improving healthcare

interactions. The use of an isiZulu-speaking chatbot (Nolwazi) for HIV self-testing, for instance, was preferred by 80% of participants over human counsellors, especially among men. These findings highlight how chatbots can facilitate sensitive healthcare discussions, including those on sexually transmitted infections (STIs), pre-exposure prophylaxis (PrEP), and sexual health, which are often stigmatized in many communities. However, despite these positive developments, challenges remain. The integration of chatbots into healthcare systems raises critical concerns about safety, privacy, and ethics. While LLM chatbots offer advanced capabilities, their deployment in healthcare environments requires rigorous validation to ensure accuracy, trustworthiness, and compliance with medical standards. Rule-based chatbots, in contrast, provide greater control over responses and can be fine-tuned to meet specific needs, making them a preferable option for certain applications in healthcare.

2.2 Future directions & challenges

The literature on healthcare chatbots reveals a growing interest in their use for both mental health support and chronic illness management. However, the field still faces several barriers, particularly in the areas of clinical validation and user engagement. A more extensive body of research is required to establish the effectiveness of these technologies in real-world healthcare settings, particularly in underserved and rural areas where chatbots could have the greatest impact. Future research should focus on developing robust evaluation frameworks and conducting large-scale clinical trials to assess the long-term efficacy and safety of healthcare chatbots. Moreover, the integration of AI and machine learning into healthcare chatbots raises important questions about data security, user trust, and the potential for algorithmic biases. As chatbots become increasingly sophisticated, ensuring that they operate ethically and safely will be paramount. Researchers should also explore how to optimize chatbot interfaces to better meet the needs of diverse patient populations, including those with low health literacy or those from marginalized communities. In conclusion, while chatbots in healthcare present a promising solution to many challenges faced by healthcare systems worldwide, significant work remains to ensure their effectiveness and safety. The development of comprehensive evaluation framework, coupled with a focus on user engagement and trust, will be essential in realizing the full potential of chatbots in healthcare.

3. Methodology

For this review, the research methodology used is Systematic Literature Review (SLR), with a focus on inclusion/exclusion criteria and quality assessment.

Table 1: Overview of Chatbots tested, and/or publicly available for use.

Genre	Chatbot	Testing	Available for Public
Mental Health	Fido	Randomized Control Trial (RCT) conducted with 81 participants	No
	Chatbot in Malawi	RCT conducted, 355 participants used the chatbot	No
	Aroha	RCT conducted with 15 participants	No
	Chatbot for psychiatric guidance for Methamphetamine addicted patients	RCT conducted, with 6 months follow-up on 55 participants	No
	Besty	RCT conducted on 45 participants	No
	Vickybot	Clinical scenarios tested with 17 participants	No
	Emohaa	Testing conducted, with 142 participants using the chatbot	No
	Chatpal	RCT was conducted previously, Conducted log analysis recently of 1403 participants	Available through official website (Mobile application)
	Wysa	RCT conducted with 68 participants	Available through official website (Mobile application)
	Woebot	RCT conducted, 68 participants utilized the app	Available through official website (Mobile application)
Physical Health	Moodfit	Unknown	Available through official website (Mobile application)
	Nena	Acceptance testing conducted with 301 participants	No
	Saytù Hemophilie	Usability testing conducted with 57 participants	No
	Lucy Liverbot	Beta Testing conducted on 20 participants	No
	PROSCA	Testing conducted on 10 participants	No
	Haris	Beta testing conducted with 14 participants	No
	Smart Monitoring Tool (SMT, IOT based intervention integrated chatbot)	Testing conducted with 13 participants	No

Given brevity, the exact methodology used that aligns with the core aspects of SLR, encompassing the search strategy, curation, quality assessment, and final selection criteria, is as follows.

For this review, databases such as Google Scholar, OpenAlex, CrossRef and Scopus were searched using keywords: “Chatbot”, “Artificial Intelligence”, “Health” along with filters applied for time i.e. 2023-24. A total of 3003 records matched our keyword search across all the above-mentioned databases. Further curation was done based on citation score while also eliminating duplicates, bringing the total number down to 900. We discarded 420 publications that were not relevant to our objectives for this review, this scrutiny was done based on the title. To ensure a rigorous quality assessment process, we used a custom written python code to retain publications from

well recognized Journals and having good impact factors viz, Nature, Elsevier, Springer, Frontiers, PLOS, IEEE, ACM, JMIR, Taylor & Francis, Sage, Jama network, Wiley, MDPI. Eligibility criteria for curation and selection of research publications were whether, in the bespoke domain, the developed chatbot provide any assistance to users in terms of education, assistive care or counselling and the use of AI technology in the development of the chatbot has been explicitly mentioned in the research article. The publications that did not belong to any of these were discarded. Finally, only 44 articles were retained that fell directly into our topic of interest.

Table 1 gives details on the different chatbots explored in this review, for the benefit of the readers only those chatbots on which some form of testing is conducted and/or available for public usage are included.

4. Chatbots in healthcare

Chatbots and virtual humans are becoming increasingly prevalent in healthcare, especially within mental health interventions. When evaluating these technologies, it is essential to consider these key aspects: a) AI Enabled: Are the interventions AI based? b) Evidence from Testing Trials: Have randomized or clinical trials provided evidence of the effectiveness of these interventions? c) Public Deployment: Are these chatbot deployed for public use? In the following subsections we are going to explore different advancements in mental health and general health.

4.1 Chatbots in mental health

Recent novel advances in mental health include a chatbot developed using Reinforcement Learning with Human Feedback (RLHF) aimed at enhancing mental health therapy.^[12] Although many chatbots are grounded in Cognitive Behavioral Therapy (CBT), which effectively addresses issues such as depression and anxiety, the RLHF approach allows for more adaptive and nuanced interactions. This method enables the chatbot to improve its response accuracy overtime through incremental conversations, contrasting with traditional rule-based chatbots that rely on fixed responses. Although the proposed chatbot demonstrated acceptable performance metrics such as naturalness, coherence, engagement, and understandability, the authors noted a lack of actionable plans for medical implementation. The study also raised concerns regarding the absence of clinical trial data or user demographic testing, indicating a need for further validation before the chatbot can be deemed suitable for medical use in community settings. For the evaluation of this proposed work a novel “Unieval-dialog” technique was used which helps measure the naturalness (response generated is natural in dialogue), coherence (response is coherent in dialogue history) and understandability (is response understandable). The naturalness score obtained post evaluation is 0.94, coherence is 0.96 and understandability is 0.93. Since the authors have claimed that no other research has used this matrix for evaluation, it is difficult to compare it with other work.

Several chatbots have been evaluated through randomized controlled trials (RCTs). With participants recruited within age group of 18-35 years, Karkosz *et al.*^[13] present Fido, a machine learning-based chatbot designed to assist young adults with anxiety and depressive symptoms using cognitive behavioral therapy (CBT) techniques. Developed through iterative co-development with therapists and potential users, Fido underwent rigorous quality assurance and testing, although no features were modified during the trial. The chatbot employs the ABC technique from CBT to help users distinguish between activating events, beliefs, and their emotional or behavioral consequences. In a

randomized controlled trial involving 81 participants, those using Fido reported significant reductions in depression, anxiety, and worry symptoms, alongside increases in life satisfaction and positive effects, with effects lasting for at least a month. Despite its promising results, the study does not clarify whether Fido is available as a standalone mobile app or a web-based service. Additionally, there is no indication of plans for public release, as participants were individually added as testers via an email link, raising questions about its broader accessibility.

With participants in age group 18-29 years, Kleinau^[14] conducted a randomized controlled trial in Malawi to evaluate during the COVID-19 pandemic. The study targeted various professional cadres, including doctors, nurses, and clinical officers, to address mental health challenges such as depression, anxiety, and stress in a region with limited access to mental health resources. Out of the 481 participants in the control group, only 355 received the actual treatment via the chatbot. The trial’s design aimed to bridge the gap in mental health care in Malawi, where access to psychiatrists and other mental health professionals is scarce. A total of 820 participants completed a participant experience questionnaire, with 37% from the control group and 63% from the treatment group. In the treatment group, 50% used the app for over 28 days, with 91% finding it easy to use and 92% deriving benefits. Common issues included confusion with the trial welcome email (32%) and difficulties with app setup (27%). The Net Promoter Score (NPS) was 55. In the control group, 52% used the Internet resources for over 15 days, with 87% finding them easy to use. The NPS for the control group was 51. Both groups encountered similar challenges with time and content complexity. Although the study focused on healthcare workers, it highlights the potential for chatbot interventions to address mental health needs in under-resourced areas, emphasizing the importance of accessible mental health support for those on the front lines.

Tested with participants in mean age group of 20, Kang *et al.*^[15] explored the Aroha chatbot, developed during the COVID-19 pandemic to assist youth (ages 13-24) in managing stress during lockdowns. Aroha employs a rule-based or decision tree algorithm to suggest calming activities and practical advice. In a study involving 15 participants (2 males, 13 females), feedback was gathered to refine the application. However, the authors noted that most participants did not use Aroha in their daily lives outside the controlled environment. This raises questions about its usability and effectiveness on a broader scale for stress management among the general population. Aroha was generally well-received by participants for its conversational tone, use of Kiwi slang, and relatable language, which helped users feel more comfortable. However, participants found the chatbot

overwhelming with too much text at once and expressed frustration with its limited understanding of free-text responses. Despite these limitations, Aroha's interactive features, holistic well-being advice, and accessibility were praised, with participants appreciating its ability to offer support without the stigma or barriers associated with traditional mental health services. The insights gained from this study highlight the potential for improvement and the importance of real-world applicability in developing chatbot for mental health support.

Participants between 18 and 65 years were recruited by Chun-Hung *et al.*^[16] developed a chatbot using machine learning and natural language processing (NLP), deployed on the Line chatbot platform, to assist in the treatment of patients at Jianan Psychiatric Centers. During the preliminary deployment, focus group members and case managers interacted with the chatbot in a realistic setting to simulate the eventual user experience. After analysing feedback, the chatbot underwent refinements to optimize functionality and user interaction. A total of 33 participants were included in the treatment group, with 25 remaining in the control group out of an initial 137. In a study with 50 participants in the (chatbot-assisted treatment) CAT group and 49 in the control group, the CAT group had fewer (Methamphetamine) MA-positive urine samples (19.5% vs. 29.6%). MA-positive samples were positively correlated with frequency of MA use, severity of use disorder, and polysubstance use, and negatively correlated with readiness to change. At the 6-month follow-up, 55 participants completed the study, with 60% reporting relative satisfaction. While the experimental group showed slightly higher treatment retention and significantly fewer MA-positive urine samples than the control group, no significant clinical differences were observed. The study suggests that chatbots can provide immediate support, collect valuable clinical data, and monitor outcomes without significantly burdening patients or providers. Participants generally expressed satisfaction with receiving CAT.

Participants, aged 24 to 68 were recruited by Thunström *et al.*^[17] who conducted a randomized controlled trial comparing usability between an anthropomorphic digital human and a text-based chatbot, BETSY (Behaviour, Emotion, Therapy System, and You), among healthy participants ($n = 45$). Participants were selected based on their scores on the Generalized Anxiety Disorder (GAD-7) scale, were divided into two groups: one interacted with a text-only version of BETSY and the other with a voice-activated digital human. Notably, men were less likely to report annoyance with BETSY compared to women. Overall, the trial found a slight bias toward the text-only interface in terms of acceptability and usability; however, the digital

voice-based interface was still highly rated among participants. This study contributes to understanding user preferences in chatbot, suggesting that while text interfaces may be favoured for usability, voice-based interactions hold significant potential.

In the study by Anmella *et al.*^[18], participants with a mean age of 35–37 years were recruited to evaluate the Vickybot chatbot. Vickybot is designed to assist healthcare professionals and patients experiencing anxiety-depressive symptoms and work-related burnout. This mobile intervention included self-administered scales for monitoring anxiety (GAD-7), depression (PHQ-9), and burnout (using items from the Maslach Burnout Inventory) every two weeks. Psychological modules tailored to assessment severity were delivered, covering anxiety, depression, and work-related stress, based on eclectic therapy, including CBT, mindfulness, and dialectical behavioral therapy. A chatbot guided users through modules, addressed queries, and identified emergencies like suicide thoughts, triggering alerts for immediate assistance. Reminders supported weekly objectives and biweekly assessments, while users could also record audio reflections for potential voice analysis. This comprehensive system ensured personalized, proactive mental health management and emergency response. This research is part of the PRESTO project, which aims to combine machine learning models for severity assessment with a smartphone-based intervention for screening, monitoring, and treatment delivery. The primary objective of the study was to evaluate the feasibility of the intervention, while secondary aims focused on its effectiveness in reducing symptoms and detecting suicide risk. During the setup phase, 40 users tested Vickybot, confirming reliable data transmission and server performance. In the simulation phase, 17 (76% female) users tested clinical scenarios, with 98.5% of expected functions and 98.8% of expected modules successfully applied. Usability scored high (mean 6.39/7), with improvements needed in reminders, personalization, and chatbot comprehension. In the feasibility and effectiveness study conducted, from among 130 invited participants, only 34 signed up, reporting anxiety (100%), depression (94%), and burnout (65%). Vickybot demonstrated usability, satisfaction, and acceptability but highlighted areas for enhancement. Notably, the authors report that Vickybot successfully identified emergency situations involving suicidal thoughts, facilitating timely interventions. However, while the chatbot showed effectiveness in alleviating work-related burnout, its impact on anxiety and depression was less pronounced. Importantly, the Vickybot app does not appear to be publicly available for download, indicating that further development and testing are required before it can be widely implemented as a mental health support tool.

The average age of the studied sample was 30.90 years for Emohaa explored by Sabour *et al.*^[19]. Emohaa is a mental health chatbot designed to reduce mental distress among users in China, available on WeChat. The chatbot comprises two main platforms: Cognitive Behavioral Therapy Chatbot (CBT-Bot): This rule-based component follows CBT principles, providing users with exercises like automatic thinking corrections and guided expressive writing. Users select options in scenarios and report their mood after completing exercises. Emotional Support Chatbot (ES-Bot): This AI-driven version employs a BERT-based model, generating responses tailored to users' emotional needs. It allows discussions about personal issues and classifies messages to identify signs of suicidal thoughts, prompting appropriate emergency responses. The study found significant reductions in depression, negative affect, and insomnia among users of Emohaa, measured by the PHQ-9, PANAS, and ISI questionnaires. Participants, all from Mainland China, had an average of 7.87 years of work experience (SD = 8.45). Baseline mental distress levels were moderate, with depression (PHQ-9: M = 16.43, SD = 5.01), anxiety (GAD-7: M = 16.23, SD = 4.37), and insomnia (M = 16.45, SD = 5.38). Positive and negative effects were assessed using the PANAS, with participants showing moderate positive effect (M = 24.76, SD = 7.20) and negative effect (M = 22.34, SD = 6.35). ANOVA and chi-squared tests were used to examine differences in baseline variables (age, gender, PHQ-9, GAD-7, PA, NA, insomnia) among the three groups: control, CBT bot, and ES bot. The results indicated no significant differences in baseline demographics (age: $F = 2.17$, $p = 0.117$; gender: $X^2 = 3.56$, $p = 0.173$) or mental distress variables (PHQ-9: $F = 2.45$, $p = 0.088$; GAD-7: $F = 0.93$, $p = 0.396$). These results indicate the chatbot's effectiveness and its potential as a valuable resource for mental health support in the public domain. Emohaa's dual functionality and accessibility underscore its relevance in addressing mental health issues, making it an essential reference for our work.

The chatbots available for public use are as follows: In a series of studies conducted under the Northern Periphery and Arctic Programme [NPAP],^[20] the ChatPal project proposed a non-commercial chatbot available as an Android and iOS app, primarily targeting the mental health and well-being of rural populations. Although ChatPal was developed prior, several recent studies have been submitted regarding the chatbot, as follows: In study published by Potts *et al.*^[21] on ChatPal, a multilingual digital mental health chatbot available in English, Scottish Gaelic, Swedish, and Finnish, involved a multicenter pre-post intervention design with 348 participants, utilizing standardized outcome measures such as the Short Warwick-Edinburgh Mental Well-Being Scale and the World Health Organization-Five Well-Being

Index. Evaluated at baseline, midpoint, and endpoint, the results indicated that ChatPal has the potential to complement other digital and face-to-face services in promoting mental well-being. However, the authors emphasized the need for further research to assess the effectiveness of the methods employed. This highlights the growing trend toward multilingual and culturally inclusive mental health support solutions, catering to diverse populations. Booth *et al.*^[22] conducted a study analysing user event logs for the ChatPal mental health and well-being chatbot, focusing on usage patterns and feature associations. Utilizing a k-means clustering algorithm, the researchers examined anonymized login data from 1,403 users between January 24, 2022, and June 22, 2022, ultimately narrowing their analysis to 579 adult users over 18 years old. Among these, 348 participants were specifically recruited for a 12-week pre-post study, with approximately 67 percent identifying as female. The analysis revealed three distinct user clusters: abandon, sporadic, and transient users, providing insights into engagement levels and usage behaviours. Notably, the feature "Treat yourself as a friend" received the highest positive ratings, suggesting that personal and relatable features may enhance user satisfaction and engagement. This research underscores the importance of understanding user interaction patterns to improve chatbot functionality and efficacy in delivering mental health support.

MacNeill *et al.*^[23] examined the effectiveness of a mental health chatbot, Wysa, a commercially available conversational agent for public use via Website for Wysa,^[24] for individuals with chronic diseases, specifically arthritis and diabetes. In their randomized controlled trial involving 68 participants, those using the Wysa chatbot formed the treatment group, while others served as the control group. The findings indicate that mental health chatbots can provide effective support for individuals managing chronic conditions. Despite being cost-effective and accessible, the study notes limitations in these programs, suggesting they may not be suitable for everyone. This underscores the need for tailored approaches in digital mental health interventions to better meet diverse user needs.

Suharwardy *et al.*^[25] conducted a randomized controlled trial to assess the feasibility and impact of the Woebot app, another non-commercial smartphone application, available on Android and iOS, on postpartum mental health among women. Out of 192 participants, 68 utilized the chatbot for mental health assessment during their postpartum period. The findings indicated that there was not a significant change in mental health outcomes for those who used the app compared to those who did not. The study concluded that while the use of the chatbot was acceptable among women in the early postpartum period, the lack of positive screening for

depression at baseline limited the chatbot's ability to demonstrate effectiveness in reducing depressive symptoms

In the book chapter, Negi^[26] explores various AI-powered interventions aimed at improving women's mental health, highlighting tools such as Moodfit. Moodfit is a mobile application that delivers a personalized mental wellness program, encompassing mindfulness exercises, mood tracking, and goal-setting features. Accessible via getmoodfit.com and available on both iOS and Android platforms, Moodfit offers a range of activities, including breathing exercises and Cognitive Behavioral Therapy (CBT) thought records. Users can also maintain a mood journal and receive reminders, with data visualizations like scatterplots to track mood, nutrition, and other variables over time. Additionally, the app includes positive quotes and educational resources aimed at fostering a positive mindset. Its user-friendly interface contributes to a satisfying overall user experience, making it a valuable tool for many. Although it is available for public use, no research publications for Moodfit were found that provide evidence on the evaluation metric and the usability of the chatbot.

In conclusion it is noted that only a limited number of chatbots have been developed within the timeframe of this review, with very few undergoing public trials such as randomized controlled trials (RCTs). The ones that are currently available for public use include ChatPal, Happify, Moodfit, Wysa, and Woebot. Each of these chatbot has shown promise in addressing mental health needs, leveraging evidence-based techniques to provide support and engage users effectively. ChatPal, for instance, targets the mental health and well-being of rural populations, offering tailored interventions. Happify employs evidence-based activities and games to improve emotional health. Moodfit focuses on tracking moods and providing tools for mental wellness. Wysa and Woebot integrate cognitive-behavioral therapy principles to help users manage anxiety and depression through conversational interactions. In conclusion, while there is a growing number of chatbots developed for mental health support, the evidence of their efficacy is still emerging. The limited public trials and the variability in their availability underscore the need for further research and evaluation. As these technologies evolve, ongoing assessment will be crucial to ensure they effectively meet the needs of users and contribute positively to mental health interventions.

4.2 Chatbot in general health

In this sub-section, we explore the development of various AI-powered chatbots designed for different healthcare applications, ranging from preventive health monitoring to disease-specific information delivery. Despite the promising potential of these chatbots, a

common theme emerges regarding their availability and readiness for general use. For instance, the AI-powered educational chatbot for radiotherapy,^[27] which utilizes Natural Language Processing (NLP) to predict user intent and offers a list of suggested questions for improved interaction, shows no evidence of pilot testing or public rollout. This limits its applicability to local or experimental settings. Additionally, the chatbot employs IBM Watson Assistant and integrates with third-party platforms like WhatsApp, Slack, and Facebook Messenger, though there is no information regarding wider deployment or user feedback. Further, Text-to-GraphQL^[28] aimed at optimizing medical question-answering systems, stands out for its algorithmic approach that converts natural language into GraphQL queries for graph databases like Neo4j. This system is designed to handle complex medical relationships and allow for easier updates compared to traditional relational databases. However, the study focuses predominantly on algorithmic development, without addressing pilot testing or user deployment, leaving its real-world application unclear. In a similar vein, GastroBot,^[29] a chatbot designed to answer knowledge-based questions on gastrointestinal diseases, leverages a fine-tuned gte-base-zh model adapted specifically for this domain. The chatbot achieves high context recall (95%), faithfulness (93.73%), and relevance (92.28%) based on the RAGAS framework, demonstrating its strong technical performance. Despite this, there is no mention of pilot testing or plans for public deployment, limiting its use to controlled environments rather than broad public access. Lastly, the Hungarian-language medical chatbot,^[30] which assists with health-related inquiries by recognizing symptoms and suggesting relevant treatments, uses over 1,500 symptom-disease records trained with LSTM algorithms. The research highlights its potential to expand with a larger dataset and more advanced models like BERT, but it lacks discussion of broader deployment or further testing, underscoring the gap between algorithmic advancements and real-world applications. In terms of current ongoing work, Chinese-speaking MSM aged 18 years or more, having access to live chat applications are being recruited to participate in a parallel-group, noninferiority randomized controlled trial, through outreach, online advertisements, and peer referrals.^[31] aims to evaluate an HIV self-testing chatbot designed to address HIV testing concerns and improve uptake. The chatbot employs a knowledge graph and machine learning for adaptive interaction, making it highly adaptable. This study's rigor and use of advanced AI tools set it apart, but its impact on HIV testing uptake and scalability will determine its success. Maia *et al.* proposes the GECA chatbot, powered by artificial intelligence, is designed to address the gap in preventive healthcare chatbots by providing medical information

and helping monitor users' health.^[32] Specifically, it focuses on two diseases, COVID-19 and dementia. The chatbot is said to be available as a mobile app for the Android platform, supporting both English and Portuguese language interactions. However, despite searching on the Play Store, the app could not be found, and there is no information regarding any randomized controlled trials (RCTs), pilot tests, or the same. Although the chatbot aims to improve health monitoring, its availability and readiness for general use are unclear. Chagas *et al.* aimed to evaluate the quality of user experience with a COVID 19 chatbot developed by a telehealth service in Brazil, focusing on its usability and identifying strengths and shortcomings through real-user feedback in simulated scenarios.^[33] The chatbot, named Ana, was shown as a widget on the portal.^[34] The chatbot, designed to assist with COVID-19 symptom severity screening and provide evidence-based health information, was integrated into a local public health service workflow. Between October 2020 and January 2021, they conducted a mixed-methods evaluation, combining a post task usability survey with interviews from 63 users and 15 volunteers engaged in simulated interactions. The usability survey revealed high satisfaction with the chatbot's usefulness, ease of use, and user satisfaction. However, interviews identified 6 positive aspects and 15 issues, categorized into usability and health support. While users found the chatbot beneficial, particularly in its health support role, issues in design and usability were highlighted. Despite the chatbot's successful use during the pandemic's early stages, some gaps are revealed in its design and a lack of a long-term improvement strategy. These results suggest that while the chatbot was effective in addressing immediate public health needs, its future integration into health care systems requires further development.

These studies, while rich in algorithmic and technical advancements, share the challenge of limited public deployment, lack of pilot testing, and unclear plans for widespread availability, which hinder their broader application in real-world healthcare settings. Furthermore, many of these chatbots focus on specific domains or localized implementations, highlighting the need for further development and evaluation before they can be scaled for public use.

The development and evaluation of AI-powered healthcare chatbots span diverse applications, including patient education, disease management, and health monitoring, with a shared focus on improving healthcare access and outcomes. Here, we explore chatbots that have undergone preliminary testing, and/or randomized controlled trials (RCTs), or user acceptance studies to assess their effectiveness and/or usability in real world or trial settings. 301 participants aged 18-29 years were enrolled at baseline for the study of Nena, explored in

Njogu *et al.* study.^[35] Nena is a chatbot designed to enhance sexual and reproductive health and rights (SRHR) awareness among young people in Kenya. The chatbot provides useful services, such as a health facility geo-locator, and uses Google DialogFlow to deliver SRHR information via platforms like Facebook Messenger and WhatsApp. The chatbot's structured decision-tree interface allows users to navigate topics through numeric responses, promoting an engaging and informative experience. Despite its strengths, the chatbot faces limitations, such as the lack of verification for the accuracy of its domain information and uncertainty around its availability, which could impact its credibility and broader reach. In a mixed-methods study conducted between November 2021 and January 2022, the chatbot demonstrated high acceptability and led to improvements in SRHR attitudes and behaviors among Kenyan young adults. Participants reported increased confidence in discussing contraception and sexual feelings with partners, as well as improved sex-positive communication and safer sex practices. The study highlights that integrating sexual pleasure into SRHR content through digital tools could be a promising strategy for improving SRHR knowledge, empowerment, and behaviors among youth.

With a median age of 32 years, 57 (32 were PWH (People with Haemophilia), nearly all PWH had severe haemophilia A and 25 were parents or relatives) participants were enrolled in the study for Saytù Hemophilie. Babington-Ashaye *et al.* developed Saytù Hemophilie, to improve health literacy for people with hemophilia in Senegal, the mobile app employs both text and speech-based communication to answer health-related questions.^[36] With high usability ratings and a potential for cross-country adaptation, Saytù stands out for its strong cultural and linguistic adaptability, offering a scalable solution for health education in underserved regions. While the app shows promise for widespread deployment in various African countries, the study did not provide details on future for scaling or public access, leaving questions about its broader application.

With Participants in age group 46.0-60.5, Lucy LiverBot was tested in Au *et al.* study.^[37] Lucy Liverbot developed to provide information on disease, medication, and nutrition for patients with decompensated chronic liver disease, integrating voice input and visual elements, offering a multifaceted approach to patient education. During its beta testing, 20 participants (out of which 11 were female) provided feedback, highlighting the chatbot's potential to improve health literacy, though some accessibility issues were reported. While promising, Lucy LiverBot's readiness for broad public use remains uncertain as it is still in the prototype phase, and there is no mention of a planned wider rollout. Ten men aged 49 to 81 years with suspicion of prostate cancer

(PC) were enrolled in the study for PROSCA (prostate cancer communication assistant), developed for provisioning patient information about early detection of PC, as cited in Görtz *et al.* study.^[38] PROSCA, a health information chatbot, designed to support communication between doctors and prostate cancer patients, shows strong usability with high satisfaction rates among older patients (median age 68). Despite its success in improving patient engagement, PROSCA's deployment appears to be restricted to specific user groups, with no mention of broader availability or implementation. The findings, though positive, point to the chatbot's potential for expanding its role in healthcare communication, yet questions about scalability and public access remain.

Haris, a chatbot tested with men who have sex with men (MSM) in Malaysia^[39], was developed to provide information on HIV testing, mental health, and preexposure prophylaxis (PrEP). Beta testing was conducted with 14 MSM from February to April 2022 using Zoom application and involved three steps: a 45-minute human-chatbot interaction using the think-aloud method, a 35-minute semi structured interview, and a 10-minute web-based survey. The first two steps were recorded, transcribed verbatim, and analyzed using the Unified Theory of Acceptance and Use of Technology. Emerging themes from the qualitative data were mapped onto the four domains of the framework: performance expectancy, effort expectancy, facilitating conditions, and social influence. Despite the small sample size and limited testing group, Haris demonstrated high acceptability, with 93% of participants rating its design and usability positively. However, like many other chatbots reviewed, Haris faces limitations related to availability, with no planned public rollout and concerns regarding long-term accessibility. The study suggests that while the chatbot can be useful tool for MSM in Malaysia, its future implementation remains uncertain, and further testing is required.

Lastly, a study by de Queiroz *et al.* evaluated the Smart Monitoring Tool (SMT), a wearable-integrated chatbot designed to monitor colorectal cancer patients undergoing chemotherapy.^[40] This prospective, non-randomized clinical study employed AI and the Internet of Things (IoT) to assess the effectiveness of a new computational model in the active treatment phase. Over eight weeks, patients self-reported symptoms, adverse effects, physical activity, and diet. The results showed that the intervention group reported symptoms with higher accuracy (92.3% vs. 64.7% in the control group) and engaged more actively with physical activity. Most patients (61.5%) interacted with the chatbot for over 62.5% of the study period. The model contributed to more accurate data collection and increased patient involvement in managing symptoms and treatment side effects. Additionally, the integration of light physical

activity was enhanced. The User Experience Questionnaire (UEQ) and System Usability Scale (SUS) scores indicated that the model met patient expectations and demonstrated acceptable usability. However, the study's small sample size (13 patients) and lack of information on broader deployment raise questions about the system's scalability and generalization to larger populations.

Across all these studies, a recurring challenge is the lack of clarity regarding public deployment and limited sample sizes, which restrict the ability to assess the effectiveness of these chatbots on a larger scale. Despite promising AI enabled features and adaptability to various health conditions, many of these chatbots are still in prototype or early testing stages. While some show significant potential for enhancing health literacy, improving patient engagement, and providing real-time health support, the future of these systems depends on their ability to overcome usability challenges, scale for broader implementation, and ensure long-term accessibility. These studies collectively underscore the need for further development, pilot testing, and evaluation before these chatbots can be widely adopted and integrated into public healthcare settings.

5. Role of generative AI in healthcare: potential and limitations

Although large language models (LLMs) are not the primary focus of this review, their growing popularity, coupled with the increasing research in this area, warrants brief attention. While the exploration of LLMs in healthcare is not yet extensive, we have tried to highlight their potential applications and provide an overview of their use in healthcare settings. The integration of generative AI, particularly large language models (LLMs) like ChatGPT, into healthcare has garnered increasing interest, with studies evaluating their performance across diverse clinical scenarios. Research shows that while AI systems can provide valuable support, they face limitations in accuracy and reliability, particularly in complex, high-stakes medical conditions. For instance, studies in emergency medicine,^[41] ophthalmology,^[42] and voice treatment decision-making. Dronkers *et al.* highlight that AI's clinical accuracy often falls short, especially in specialized conditions.^[43] However, advancements in AI for domain-specific applications, such as gastrointestinal imaging, offer promise. A recent study by Rau *et al.*^[44] demonstrated that a gastrointestinal imaging-aware chatbot (GIA-CB), powered by GPT-4 and enhanced via the Llama Index framework, significantly outperformed a generic GPT-4 chatbot in providing differential diagnoses based on imaging descriptions. The GIA-CB delivered accurate primary and differential diagnoses in 90% of cases, with formulated rationale and source references,

emphasizing the importance of knowledge retrieval for explainable and trustworthy AI in clinical decision-making. The findings from Goodman *et al.* work^[45] align with the growing body of research evaluating the accuracy and reliability of AI systems in healthcare. In their study, GPT-based chatbots (GPT-3.5 and GPT-4) were tested for accuracy and completeness in answering 284 medical queries across 17 specialties, with promising results. Median accuracy and completeness scores were high, especially for GPT-4, which demonstrated significant improvements over GPT-3.5 and over time upon re-evaluation. While easy questions received near-perfect scores, harder queries showed more variability, highlighting areas where AI systems still require refinement. The results reinforce the notion that while AI has potential to augment clinical decision-making, its limitations, particularly in nuanced or complex scenarios, necessitate ongoing validation and oversight. Parikh *et al.* study highlights the variability and inaccuracies in AI chatbot-generated recommendations for oculoplastic surgeons.^[46] Among 539 suggested physicians, only 64.1% were oculoplastic specialists, with 38% being either non-existent or not practicing in the specified city. Gender representation was skewed, with only 27.7% of recommendations being female. Prompt phrasing significantly influenced results, with terms like “eyelid lifts” yielding more general plastic surgeons and ENTs instead of oculoplastic specialists. These findings underscore the need to address biases and inaccuracies in AI-driven healthcare recommendations, especially as patient reliance on such tool’s increases. Despite these advancements, challenges remain, as AI-generated recommendations can lack depth or consistency with guidelines, as seen in comparisons with established resources like UpToDate.^[47] Nevertheless, AI tools demonstrate strong potential to enhance healthcare access, particularly in resource-limited settings^[48] and increase accessibility for stigmatizing health concerns.^[49] Future developments will require rigorous validation, continuous learning, and collaboration between developers and healthcare professionals to refine accuracy, mitigate biases, and optimize patient outcomes.

6. Discussion

As Friedman ^[50] pointed out in his article on Clark Stanley's snake oil scam remedy, the authors point to the similarly opaque, 'black box' character of sophisticated machine learning algorithms, which tend to defy simplistic interpretation because of their intricate parameters. Given the very delicate nature of the health domain, proper caution must be taken to ensure no casualties are caused as an effect of using such technologically advanced interventions. Moreover, issues like data security and protection, and other safety ethics must be prioritized for any of such interventions.

Advancements in large language models have led to the growing use of chatbot in various fields, including healthcare. One example is the Simsimi chatbot, which uses an AI engine for communication but is not specifically tailored for domains like mental health. Chin *et al.* study explores the potential of using this chatbot for mental health support by analysing user conversations that include terms like “sad” and “depressed.”^[51] While the chatbot demonstrates promise, it does not yet exhibit the level of intelligence needed for such sensitive topics. Notably, users of chatbots like Simsimi do not typically seek social support for life difficulties as they might on social media, yet they expect features like active listening skills and a safe space to express emotions like sadness or depression. The study analysing 152,783 conversations from both Western (Canada, the UK, the US) and Eastern countries (Indonesia, India, Malaysia, the Philippines, and Thailand) found that Eastern users were more likely to use words related to sadness, while Western users discussed more vulnerable topics like mental health and sensitive issues such as death and swear words. While these findings point to the potential of chatbots in the mental health space, they also highlight the need for professional intervention and robust privacy protections. These studies reflect a broader trend in the research on chatbot across various domains, where user preferences and the adaptability of technology remain central to its effectiveness.

Several other studies underscore the importance of tailoring chatbot to specific user needs while addressing challenges like emotional engagement and privacy.^[52] For example, investigates the use of chatbots in delivering genetic services, emphasizing that while patients are generally receptive to chatbots for some aspects of genetic testing, a “safety net” in the form of a care provider is essential when more complex issues arise. Similarly, Park *et al.* examined how emotional disclosure through AI chatbots impacts user satisfaction, finding that chatbots that facilitate emotional expression were perceived as more engaging.^[53] However, the lack of publicly available information on these tools limits the generalizability of the results. Bowman *et al.* also explored how the tone of chatbot interactions, specifically politeness, influences user experiences in mood logging for mental health.^[54] They found that while politeness could enhance supportive interactions, it could also lead to feelings of distrust or condescension if not carefully balanced. Other studies, such as Schmitz and Becker highlight the utility of chatbots in niche areas like dementia caregiving,^[55] where personalized information and user engagement are key to their success. While some chatbots, such as PharmindBot^[56] have proven effective for educational and data collection purposes in healthcare, their impact on patient care remains unclear. Trials such as Al-Hilli *et al.*^[57] randomized study on

genetic counselling chatbots suggest that AI can match traditional counselling in certain contexts, but further research is needed to ensure broader applicability. Finally, emerging technologies, such as mindfulness-based stress reduction via chatbots and virtual humans, also show potential. While virtual humans demonstrated better engagement, chatbots lagged in adherence to tasks, underscoring the need for continued development to enhance the perceived empathy and effectiveness of these tools in diverse healthcare settings. Together, these studies suggest that while chatbot have significant promise in healthcare, their development must carefully consider user preferences, emotional needs, and the complexities of sensitive issues like mental health, ensuring that these tools are both effective and ethically sound. Mobile applications, particularly chatbot-based platforms like Woebot, Wysa, and Anna, are gaining popularity for providing mental health support,^[58] with Anna being developed by Happify Health to offer innovative digital mental health care.^[59] Anna, available on happify.com and the Play Store, features a user-friendly interface with various activities like guided meditations, quizzes, and reflective prompts aimed at fostering mindfulness. However, despite its accessibility and ease of use, the app's simplistic structure and limited therapeutic depth suggest it may not replace traditional therapy. This trend highlights the growing popularity of mobile mental health tools, but also underscores the need for appropriate security measures, especially considering the risk of young children accessing such apps. To ensure user privacy and data protection, implementing features like numeric lock or other forms of authentication is strongly recommended.

7. Conclusion

In conclusion, the increasing interest in utilizing chatbots for mental and general health care management is evident, with several publicly available chatbots for mental health support, such as Chatpal, Wysa, and Woebot. Numerous publications highlight the ongoing efforts to improve these systems, particularly in the mental health space, where randomized controlled trials (RCTs) have demonstrated their potential for broader public use, provided they undergo necessary advancements and refinements. However, in the domain of physical health care, we found no chatbots that are ready for public use, with most still in user acceptance or beta testing phases. This underscores a significant gap in the development and implementation of chatbots for physical health care, highlighting the need for further innovation and progress in this area.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-

profit sectors.

Data Availability Statement

No data were generated or analyzed during the current study. Therefore, data sharing is not applicable to this article.

Conflict of Interest

There is no conflict of interest.

Artificial Intelligence (AI) Use Disclosure

The authors declare that artificial intelligence (AI)-assisted tools were used only for language refinement, grammar improvement, and manuscript structuring purposes during the preparation of this work. All technical content, experimental implementation, results, and interpretations were independently developed and verified by the authors.

Supporting Information

Not applicable.

References

- [1] U.S. Government Accountability Office, Artificial intelligence in health care: Benefits and challenges of technologies to augment patient care, 2021, <https://www.gao.gov/products/gao-21-7sp>, Accessed 10 October 2024.
- [2] GOV.UK, Guidance software and artificial intelligence (ai) as a medical device, 2024, <https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device>, Accessed 10 October 2024.
- [3] W. A. Haseltine, The chatbot revolution: Transforming healthcare with ai language models, 2023, <https://www.forbes.com/sites/williamhaseltine/2023/10/18/the-chatbot-revolution-transforming-healthcare-with-ai-language-models/>, Accessed 2 November 2024.
- [4] Katendejericho, The impact of ai-powered chatbots on mental health: An evidence-based analysis, 2024, <https://medium.com/@katendejericho5/the-impact-of-ai-powered-chatbots-on-mental-health-an-evidence-based-analysis-884b6717e656>, Accessed 2 November 2024.
- [5] J. Weizenbaum, Eliza-a computer program for the study of natural language communication between man and machine, *Communications of the ACM*, 1966, 9, 36-45, doi: 10.1145/365153.365168.

- [6] K. M. Colby, Artificial paranoia; a computer simulation of paranoid processes, 1975, doi: 10.1016/C2013-0-02631-X.
- [7] R. S. Wallace, The alice artificial intelligence foundation, Proceedings of the 14th International Conference on Artificial Intelligence, 1995.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems, 2017, doi: 10.48550/arXiv.1706.03762.
- [9] A. Ahmed, A. Hassan, S. Aziz, A. A. Abd-alrazaq, N. Ali, M. Alzubaidi, D. Al-Thani, B. Elhusein, M. A. Siddig, M. Ahmed, M. Househ, Chatbot features for anxiety and depression: A scoping review, *Health Informatics Journal*, 2023, **29**, 146045822211467, doi: 10.1177/14604582221146719.
- [10] M. Phiri, A. Munoriyarw, Health chatbots in africa: scoping review, *Journal of Medical Internet Research*, 2023, **25**, e35573, doi: 10.2196/35573.
- [11] A. van Heerden, S. Bosman, D. Swendeman, and W. S. Comulada, Chatbots for hiv prevention and care: a narrative review, *Current HIV/AIDS Reports*, 2023, **20**, 481–486, doi: 10.1007/s11904-023-00681-x.
- [12] A. Mukhtar, D. Gupta, S. Parida, A reinforcement learning approach for intelligent conversational chatbot for enhancing mental health therapy, *Procedia Computer Science*, 2024, **235**, 916–925, doi: 10.1016/j.procs.2024.04.087.
- [13] S. Karkosz, R. Szymański, K. Sanna, J. Michałowski, Effectiveness of a web-based and mobile therapy chatbot on anxiety and depressive symptoms in subclinical young adults: Randomized controlled trial, *JMIR Formative Research*, 2024, **8**, e47960, doi: 10.2196/47960.
- [14] E. Kleinau, T. Lamba, W. Jaskiewicz, K. Gorentz, I. Hungerbuehler, D. Rahimi, D. Kokota, L. Maliwichi, E. Jamu, A. Zumazuma, M. Negrão, R. Mota, Y. Khouri, M. Kapps, Effectiveness of a chatbot in improving the mental wellbeing of health workers in malawi during the covid-19 pandemic: A randomized, controlled trial, *PLOS ONE*, 2024, **19**, e0303370, doi: 10.1371/journal.pone.0303370.
- [15] A. Kang, S. Hetrick, T. Cargo, S. Hopkins, N. Ludin, S. Bodmer, K. Steven son, C. Holt-Quick, K. Stasiak, Exploring young adults' views about aroha, a chatbot for stress associated with the covid-19 pandemic: Interview study among students, *JMIR Formative Research*, 2023, **7**, e44556, doi: 10.2196/44556.
- [16] L. Chun-Hung, L. Guan-Hsiung, Y. Wu-Chuan, L. Yu-Hsin, Chatbot-assisted therapy for patients with methamphetamine use disorder: a preliminary randomized controlled trial, *Frontiers in Psychiatry*, 2023, **14**, 1159399, doi: 10.3389/fpsyt.2023.1159399.
- [17] A. O. Thunström, H. K. Carlsen, L. Ali, T. Larson, A. Hellström, S. Steingrimsson, Usability comparison among healthy participants of an anthropomorphic digital human and a text-based chatbot as a responder to questions on mental health: Randomized controlled trial, *JMIR Human Factors*, 2024, **11**, e54581, doi: 10.2196/54581.
- [18] G. Anmella, M. Sanabra, M. Primé-Tous, X. Segú, M. Caverio, I. Morilla, I. Grande, V. Ruiz, A. Mas, I. Martín-Villalba, A. Caballo, J. P. Esteva, A. Rodríguez-Rey, F. Piazza, F. J. Valdesoiro, C. Rodríguez-Torrella, M. Espinosa, C. Virgili, C. Sorroche, A. Ruiz, A. Solanes, J. Radua, M. A. Also, E. Sant, S. Murgui, M. Sans-Corrales, A. H. Young, V. Vicens, J. Blanch, E. Caballeria, H. López-Pelayo, C. López, V. Olivé, L. Pujol, S. Quesada, Solé B, Torrent C, Martínez-Aran A, Guarch J, Navinés R, Murru A, Fico G, de Prisco M, Oliva V, Amoretti S, Pio-Carrino C, Fernandez-Canseco M, Villegas M, Vieta E, Hidalgo-Mazzei D, Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and health care professionals: development, feasibility, and potential effectiveness studies, *Journal of Medical Internet Research*, 2023, **25**, e43293, doi: 10.2196/43293.
- [19] S. Sabour, W. Zhang, X. Xiao, Y. Zhang, Y. Zheng, J. Wen, J. Zhao, M. Huang, A chatbot for mental health support: exploring the impact of emohaa on reducing mental distress in china, *Frontiers in Digital Health*, 2023, **5**, doi: 10.3389/fdgth.2023.1133987.
- [20] Welcome to chatpal, <https://chatpal.interreg-npa.eu/>, Accessed 24 October 2024.
- [21] C. Potts, F. Lindström, R. Bond, M. Mulvenna, F. Booth, E. Ennis, K. Parding, C. Kostenius, T. Broderick, K. Boyd, A.-K. Vartiainen, H. Nieminen, C. Burns, A. Bickerdike, L. Kuosmanen, I. Dhanapala, A. Vakaloudis, B. Cahill, M. MacInnes, M. Malcolm, S. O'Neill, A multilingual digital mental health and well-being chatbot (chatpal): Pre-post multicenter intervention study, *Journal of Medical Internet Research*, 2023, **25**, e43051, doi: 10.2196/43051.
- [22] F. Booth, C. Potts, R. Bond, M. Mulvenna, C. Kostenius, I. Dhanapala, A. Vakaloudis, B. Cahill, L. Kuosmanen, E. Ennis, A mental health and well-being chatbot: user event log analysis, *JMIR mHealth and uHealth*, 2023, **11**, e43052, doi: 10.2196/43052.
- [23] L. MacNeill, S. Doucet, A. Luke, Effectiveness of a mental health chatbot for people with chronic diseases: Randomized controlled trial, *JMIR Formative Research*, 2024, **8**, e50025, doi:

- 10.2196/50025.
- [24] Your AI companion for mental health, <https://www.wysa.com>, Accessed 24 November 2024.
- [25] S. Suharwardy, M. Ramachandran, S. A. Leonard, A. Gunaseelan, D. J. Lyell, A. Darcy, A. Robinson, A. Judy, Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial, *AJOG Global Reports*, 2023, **3**, 100165 doi: 10.1016/j.xagr.2023.100165.
- [26] R. Negi, Improving women's mental health through AI-powered interventions and diagnoses, In *Artificial Intelligence and Machine Learning for Women's Health Issues*, Academic Press, 2024, 173-191.
- [27] J. C. L. Chow, L. Sanders, K. Li, Design of an educational chatbot using artificial intelligence in radiotherapy, *AI*, 2023, **4**, 319-332, doi: 10.3390/ai4010015.
- [28] P. Ni, R. Okhrati, S. Guan, V. Chang, Knowledge graph and deep learning-based text to graphql model for intelligent medical consultation chatbot, *Information Systems Frontiers*, 2024, **26**, 137-156, doi: 10.1007/s10796-022-10295-0.
- [29] Q. Zhou, C. Liu, Y. Duan, K. Sun, Y. Li, H. Kan, Z. Gu, J. Shu, J. Hu, Gastrobot: a chinese gastrointestinal disease chatbot based on the retrieval-augmented generation, *Frontiers in Medicine*, 2024, **11**, 1392555, doi: 10.3389/fmed.2024.1392555.
- [30] B. Simon, A. Hartveg, L. Dénes-Fazakas, G. Eigner, L. Szilagyi, Advancing medical assistance: Developing an effective hungarian-language medical chatbot with artificial intelligence, *Information*, 2024, **15**, 297, doi: 10.3390/info15060297.
- [31] S. Chen, Q. Zhang, C. kit Chan, F. yuen Yu, A. Chidgey, Y. Fang, P. K. H. Mo, Z. Wang, Evaluating an innovative hiv self-testing service with web-based, real-time counselling provided by an artificial intelligence chatbot (hivst-chatbot) in increasing HIV self-testing use among Chinese men who have sex with men: Protocol for a noninferiority randomized controlled trial, *JMIR Research Protocols*, 2024, **12**, e48447, doi: 10.2196/48447.
- [32] E. Maia, P. Vieira, I. Praça, Empowering preventive care with geca chatbot, *Healthcare*, 2023, **11**, 2532, doi: 10.3390/healthcare11182532.
- [33] B. A. Chagas, A. S. Pagano, R. O. Prates, E. C. Praes, K. Ferregueti, H. Vaz, Z. S. N. Reis, L. B. Ribeiro, A. L. P. Ribeiro, T. M. Pedroso, A. Beleigoli, C. R. A. Oliveira, M. S. Marcolino, Evaluating user experience with a chatbot designed as a public health response to the covid-19 pandemic in brazil: Mixed methods study, *JMIR Human Factors*, 2023, **10**, e43135, doi: 10.2196/43135.
- [34] Control de telessaude, <https://telessaude.hc.ufmg.br>, Accessed 10 October 2024.
- [35] J. Njogu, G. Jaworski, C. Oduor, A. Chea, A. Malmqvist, C. W. Rothschild, Assessing acceptability and effectiveness of a pleasure-oriented sexual and reproductive health chatbot in kenya: an exploratory mixed-methods study, *Sexual and Reproductive Health Matters*, 2023, **31**, doi: 10.1080/26410397.2023.2269008.
- [36] A. Babington-Ashaye, P. de Moerloose, S. Diop, A. Geissbuhler, Design, development and usability of an educational AI chatbot for people with haemophilia in Senegal, *Clinical Haemophilia*, 2023, **29**, 1063-1073, doi: 10.1111/hae.14815.
- [37] J. Au, C. Falloon, A. Ravi, P. Ha, S. Le, A beta-prototype chatbot for increasing health literacy of patients with decompensated cirrhosis: Usability study, *JMIR Human Factors*, 2023, **10**, e42506, doi: 10.2196/42506.
- [38] M. Görtz, K. Baumgärtner, T. Schmid, M. Muschko, P. Woessner, A. Gerlach, M. Byczkowski, H. Sülmann, S. Duensing, M. Hohenfellner, An artificial intelligence-based chatbot for prostate cancer education: Design and patient evaluation study, *Digital Health*, 2023, **9**, 20552076231173304, doi: 10.1177/20552076231173304.
- [39] M. H. Cheah, Y. N. Gan, F. L. Altice, J. A. Wickersham, R. Shrestha, N. A. M. Salleh, K. S. Ng, I. Azwa, V. Balakrishnan, A. Kamarulzaman, Z. Ni, Testing the feasibility and acceptability of using an artificial intelligence chatbot to promote hiv testing and pre-exposure prophylaxis in Malaysia: Mixed methods study, *JMIR Human Factors*, 2024, **1**, e52055, doi: 10.2196/52055.
- [40] D. A. de Queiroz, R. S. Passarello, V. V. de Moura Fé, A. Rossini, E. F. da Silveira, E. A. I. F. de Queiroz, C. A. da Costa, A wearable chatbot-based model for monitoring colorectal cancer patients in the active phase of treatment, *Healthcare Analytics*, 2023, 100257, doi: 10.1016/j.health.2023.100257.
- [41] B. Arslan, G. Eyupoglu, S. Korkut, K. A. Turkdogan, E. Altinbilek, The accuracy of ai-assisted chatbots on the annual assessment test for emergency medicine residents, *Journal of Medicine, Surgery, and Public Health*, 2024, **3**, 100070, doi: 10.1016/j.glmedi.2024.100070.
- [42] M. Botross, S. O. Mohammadi, K. Montgomery, C. Crawford, Performance of Google's artificial intelligence chatbot "bard" (now "Gemini") on ophthalmology board exam practice questions,

- Cureus*, 2024, **16**, e57348, doi: 10.7759/cureus.57348.
- [43] E. A. Dronkers, A. Geneid, C. A. Yaghchi, J. R. Lechien, Evaluating the potential of AI chatbots in treatment decision-making for acquired bilateral vocal fold paralysis in adults, *Journal of Voice*, 2024, **39**, 871-881, doi: 10.1016/j.jvoice.2024.02.020.
- [44] S. Rau, A. Rau, J. Nattenmüller, A. Fink, F. Bamberg, M. Reisert, M. F. Russe, A retrieval-augmented chatbot based on gpt-4 provides appropriate differential diagnosis in gastrointestinal radiology: a proof-of-concept study, *European Radiology Experimental*, 2024, **8**, 60, doi: 10.1186/s41747-024-00457-x.
- [45] R. S. Goodman, J. R. Patrinely, C. A. S. Jr, E. Zimmerman, R. R. Donald, S. S. Chang, S. T. Berkowitz, A. Finn, E. Jahangir, E. A. Scoville, T. S. Reese, D. L. Friedman, J. A. Bastarache, Y. F. van der Heijden, J. J. Wright, F. Ye, N. Carter, M. R. Alexander, J. H. Choe, C. A. Chastain, J. A. Zic, S. N. Horst, I. Turker, R. Agarwal, E. Osmundson, K. Idrees, C. M. Kiernan, C. Padmanabhan, C. E. Bailey, C. E. Schlegel, L. B. Chambless, M. K. Gibson, T. J. Osterman, L. E. Wheless, D. B. Johnson, Accuracy and reliability of chatbot responses to physician questions, *JAMA Network Health Informatics*, 2023, **10**, e2336483, doi: 10.1001/jamanetworkopen.2023.36483.
- [46] A. O. Parikh, M. C. Oca, J. R. Conger, A. McCoy, J. Chang, S. Zhang-Nunes, Accuracy and bias in artificial intelligence chatbot recommendations for oculoplastic surgeons, *Cureus*, 2024, **16**, e57611, doi: 10.7759/cureus.57611.
- [47] Z. Karimov, I. Allahverdiyev, O. Y. Agayarov, D. Demir, E. Almu Radova, Chatgpt vs uptodate: comparative study of usefulness and reliability of chatbot in common clinical presentations of otorhinolaryngology-head and neck surgery, *European Archives of Oto-Rhino-Laryngology*, 2024, **281**, 2145-2151, doi: 10.1007/s00405-023-08423-w.
- [48] W. B. Weeks, B. Taliesin, J. M. Lavista, Using artificial intelligence to advance public health, *International Journal of Public Health*, 2023, **68**, 1606716, doi: 10.3389/ijph.2023.1606716.
- [49] D. Branley-Bell, R. Brown, L. Coventry, E. Sillence, Chatbots for embarrassing and stigmatizing conditions: Could chatbots encourage users to seek medical advice?, *Frontiers in Communication*, 2023, **8**, 1275127, doi: 10.3389/fcomm.2023.1275127.
- [50] J. Friedman, How snake oil became a symbol of fraud and deception, *Smithsonian Magazine*, 2024, <https://www.smithsonianmag.com/innovation/how-snake-oil-became-a-symbol-of-fraud-and-deception-180985300/>, Accessed 10 November 2024.
- [51] H. Chin, H. Song, G. Baek, M. Shin, C. Jung, M. Cha, J. Choi, C. Cha, The potential of chatbots for emotional support and promoting mental well-being in different cultures: Mixed methods study, *Journal of Medical Internet Research*, 2023, **25**, e51712, doi: 10.2196/51712.
- [52] S. Luca, M. Clausen, A. Shaw, W. Lee, S. Krishnapillai, E. Adi Wauran, H. Faghfoury, G. Costain, R. Jobling, M. Aronson, E. Liston, J. Silver, C. Shuman, L. Chad, R. Z. Hayeems, Y. Bombard, Finding the sweet spot: a qualitative study exploring patients' acceptability of chatbots in genetic service delivery, *Human Genetics*, 2023, **142**, 321-330, doi: 10.1007/s00439-022-02512-2.
- [53] G. Park, J. Chung, S. Lee, Effect of ai chatbot emotional disclosure on user satisfaction and reuse intention for mental health counselling: a serial mediation model, *Current Psychology*, 2023, **42**, 28663-28673, doi: 10.1007/s12144-022-03932-z.
- [54] R. Bowman, O. Cooney, J. W. Newbold, A. Thieme, L. Clark, G. Doherty, B. Cowan, Exploring how politeness impacts the user experience of chatbots for mental health support, *International Journal of Human-Computer Studies*, 2024, **184**, 103181, doi: 10.1016/j.ijhcs.2023.103181.
- [55] D. Schmitz, B. Becker, Chatbot-mediated learning for caregiving relatives of people with dementia: Empirical findings and didactical implications for multidisciplinary health care, *Journal of Multidisciplinary Healthcare*, 2024, **17**, 219-228, doi: 10.2147/JMDH.S424790.
- [56] R. M. Alkoudmani, G. S. Ooi, M. L. Tan, Implementing a chatbot on facebook to reach and collect data from thousands of health care providers: Pharmindbot as a case, *Journal of the American Pharmacists Association*, 2023, **63**, 1634-1642.e3, doi: 10.1016/j.japh.2023.06.007.
- [57] Z. Al-Hilli, R. Noss, J. Dickard, W. Wei, A. Chichura, V. Wu, K. Renicker, H. J. Pederson, C. Eng, A randomized trial comparing the effectiveness of pre-test genetic counselling using an artificial intelligence automated chatbot and traditional in-person genetic counselling in women newly diagnosed with breast cancer, *Annals of Surgical Oncology*, 2023, **30**, 5990-5996, doi: 10.1245/s10434-023-13888-4.
- [58] M. D. R. Haque, S. Rubya, An overview of chatbot-based mobile mental health apps: insights from app description and user reviews, *JMIR Mhealth Uhealth*, 2023, **11**, e44838, doi: 10.2196/44838.

- [59] Z. Khawaja, J.-C. Bélisle-Pipon, Your robot therapist is not your therapist: understanding the role of ai-powered mental health chatbots, *Frontiers in Digital Health*, 2024, **5**, doi: 10.3389/fdgth.2023.1278186.

Publisher Note: The views, statements, and data in all publications solely belong to the authors and contributors. GR Scholastic is not responsible for any injury resulting from the ideas, methods, or products mentioned. GR Scholastic remains neutral regarding jurisdictional claims in published maps and institutional affiliations.