Accepted Manuscript

Submission Date: 2025-03-07
Accepted Date: 2025-07-01
Accepted Manuscript online: 2025-07-02

Applied Clinical Informatics

Health consumers’ use and perceptions of health information from generative artificial intelligence chatbots: A scoping review

John Robert Bautista, Drew Herbert, Matthew Farmer, Ryan Q De Torres, Gil P Soriano, Charlene Ronquillo.

Affiliations below.

DOI: 10.1055/a-2647-1210

Please cite this article as: Bautista J, Herbert D, Farmer M et al. Health consumers’ use and perceptions of health information from generative artificial intelligence chatbots: A scoping review. ACI 2025. doi: 10.1055/a-2647-1210

Conflict of Interest: The authors declare that they have no conflict of interest.

This study was supported by MU Sinclair School of Nursing, Start-up grant

Abstract:

Background

Health consumers can use generative artificial intelligence (GenAI) chatbots to seek health information. As GenAI chatbots continue to improve and be adopted, it is crucial to examine how health information generated by such tools is used and perceived by health consumers.

Objective

To conduct a scoping review of health consumers’ use and perceptions of health information from GenAI chatbots.

Methods

Arksey and O’Malley’s five-step protocol was used to guide the scoping review. Following PRISMA guidelines, relevant empirical papers published on or after January 1, 2019 were retrieved between February and July 2024. Thematic and content analyses were performed.

Results

We retrieved 3,840 titles and reviewed 12 papers that included 13 studies (quantitative = 5, qualitative = 4, and mixed = 4). ChatGPT was used in 11 studies, while two studies used GPT-3. Most were conducted in the US (n = 4). The studies involve general and specific (e.g., medical imaging, psychological health, and vaccination) health topics. One study explicitly used a theory. Eight studies were rated with excellent quality. Studies were categorized as user experience studies (n = 4), consumer surveys (n = 1), and evaluation studies (n = 8). Five studies examined health consumers’ use of health information from GenAI chatbots. Perceptions focused on: (1) accuracy, reliability, or quality; (2) readability; (3) trust or trustworthiness; (4) privacy, confidentiality, security, or safety; (5) usefulness; (6) accessibility; (7) emotional appeal; (8) attitude; and (9) effectiveness.

Conclusion

Although health consumers can use GenAI chatbots to obtain accessible, readable, and useful health information, negative perceptions of their accuracy, trustworthiness, effectiveness, and safety serve as barriers that must be addressed to mitigate health-related risks, improve health beliefs, and achieve positive health outcomes. More theory-based studies are needed to better understand how exposure to health information from GenAI chatbots affects health beliefs and outcomes.

Corresponding Author:

Dr. John Robert Bautista, University of Missouri, Sinclair School of Nursing, 915 Hitt St, 65211-1300 Columbia, United States, jbautista@missouri.edu

Affiliations:
John Robert Bautista, University of Missouri, Sinclair School of Nursing, Columbia, United States
Drew Herbert, University of Missouri, Sinclair School of Nursing, Columbia, United States
Matthew Farmer, University of Missouri, Sinclair School of Nursing, Columbia, United States
[...]
Charlene Ronquillo, The University of British Columbia Okanagan, Kelowna, Canada

Health consumers' use and perceptions of health information from generative artificial intelligence chatbots: A scoping review

John Robert Bautista, PhD, MPH, RN[1,2*]

Drew Herbert, MSN, MA, APRN[1]

Matthew Farmer, PhD, RN[1]

Ryan Q. De Torres, RN, MA[3]

Gil P. Soriano, PhD, RN[4]

Charlene E. Ronquillo, PhD, RN[5]

[1]Sinclair School of Nursing, University of Missouri-Columbia, USA

[2]Institute for Data Science and Information, University of Missouri-Columbia, USA

[3]College of Nursing, University of the Philippines-Manila, Philippines

[4]Department of Nursing, National University, Philippines

[5]School of Nursing, University of British Columbia Okanagan, Canada

*Corresponding author

John Robert Bautista

jbautista@missouri.edu

Sinclair School of Nursing, University of Missouri-Columbia

915 Hitt St., Columbia, MO 65211

Abstract

Background

Health consumers can use generative artificial intelligence (GenAI) chatbots to seek health information. As GenAI chatbots continue to improve and be adopted, it is crucial to examine how health information generated by such tools is used and perceived by health consumers.

Objective

To conduct a scoping review of health consumers' use and perceptions of health information from GenAI chatbots.

Methods

Arksey and O'Malley's five-step protocol was used to guide the scoping review. Following PRISMA guidelines, relevant empirical papers published on or after January 1, 2019 were retrieved between February and July 2024. Thematic and content analyses were performed.

Results

We retrieved 3,840 titles and reviewed 12 papers that included 13 studies (quantitative = 5, qualitative = 4, and mixed = 4). ChatGPT was used in 11 studies, while two studies used GPT-3. Most were conducted in the US (*n* = 4). The studies involve general and specific (e.g., medical imaging, psychological health, and vaccination) health topics. One study explicitly used a theory. Eight studies were rated with excellent quality. Studies were categorized as user experience studies (*n* = 4), consumer surveys (*n* = 1), and evaluation studies (*n* = 8). Five studies examined health consumers' use of health information from GenAI chatbots. Perceptions focused on: (1)

accuracy, reliability, or quality; (2) readability; (3) trust or trustworthiness; (4) privacy, confidentiality, security, or safety; (5) usefulness; (6) accessibility; (7) emotional appeal; (8) attitude; and (9) effectiveness.

Conclusion

Although health consumers can use GenAI chatbots to obtain accessible, readable, and useful health information, negative perceptions of their accuracy, trustworthiness, effectiveness, and safety serve as barriers that must be addressed to mitigate health-related risks, improve health beliefs, and achieve positive health outcomes. More theory-based studies are needed to better understand how exposure to health information from GenAI chatbots affects health beliefs and outcomes.

confirmed by a neurosurgeon.[3] Despite this case showing both positive (the democratization of health information) and negative (possibility of false hopes) effects of relying on health information from GenAI chatbots, research is needed to identify the implications of exposure to health information from GenAI chatbots, including their significance in altering healthcare decisions.[4]

As GenAI chatbots continue to improve and be adopted, it is crucial to examine how health information generated by such tools is used and perceived by health consumers. However, reviews involving GenAI chatbots in the health domain have focused on their ethical use,[5] healthcare professionals' perspectives on information quality,[6] and ways of enhancing healthcare delivery.[7] Conversely, reviews on health consumers' health information seeking focus on health websites,[8,9] mobile health apps,[10] and social media.[11] To advance research on consumer health informatics, it is pertinent that we synthesize literature on how health consumers use and perceive health information from GenAI chatbots.

Given the novelty of this technology, no study has systematically examined health consumers' use and perceptions of health information from GenAI chatbots. To address this gap, we adopted Arksey and O'Malley's[12] five-step scoping review protocol to identify the research landscape on this topic. Overall, our results offer important insights into advancing research regarding the effect of GenAI chatbots on consumer health informatics.

MATERIALS AND METHODS

Step 1: Identifying research questions

We developed our research questions using the Population-Exposure-Outcome (PEO) Framework.[13] Our target population is health consumers (as defined by the US National Institutes of Health[14]), which includes the general public or lay people. We then focus on studies wherein health consumers were exposed to health information from GenAI chatbots directly (health consumers used a GenAI chatbot to retrieve health information as part of the study) or indirectly (researchers presented health consumers with health information from GenAI chatbots or asked about its use for health information seeking). The target outcomes include the use (including intention) and perceptions of health information from GenAI chatbots. We aimed to answer the following research questions:

RQ1: What are the characteristics of studies on health consumers' use and perceptions of health information from GenAI chatbots?

RQ2: How did health consumers use health information from GenAI chatbots?

RQ3: What are health consumers' perceptions of health information from GenAI chatbots?

Step 2: Identifying relevant studies

A health sciences librarian performed a database search in February 2024 based on search terms provided by JB. Supplementary Appendix 1 lists the ten databases and the corresponding search terms and results. The search was limited to references from January 2019 to February 2024. Although OpenAI's (developer of ChatGPT and the company that made GenAI chatbots mainstream) GPT-1 existed in 2018, they only released the 2019 version (GPT-2) to address misuse concerns.[15] This suggests that most researchers would be able to use it for research in 2019. JB, DH, and MF also performed manual searches between March and July 2024 through reference reviews and Scopus and Google Scholar searches. We used Covidence and Zotero 7 for records screening and management, respectively.

Step 3: Study selection

The inclusion and exclusion criteria were patterned based on the PEO framework described in Step 1. Peer-reviewed empirical papers (i.e., journal articles or conference proceedings) were included based on the following criteria: (1) written in English, (2) involve health consumers, (3) specified a GenAI chatbot, and (4) results reflect health consumers' use or perceptions of health information from a GenAI chatbot. Papers with unclear reference to any GenAI chatbot, results that focus only on performance testing of GenAI, or intention-based findings on using health information from GenAI chatbots were excluded. If a paper reports findings from health professionals and consumers, we included that paper and extracted results from the latter.

Step 4: Charting the data

Fig. 1 shows the PRISMA[16] diagram that illustrates the search process. The initial search yielded 3,840 references based on the database (*n* = 3,831) and manual (Google Scholar = 8; Scopus = 1) searches. After removing 795 duplicates, JB randomly selected 10%[1,17] of the unique references (305 out of 3,045) to test interrater reliability for abstract and title screening. Based on the listed inclusion/exclusion criteria, GS and RD reviewed the references and achieved a moderate agreement (Cohen's κ = 0.67). JB discussed all disagreements with GS and RD. CR was consulted for any uncertainties. Once group consensus was reached, JB, RD, GS, and CR screened 3,045 unique references, of which 3,022 were excluded. Among 23 references with full text, 11 were excluded because they did not present results about health consumers (*n* = 1), did not use GenAI chatbots (*n* = 5), focused on the technical evaluation of GenAI chatbots (*n* = 3), or only included intention-based findings (n = 2). Overall, 12 papers representing 13 studies are included in this review.[18–29]

<See Fig. 1 here>

Step 5: Collating, summarizing, and reporting the results

An initial review of the included studies revealed diverse research designs. This necessitates using the Mixed Methods Appraisal Tool (MMAT) version 2018[30] which has been used in reviews with multi-methods studies on consumers' health information interaction.[1,31,32] MMAT quality scores range from 0 to 5 (0-2 = poor; 3 = fair; 4 = good; 5 = excellent). GS and RD independently reviewed each study's quality based on MMAT. Interrater reliability is moderate (Krippendorff's α = .63 and .82). Disagreements were discussed among team members. Supplementary Appendix 2 shows the results of the quality appraisal. Eight studies (62%) were rated as excellent. Similar to previous reviews,[1,33,34] no studies were excluded based on quality.

After quality appraisal, we developed a data extraction form using Microsoft Excel. The fields were initially based on the Joanna Briggs Institute's data extraction guidelines.[35] However, the research team added fields that allow an in-depth discussion of the findings (see supplementary Appendix 3 for complete fields). All authors were assigned papers to complete the extraction form. Close-ended fields (e.g., publication year and sample size) were analyzed using content analysis in Microsoft Excel. In contrast, open-ended fields (e.g., aims and key findings) were analyzed using thematic analysis in MAXQDA 24.

RESULTS

Characteristics of the included studies (RQ1)

Table 1 shows a summary of the study characteristics. Given the novelty of GenAI chatbots, studies were recently published between 2023 (*n* = 5; 45%)[24,25,28,29] and 2024 (*n* = 8; 62%).[18–23,26,27] Although most (*n* = 11; 85%) of the studies were carried out since the public release of ChatGPT-3.5 on November 30, 2022,[18–23,25–29] Karinshak et al.[24] performed the first research work (*n* = 2; 13%) that used a GenAI chatbot (GPT-3) to extract health information (i.e., vaccine information) that was subsequently used to gather perceptions from health consumers. In general, studies were mostly conducted in high-income countries[36] (Australia,[26] Kuwait,[21] Germany,[28] Saudi Arabia,[18–20] South Korea,[29] UAE,[21] and US[22–24]), with the US having four (31%) studies reflected in three papers.[22–24]

<Insert Table 1 here>

Most (*n* = 11; 85%) of the studies focused on specific health topics, such as cancer,[20] chronic disease,[18] medical imaging,[23,28] psychological health,[19,27] vaccination,[24] surgical procedures,[26,29] and urolithiasis.[25] Studies were primarily quantitative (*n* = 5; 38%)[23–25,28] and conducted surveys (*n* = 9; 69%)[22–29] for data collection. Except for Choudhury et al.'s study[22] that used Unified Theory of Acceptance and Use of Technology (UTAUT),[37] the rest did not use a theory.[18–21,23–29]

Studies collected data from patients (*n* = 6; 46%),[18–20,25,26,28] the general adult population (*n* = 5; 47%),[21,22,24,29], informal caregivers (*n* = 1; 7%),[27] or patient advocates (*n* = 1; 7%).[23] Patients were recruited from hospitals (*n* = 6; 40%),[18–20,25,26,28] while online methods, such as survey panels (i.e., Amazon Mechanical Turk and Centiment; *n* = 3; 33%)[22,24] and social media

(Facebook, Instagram, X, Telegram; *n* = 2; 13%),[21,27] were used to reach the general adult population or informal caregivers. The analytic sample size ranged between 2 and 1,496, with a median of 24 (*SD* = 443.63). All studies referenced OpenAI's GenAI chatbots, with the majority referencing ChatGPT-3.5 (*n* = 6; 47%),[18–20,23,25,28] followed by GPT-3 (*n* = 2; 13%),[24] and ChatGPT-4 (*n* = 1; 7%).[29]

The studies can be categorized into three groups based on their study aims. The first category (user experience studies) involves investigating health consumers' experience using GenAI chatbots for health information seeking (*n* = 4; 31%).[18–21] For instance, these studies recruited participants who had used ChatGPT for health information seeking and asked for their experience with its use. The second category (consumer surveys) involves identifying health consumers' use and perceptions of GenAI chatbots for health information seeking through consumer surveys (*n* = 1; 8%).[22]

The third category (evaluation studies) involves examining health consumers' evaluation of health information from GenAI chatbots (*n* = 8; 62%).[23–29] These studies have two phases in which researchers use GenAI chatbots to generate health information (i.e., generation phase), which is then followed by an evaluation phase in which researchers ask both health consumers and professionals[23,26,27,29] (*n* = 4) or the former only (*n* = 4)[24,25,28] to evaluate GenAI-generated health information. In the generation phase, most studies generated prompts that were self-developed by the research team (*n* = 6),[24–26,28,29] followed by those generated through literature reviews (*n* = 2)[23,27] and consultation with independent experts (*n* = 1).[23] Next, six studies used zero-shot prompting,[23,25–29] while two of Karinshak et al.'s[24] studies were based on zero-shot and few-shot prompting. Moreover, only two studies specified that one member of the research team entered the prompts.[27,29] In the evaluation phase, two studies employed blinding (the source of health information was not disclosed),[24,26] three did not,[25,27,29] one did both,[24] and two were unclear.[23,28]

Health consumers' use of health information from GenAI chatbots (RQ2)

Five studies[18–22] examined health consumers' use of health information from GenAI chatbots. These include interview studies in West Asia[18–21] and a survey study in the US.[22]

Among the West Asia studies, three studies by Al-Anezi in Saudi Arabia required university hospital patients (29 chronic disease patients,[18] 24 mental health patients,[19] and 72 cancer patients[20]) to use ChatGPT-3.5 to search for health information within two weeks before conducting interviews. These studies are some of the earliest that involved health consumers using a GenAI chatbot for health information seeking. Meanwhile, Al-Shboul[21] interviewed participants from Jordan, Kuwait, and UAE who used ChatGPT to seek health information between 2022 and 2023.

Collectively, ChatGPT was primarily used by health consumers as an information hub to obtain referrals for health services and resources, address health concerns and misconceptions, and learn more about health issues.[18–21] Other uses involve intervention delivery (psychoeducation, cognitive behavioral therapy, and crisis intervention),[18–20] emotional support,[18–21] goal setting,[18–21] and language translation.[20]

Another involved a US consumer survey based on a panel survey of 607 US adults recruited from Centiment.[22] Results show that only 44 (7%) reported using ChatGPT for health information seeking.

Health consumers' perceptions of health information from GenAI chatbots (RQ3)

Table 2 presents a summary of perceptions related to health information from GenAI chatbots.

<Insert Table 2 here>

Accuracy, reliability, or quality (10 studies)

Five studies noted that health consumers are concerned about the accuracy, reliability, or quality of health information provided by GenAI chatbots.[18–21,29] Some studies allude to the socio-technical nature of ChatGPT, wherein participants recognize that it does not have the latest information due to outdated training data[18] (technical dimension) and the reliability of the output depends on the user's prompting skills (social dimension).[29] Moreover, qualitative insights suggest that consumers expect both quality and quantity, in which ChatGPT should provide comprehensive yet reliable health information.[22]

Three studies highlight differences in accuracy, reliability, or quality perceptions among health consumers and professionals.[26,27,29] For instance, a study by Lockie and Choi[26] that blinded the source of the laparoscopic cholecystectomy information leaflets found that patients (compared to doctors) gave a higher quality rating to the ChatGPT version ($M_{patients}$ = 7.5; $M_{doctors}$ = 6.7). Likewise, patients gave the ChatGPT version a higher quality rating than the leaflet created by surgeons ($M_{ChatGPT}$ = 7.5; $M_{Surgeon}$ = 7.1). These findings were consistent with two unblinded studies.[27,29] Specifically, Saeidnia et al.[27] reported that informal caregivers, on average, rated the health information from ChatGPT at a higher level of responsiveness (i.e., "Were the responses scientific enough?") than formal caregivers (i.e., neurologists and nurses) across 31 dementia-related information needs ($M_{informal\ caregivers}$ = 3.77; Informal $_{caregivers}$ = 3.13). Moreover, Yun et al.[29] used DISCERN (a validated instrument for evaluating written consumer health information[38]) and found that laypeople (compared to plastic surgeons) gave higher reliability ($M_{laypeople}$ = 3.61; $M_{plastic\ surgeon}$ = 3.47; p = .014) and information quality ($M_{laypeople}$ = 3.81; $M_{plastic\ surgeon}$ = 3.40; p < .001) scores to ChatGPT-generated mammoplasty information.

Two studies by Karinshak et al.[24] demonstrate how source and source labels affect perceptions of accuracy, reliability, or quality of health information provided by GenAI chatbots. In both studies wherein respondents were blinded from the actual source of information, the GPT-3-generated COVID-19 vaccine information received a significantly higher argument strength than the one from the Centers for Disease Control and Prevention (CDC). However, in the second study that used the same GPT-3-generated COVID-19 vaccine information but was experimentally labeled as either originating from the CDC, doctors, or AI, argument strength is lower for the AI group (M = 3.60) than the CDC (M = 3.81) or doctors group (M = 3.79).

Readability (7 studies)

Qualitative results involving patients[26] and informal caregivers[27] indicated that the health information provided by ChatGPT was well presented, used plain and simple language, and was easy to read. These findings are consistent with quantitative studies that simplified health information using ChatGPT.[25,28] Moreover, Yun et al.[29] found that laypeople (compared to plastic surgeons) gave a slightly higher understandability ($M_{laypeople}$ = 93.42; $M_{plastic\ surgeon}$ = 90.50; p = .051) score to ChatGPT-generated mammoplasty information. Conversely, qualitative studies involving chronic disease[18] and cancer[20] patients from Saudi Arabia found that ChatGPT was not effective in translating health information from English to Arabic.

Trust or trustworthiness (5 studies)

Qualitative findings from several studies highlight health consumers' concerns about the trustworthiness of ChatGPT as a source of health information.[18,21,22] A common theme is that the lack of trust stems from patients' perceived inaccuracy[18,21,22] and bias[18] of health information from ChatGPT. Some studies also found that the trustworthiness of health information from GenAI chatbots is context-dependent. For instance, consumers may distrust health information from ChatGPT if it is about a serious medical issue,[21] but may trust it if the answer is unknown (e.g., patient does not know anything about the health issue).[27] Findings from qualitative studies align with their quantitative counterparts, wherein health consumers are less likely to trust urolithiasis information (unblinded; no comparison group) from ChatGPT-3.5[25] and vaccine information (blinded; $M_{GPT\text{-}3}$ = 3.10; M_{CDC} = 3.77; M_{doctor} = 3.98; p < .001) from GPT-3.[24] One study found that ChatGPT can enhance its trustworthiness by reminding users to consult healthcare professionals for more information about their condition.[18] Another study also suggests that an overly intelligent ChatGPT would make consumers apprehensive of delegating health-related decision-making.[22]

Privacy, confidentiality, security, or safety (5 studies)

Qualitative findings from five studies emphasized health consumers' concerns about privacy, confidentiality, security, or safety of health information from ChatGPT.[18,19,21,22,29] Studies[18,19,21] found that health consumers believe ChatGPT might be misusing others' protected health information (PHI) to generate a response. In effect, they feel unsafe entering their PHI for health information seeking.[18,19,21] Although health consumers who use ChatGPT for health or non-health purposes expressed concerns about the privacy and confidentiality of their information, those who use it for health-related inquiries tend to emphasize the need to secure the safety of health information.[22]

Usefulness (4 studies)

Health consumers consider health information from ChatGPT useful as it can address their health information needs.[27] However, the extent of usefulness is context-dependent based on the difficulty of the question, user type, and extent of personalization. First, Al-Shboul et al.[21] reported that most participants expressed that ChatGPT is useful only for basic health questions and considers usefulness as a motivator to interact with ChatGPT. This is consistent with the study of Gordon et al.[23] wherein patient advocates rated most of the ChatGPT responses to radiology report questions as "at least partially relevant and of or helpful" (97%; n = 128/132) rather than "fully relevant and of or helpful" (57%; n = 75/132).

Second, although laypeople and plastic surgeons found health information about mammoplasty to be useful (usefulness was conceptualized as actionability based on the Patient Education Materials Assessment Tool [PEMAT][39]), the former had a significantly lower rating for its usefulness ($M_{laypeople}$ = 86.56; $M_{plastic\ surgeon}$ = 93.44; p = .013).[29] That study also found that the lack of visual aids limits the usefulness of text-only health information provided by ChatGPT.[29]

Finally, studies show that health information from ChatGPT was less useful because it lacked personalization.[20,29] This is supported by one study wherein most participants (n = 12; 75%) expressed that ChatGPT should provide personalized responses to be useful.[21]

Accessibility (4 studies)

Qualitative findings show that health consumers appreciate the capability to access health information from ChatGPT regardless of time or location.[18,21,29] Moreover, it enhances access to health information by being free to use[18,27] and available on multiple devices.[18] However, there is concern about how long it will remain free.[18] One study also found that most health consumers (*n* = 13; 81%) consider accessibility as a motivator to use ChatGPT for health information seeking.[21]

Emotional appeal (4 studies)

Two studies found that health information provided by ChatGPT-3.5 (free version) lacks empathy.[18,21] On the contrary, a study that used ChatGPT-4 (paid version at the time of the study) to generate mammoplasty information found that laypeople thought that it provides "emotionally appropriate counseling" and "is actually better than some doctors."[29] The same study also found that laypeople gave the ChatGPT-4-generated information a higher emotional score than plastic surgeons ($M_{laypeople}$ = 3.49; $M_{plastic\ surgeon}$ = 3.05; p = .002).[29] Given the importance of emotional appeal, one study found that 63% (*n* = 10) of health consumers consider it a motivator to use ChatGPT for health information seeking.[21]

Attitude (3 studies)

Kim et al.[25] found that patients had more negative attitudes (i.e., worry and wariness) after reading urolithiasis prevention information from ChatGPT (unblinded). This finding is consistent with Karinshak et al.'s[24] second study in which unblinded respondents had a lower attitude to COVID-19 vaccine information from GPT-3 ($M_{GPT\text{-}3}$ = 2.38) than the one from the CDC (M_{CDC} = 2.76). On the contrary, Karinshak et al.'s[24] blinded groups across two studies reported higher attitudes towards vaccine information from GPT-3 than those from the CDC.

Effectiveness (2 studies)

Two studies by Karinshak et al.[24] examined perceived message effectiveness (reflecting persuasiveness and believability) of COVID-19 vaccine information from GPT-3. In both studies wherein respondents were blinded from the actual source of information, the GPT-3-generated COVID-19 vaccine information had significantly higher perceived message effectiveness than the one from the CDC. However, in the second study that used the same GPT-3-generated COVID-19 vaccine information but was labeled as either originating from the CDC, doctors, or AI, perceived message effectiveness was significantly lower for the AI group (M = 3.28) than the CDC (M = 3.50) or doctors group (M = 3.47).

DISCUSSION

Our findings show that studies on health consumers' use and perceptions of health information from GenAI chatbots are in the early stages, as evidenced by few publications concentrated in high-income countries and the prevalence of atheoretical studies. This finding is consistent with reviews of other emerging health information technologies.[40,41] Thus, we encourage using theory to better understand and offer potential explanations of how exposure to health information from GenAI chatbots leads to health beliefs and outcomes. Broad categories of theoretical models that may offer helpful insights include behavior change models (e.g., AI Chatbot Behavior Change Model,[42] Behavior Change Wheel,[43] Health Belief Model,[44] and Theory of Planned Behavior[45]), technology acceptance models (e.g., UTAUT[37]) and implementation science models (e.g., Consolidated Framework for Implementation Research Framework[46]).

Although more than half of the studies (62%) were of excellent quality, none used standardized reporting guidelines. This is expected since GenAI-related reporting guidelines (FUTURE-AI[47] and TRIPOD-LLM[48]) were not yet available when the reviewed studies were conducted. As more scholars become aware of these guidelines, we expect greater adoption of such guidelines. Moreover, future work should provide more details of their methodology (e.g., prompt generation process, prompting technique, and number of assigned prompters) to enhance rigor and reproducibility.

Most studies used OpenAI's ChatGPT. Although Open AI's GPT-3 was the first[24] to be referenced among the reviewed studies, its limited release among developers[49] may explain why it was not used in as many studies as ChatGPT. As ChatGPT was more recently released, very few users have used it,[50] making it a novel health information source. This is evidenced by a few studies that required participants to use ChatGPT for health information seeking[18–20] and a low percentage (7%) of self-reported use for health information seeking[22] As the number of GenAI chatbots continues to grow and as health consumers increasingly use them, we expect to see studies that report greater use of GenAI chatbots for health information seeking and compare the use and perceptions of health information between GenAI chatbots.

The findings shed light on perceptions that various stakeholders (e.g., end users, healthcare providers, GenAI developers, policymakers, and scholars) should be mindful of when incorporating GenAI chatbots to support health information seeking. For instance, despite health consumers perceiving health information from GenAI chatbots to be accessible, readable, and useful, the studies we reviewed also found that health consumers have negative attitudes and distrust towards GenAI chatbots, leading them to be critical of their accuracy, safety, and effectiveness, especially when health information is explicitly mentioned to originate from them.

This is consistent with earlier consumer surveys[51] and research on user perceptions of AI-generated output in healthcare[52,53] and non-healthcare contexts.[54] Given the rapid technological development of GenAI chatbots[55] and as more health consumers become familiar with them,[56] there will be a need to longitudinally examine perceptions of health information from such tools to mitigate health-related risks and improve health outcomes. Besides, examining cultural and socioeconomic differences[57,58] could identify patterns in the use and perceptions of GenAI chatbots for health information seeking.

Limitations and future perspectives

This review has several limitations. Although we conducted a rigorous and systematic search through database and manual searches, this scoping review only represents a few studies. Given the strong scientific interest in GenAI as evidenced by an ever-increasing number of newly published papers,[59] we have missed studies that were not indexed during the search. Besides, grey literature was not included in the search. As such, the insights from this review only reflect the findings from the included studies, which limits generalizability. Since only papers published in English were considered for inclusion, otherwise qualifying non-English publications may have been missed. Despite assessing the quality of studies, this only provides the current state of study quality on this topic. It does not give a critical evaluation necessary to facilitate the development of evidence-based practices. Finally, given the availability of multiple GenAI chatbots (e.g., Claude, Copilot, DeepSeek, Gemini, Grok, Meta AI, and Perplexity) that routinely incorporate enhanced information retrieval technologies (e.g., embedded real-time web-search functionality, retrieval augmented generations, and response reasoning[60]) to help reduce hallucinations and provide dynamic, up-to-date information, it is crucial to identify how such changes affect health information seeking. Thus, future studies can use our findings as a baseline to identify changes in health consumers' use and perceptions of health information from a wide range of GenAI chatbots.

CONCLUSIONS

This scoping review provides an initial overview of health consumers' use and perceptions of health information from GenAI chatbots. Although health consumers can use GenAI chatbots to obtain accessible, readable, and useful health information, negative perceptions of their accuracy, trustworthiness, effectiveness, and safety serve as barriers that stakeholders must address to mitigate health-related risks, improve health beliefs, and achieve positive health outcomes among users of GenAI chatbots. Aside from advocating the use of theories to explain how health information provided by GenAI chatbots leads to health beliefs and outcomes, this review calls for methodological rigor by using standardized reporting guidelines that facilitate reproducibility and comparison of future work.

Clinical Relevance Statement

This scoping review identified the research landscape on health consumers' use and perceived health information from GenAI chatbots. Our findings show that health consumers distrust AI, making them critical of its accuracy, safety, and effectiveness. Healthcare providers must familiarize themselves with GenAI chatbots and work with health consumers on responsibly using them for health information seeking.

Multiple-Choice Questions

1. Who among the following authors conducted the earliest study to examine health consumers' perceptions of health information from GenAI chatbot?

    a. Karinshak et al. (2023)

    b. Kim et al. (2023)

    c. Schmidt et al. (2023)

    d. Yun et al. (2023)

    Correct Answer: The correct answer is option a. Karinshak et al.[24] used GPT-3 to obtain COVID-19 vaccine information. GPT-3 was released in 2019 and the predecessor of ChatGPT (released November 2022). The rest of the choices conducted their study using different versions of ChatGPT.

2. Most studies on health consumers' perceptions of health information from GenAI chatbot provided insights into…

    a. Usefulness

    b. Trust or trustworthiness

    c. Readability

    d. Accuracy, reliability, or quality

    Correct Answer: The correct answer is option d. Most of the studies (77%; n = 10) provided insights into the accuracy, reliability, or quality of health information from GenAI chatbots.

Author's Contributions

JB conceptualized, managed, and supervised the project. JB developed the search terms. JB, DH, and MF performed manual search. JB, GS, RD, and CR reviewed search results and screened the references. All authors extracted the data. JB, DH, and MF analyzed the extracted data. JB and DH drafted the manuscript. All authors revised and approved the final version of the manuscript.

Protection of Human and Animal Subjects

Human and/or animal subjects were not included in the project.

Data Availability Statement

The data underlying this article including the detailed search strategy are available in the article and its online supplementary appendix information.

Funding

This work was supported by a start-up grant from MU Sinclair School of Nursing. The funder had no role in the study design, data collection, data analysis, or manuscript preparation.

Conflict of interest statement

The authors have no conflicts of interest to report related to this work.

Acknowledgments

We acknowledge Rebecca Graves of the University of Missouri J. Otto Lottes Health Sciences Library for her assistance in the database search.

References

1. Bautista JR, Zhang Y, Gwizdka J, Chang YS. Consumers' longitudinal health information needs and seeking: a scoping review. *Health Promot Int*. 2023;38(4):daad066. doi:10.1093/heapro/daad066
2. Marr B. A Short History Of ChatGPT: How We Got To Where We Are Today. Forbes. 2023. Accessed January 29, 2025. https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/
3. Holohan M. A boy saw 17 doctors over 3 years for chronic pain. ChatGPT found the diagnosis. TODAY.com. September 12, 2023. Accessed January 29, 2025. https://www.today.com/health/mom-chatgpt-diagnosis-pain-rcna101843
4. Ayo-Ajibola O, Davis RJ, Lin ME, Riddell J, Kravitz RL. Characterizing the Adoption and Experiences of Users of Artificial Intelligence–Generated Health Information in the United States: Cross-Sectional Questionnaire Study. *J Med Internet Res*. 2024;26(1):e55138. doi:10.2196/55138
5. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). *Npj Digit Med*. 2024;7(1):1-11. doi:10.1038/s41746-024-01157-x
6. Wei Q, Yao Z, Cui Y, Wei B, Jin Z, Xu X. Evaluation of ChatGPT-generated medical responses: A systematic review and meta-analysis. *J Biomed Inform*. 2024;151:104620. doi:10.1016/j.jbi.2024.104620

7. Yim D, Khuntia J, Parameswaran V, Meyers A. Preliminary Evidence of the Use of Generative AI in Health Care Clinical Services: Systematic Narrative Review. *JMIR Med Inform*. 2024;12(1):e52073. doi:10.2196/52073
8. Ferraris G, Monzani D, Coppini V, et al. Barriers to and facilitators of online health information-seeking behaviours among cancer patients: A systematic review. *Digit Health*. 2023;9:20552076231210663. doi:10.1177/20552076231210663
9. Zhang Y, Kim Y. Consumers' Evaluation of Web-Based Health Information Quality: Meta-analysis. *J Med Internet Res*. 2022;24(4):e36463. doi:10.2196/36463
10. Wang C, Qi H. Influencing Factors of Acceptance and Use Behavior of Mobile Health Application Users: Systematic Review. *Healthcare*. 2021;9(3):357. doi:10.3390/healthcare9030357
11. Freeman JL, Caldwell PHY, Scott KM. How Adolescents Trust Health Information on Social Media: A Systematic Review. *Acad Pediatr.* 2023;23(4):703-719. doi:10.1016/j.acap.2022.12.011
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19-32. doi:10.1080/1364557032000119616
13. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol*. 2018;18:5. doi:10.1186/s12874-017-0468-4
14. National Institutes of Health. Health consumer. Toolkit. Accessed January 30, 2025. https://toolkit.ncats.nih.gov/glossary/health-consumer
15. Solaiman I, Brundage M, Clark J, et al. Release Strategies and the Social Impacts of Language Models. Published online November 13, 2019. doi:10.48550/arXiv.1908.09203

16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int J Surg*. 2010;8(5):336-341. doi:10.1016/j.ijsu.2010.02.007

17. Graafsma J, Murphy RM, van de Garde EMW, et al. The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: a scoping review. *J Am Med Inform Assoc*. 2024;31(6):1411-1422. doi:10.1093/jamia/ocae076

18. Al-Anezi FM. Exploring the use of ChatGPT as a virtual health coach for chronic disease management. *Learn Health Syst*. Published online 2024. doi:10.1002/lrh2.10406

19. Al-Anezi F. Assessing the Effectiveness of ChatGPT in Delivering Mental Health Support: A Qualitative Study. *J Multidiscip Healthc*. 2024;17:461-471. doi:10.2147/JMDH.S447368

20. Al-Anezi F. Examining the role of ChatGPT in promoting health behaviors and lifestyle changes among cancer patients. *Nutr Health*. Published online 2024:02601060241244563. doi:10.1177/02601060241244563

21. Al Shboul MKI, Alwreikat A, Alotaibi FA. Investigating the Use of ChatGpt as a Novel Method for Seeking Health Information: A Qualitative Approach. *Sci Technol Libr.* 2024;43(3):225-234. doi:10.1080/0194262X.2023.2250835

22. Choudhury A, Elkefi S, Tounsi A. Exploring factors influencing user perspective of ChatGPT as a technology that assists in healthcare decision making: A cross sectional survey study. *PLOS ONE*. 2024;19(3):e0296151. doi:10.1371/journal.pone.0296151

23. Gordon EB, Towbin AJ, Wingrove P, et al. Enhancing Patient Communication With Chat-GPT in Radiology: Evaluating the Efficacy and Readability of Answers to Common Imaging-Related Questions. *J Am Coll Radiol*. 2024;21(2):353-359. doi:10.1016/j.jacr.2023.09.011

24. Karinshak E, Liu SX, Park JS, Hancock JT. Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. *Proc ACM Hum-Comput Interact*. 2023;7(CSCW1):1-29. doi:10.1145/3579592

25. Kim SH, Tae JH, Chang IH, et al. Changes in patient perceptions regarding ChatGPT-written explanations on lifestyle modifications for preventing urolithiasis recurrence. *Digit Health*. 2023;9:20552076231203940. doi:10.1177/20552076231203940

26. Lockie E, Choi J. Evaluation of a chat GPT generated patient information leaflet about laparoscopic cholecystectomy. *ANZ J Surg*. 2024;94(3):353-355. doi:10.1111/ans.18834

27. Saeidnia HR, Kozak M, Lund BD, Hassanzadeh M. Evaluation of ChatGPT's responses to information needs and information seeking of dementia patients. *Sci Rep*. 2024;14(1):10273. doi:10.1038/s41598-024-61068-5

28. Schmidt S, Zimmerer A, Cucos T, Feucht M, Navas L. Simplifying radiologic reports with natural language processing: a novel approach using ChatGPT in enhancing patient understanding of MRI results. *Arch Orthop TRAUMA Surg*. Published online 2023. doi:10.1007/s00402-023-05113-4

29. Yun JY, Kim DJ, Lee N, Kim EK. A comprehensive evaluation of ChatGPT consultation quality for augmentation mammoplasty: A comparative analysis between plastic surgeons and laypersons. *Int J Med Inf*. 2023;179. doi:10.1016/j.ijmedinf.2023.105219

30. Hong QN, Pluye P, Fàbregues S, et al. Mixed methods appraisal tool (MMAT), version 2018. *Regist Copyr.* 2018;1148552.

31. Akhlaq A, McKinstry B, Muhammad KB, Sheikh A. Barriers and facilitators to health information exchange in low-and middle-income country settings: a systematic review. *Health Policy Plan*. 2016;31(9):1310-1325.

32. Hurley D, Swann C, Allen MS, Ferguson HL, Vella SA. A systematic review of parent and caregiver mental health literacy. *Community Ment Health J*. 2020;56(1):2-21.

33. Derksen ME, van Strijp S, Kunst AE, Daams JG, Jaspers MWM, Fransen MP. Serious games for smoking prevention and cessation: A systematic review of game elements and game effects. *J Am Med Inform Assoc*. 2020;27(5):818-833.

34. Moore EC, Tolley CL, Bates DW, Slight SP. A systematic review of the impact of health information technology on nurses' time. *J Am Med Inform Assoc*. 2020;27(5):798-807.

35. Joanna Briggs Institute. 10.2.7 Data extraction - JBI Manual for Evidence Synthesis - JBI Global Wiki. Accessed December 20, 2024. https://jbi-global-wiki.refined.site/space/MANUAL/355862769/10.2.7+Data+extraction

36. World Bank. World Bank Country and Lending Groups. 2025. Accessed January 6, 2025. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

37. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. *MIS Q*. 2003;27(3):425-478. doi:10.2307/30036540

38. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health*. 1999;53(2):105-111. Accessed January 20, 2025. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756830/

39. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ Couns*. 2014;96(3):395-403. doi:10.1016/j.pec.2014.05.027

40. Baretta D, Bondaronek P, Direito A, Steca P. Implementation of the goal-setting components in popular physical activity apps: Review and content analysis. *Digit Health*. 2019;5:2055207619862706. doi:10.1177/2055207619862706

41. Maddison R, Rawstorn JC, Shariful Islam SM, et al. mHealth Interventions for Exercise and Risk Factor Modification in Cardiovascular Disease. *Exerc Sport Sci Rev*. 2019;47(2):86. doi:10.1249/JES.0000000000000185

42. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. *J Med Internet Res*. 2020;22(9):e22845. doi:10.2196/22845

43. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implement Sci*. 2011;6(1):42. doi:10.1186/1748-5908-6-42

44. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. *Health Educ Q*. 1988;15(2):175-183. doi:10.1177/109019818801500203

45. Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process*. 1991;50(2):179-211. doi:10.1016/0749-5978(91)90020-T

46. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci*. 2009;4(1):50. doi:10.1186/1748-5908-4-50

47. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ*. 2025;388:e081554. doi:10.1136/bmj-2024-081554

48. Gallifant J, Afshar M, Ameen S, et al. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat Med*. 2025;31(1):60-69. doi:10.1038/s41591-024-03425-5

49. Weiss TR. OpenAI GPT-3 Waiting List Dropped as GPT-3 Is Fully Released for Developer and Enterprise Use. AIwire. November 18, 2021. Accessed February 11, 2025. https://www.aiwire.net/2021/11/18/openai-gtp-3-waiting-list-is-gone-as-gtp-3-is-fully-released-for-use/

50. McClain C. Americans' use of ChatGPT is ticking up, but few trust its election information. Pew Research Center. March 26, 2024. Accessed February 18, 2025. https://www.pewresearch.org/short-reads/2024/03/26/americans-use-of-chatgpt-is-ticking-up-but-few-trust-its-election-information/

51. Funk AT Giancarlo Pasquini, Alison Spencer and Cary. 60% of Americans Would Be Uncomfortable With Provider Relying on AI in Their Own Health Care. Pew Research Center. February 22, 2023. Accessed February 18, 2025. https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/

52. Lee MK, Rich K. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Association for Computing Machinery; 2021:1-14. doi:10.1145/3411764.3445570

53. Spotnitz M, Idnay B, Gordon ER, et al. A Survey of Clinicians' Views of the Utility of Large Language Models. *Appl Clin Inform*. 2024;15:306-312. doi:10.1055/a-2281-7092

54. Lee G, Kim HY. Human vs. AI: The battle for authenticity in fashion design and consumer response. *J Retail Consum Serv*. 2024;77:103690. doi:10.1016/j.jretconser.2023.103690

55. Salmi L, Lewis DM, Clarke JL, et al. A proof-of-concept study for patient use of open notes with large language models. *JAMIA Open*. 2025;8(2):ooaf021. doi:10.1093/jamiaopen/ooaf021

56. Presiado M, Montero A, Lopes L, Published LH. KFF Health Misinformation Tracking Poll: Artificial Intelligence and Health Information. KFF. August 15, 2024. Accessed May 26, 2025. https://www.kff.org/health-information-and-trust/poll-finding/kff-health-misinformation-tracking-poll-artificial-intelligence-and-health-information/

57. Albashayreh A, Zeinali N, Gusen NJ, Ji Y, Gilbertson-White S. An Informatics Approach to Characterizing Rarely Documented Clinical Information in Electronic Health Records: Spiritual Care as an Exemplar. *Appl Clin Inform*. Published online May 5, 2025. doi:10.1055/a-2599-6300

58. Langevin R, Berry ABL, Zhang J, et al. Implementation Fidelity of Chatbot Screening for Social Needs: Acceptability, Feasibility, Appropriateness. *Appl Clin Inform*. 2023;14:374-391. doi:10.1055/a-2035-5342

59. World Intellectual Property Organization. Patent Landscape Report - Generative Artificial Intelligence (GenAI) - Key findings and insights. Published online 2024. Accessed February 11, 2025. https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/en/key-findings-and-insights.html

60. Menz BD, Modi ND, Abuhelwa AY, et al. Generative AI chatbots for reliable cancer information: Evaluating web-search, multilingual, and reference capabilities of emerging large language models. *Eur J Cancer*. 2025;218:115274. doi:10.1016/j.ejca.2025.115274

Table 1 Study characteristics

Characteristics	*n* of studies (%)
Year published	
2024 (up to July)	8 (62%)
2023	5 (38%)
Study period	
After ChatGPT-3.5 release (since November 30, 2022)	11 (85%)
Before ChatGPT-3.5 release (before November 30, 2022)	2 (15%)
Country conducted	
United States (US)	4 (31%)
Saudi Arabia	3 (23%)
South Korea	2 (15%)
Australia	1 (8%)
Germany	1 (8%)
Iran	1 (8%)
Jordan	1 (8%)
Kuwait	1 (8%)
United Arab Emirates (UAE)	1 (8%)
Health topic	
General health	2 (15%)
Medical imaging	2 (15%)
Psychological health	2 (15%)
Vaccination	2 (15%)
Surgical procedures	2 (15%)
Cancer	1 (8%)
Chronic Disease	1 (8%)
Urolithiasis	1 (8%)
Use of theory	
No	14 (92%)
Yes	1 (8%)
Design	
Quantitative	5 (38%)
Qualitative	4 (31%)
Mixed	4 (31%)
Data collection method	
Survey	9 (69%)
Interview	4 (31%)
Focus group	1 (8%)
Health consumer category	
Patients	6 (46%)
General adult population	5 (38%)
Informal caregivers	1 (8%)
Patient advocates	1 (8%)
Recruitment site	
Hospital	6 (46%)

Survey panel	3 (23%)
Social media	2 (15%)
Unspecified	2 (15%)
Analytic sample size	
Less than 10	2 (15%)
10-99	8 (62%)
More than 100	3 (23%)
GenAI chatbot	
ChatGPT-3.5	6 (46%)
ChatGPT (unspecified version)	4 (31%)
GPT-3	2 (15%)
ChatGPT-4	1 (8%)
Study type	
Evaluation studies	8 (62%)
User experience studies	4 (31%)
Consumer surveys	1 (8%)

Note: Results can exceed 100% due to overlap or rounding.

Table 2 Summary of studies depicting health consumers' perceptions of health information from GenAI chatbots

Study	Accuracy, reliability, or quality (*n* = 10)	Readability (*n* = 7)	Trust or trustworthiness (*n* = 5)	Privacy, confidentiality, security, or safety (*n* = 5)	Usefulness (*n* = 4)	Accessibility (*n* = 4)	Emotional appeal (*n* = 4)	Attitude (*n* = 3)	Effectiveness (*n* = 2)
1. Al-Anezi (2024)[18]	X	X	X	X		X	X		
2. Al-Anezi (2024)[19]	X			X					
3. Al-Anezi (2024)[20]	X	X			X				
4. Al-Shboul et al. (2023)[21]	X		X	X	X	X	X		

5. Choudhury et al. (2024)[22]	X		X	X					
6. Gordon et al. (2024)[23]									
7. Karinshak et al. (2023) – study 1[24]	X							X	X
8. Karinshak et al. (2023) – study 2[24]	X							X	X
9. Kim et al. (2023)[25]		X	X					X	
10. Lockie and Choi (2024)[26]	X	X							
11. Saeidnia et al. (2024)[27]	X	X			X	X			
12. Schmidt et al. (2023)[28]		X							
13. Yun et al. (2023)[29]	X	X		X	X	X	X		

Fig. 1. PRISMA diagram

Supplementary Appendix 1. Search terms used in database search

Database: ACM Digital Library
Search date: 2/28/2024
Results: 924
"query": { AllField:("Artificial intelligence" OR "Generative AI" OR "generative intelligence" OR ChatGPT OR Bard OR Copilot OR "bing chat" OR "bing ai" OR gemini) AND ("Healthcare information" OR "Health information" OR "medical information" OR "Drug information" OR "information seeking" OR "seeking behavior" OR "seeking behaviour" OR "seeking behaviors" OR "Seeking Behaviours" OR "information behavior" OR "information Behaviour" OR "information behaviors" OR "information Behaviours") AND (patients OR patient) }
"filter": { E-Publication Date: Past 5 years, ACM Content: DL }

Database: CINAHL
Search date: 2/26/2024
Results: 158
((AI OR Artificial intelligence OR Generative AI OR generative intelligence OR ChatGPT OR Copilot OR bing chat OR bing AI OR (TI Bard OR AB Bard OR SU Bard))) AND (Healthcare information OR Health information OR medical information OR information seeking OR seeking behavior OR Seeking Behaviour OR seeking behaviors OR Seeking Behaviours OR information behavior OR information Behaviour OR information behaviors OR information Behaviours OR (MH "Consumer Health Information") OR (MH "Drug Information")) AND (patient OR patients OR consumers OR consumer OR general public OR layperson OR laypersons OR laypeople)
Limiters - Publication Date: 20190101-20240231

Database: Cochrane DSR
Search date: 2/26/2024
Results: 7
(Artificial intelligence or Generative AI or generative intelligence or ChatGPT or Copilot or bing chat or bing AI or bard or gemini).ti,ab,ct,kw.
No limiters as there were so few results.

Database: Communication & Mass Media Complete
Search date: 2/26/2024
Results: 78
((Healthcare information OR Health information OR medical information OR drug information OR information seeking OR seeking behavior OR seeking Behaviour OR seeking behaviors OR seeking Behaviours OR information behavior OR information Behaviour OR information behaviors OR information Behaviours) OR (patients OR patient)) AND (AI OR Artificial intelligence OR Generative AI OR generative intelligence OR ChatGPT OR Copilot OR bing chat OR bing ai OR Bard OR gemini)
Limiters - Publication Date: 20190101-20241231

Database: Library Literature & Information Science Full Text
Search date: 2/26/2024
Results: 180
(AI OR Artificial intelligence OR Generative AI OR generative intelligence OR ChatGPT OR Copilot OR bing chat OR bing ai OR Bard OR gemini) AND ((Healthcare Information OR Health information OR medical information OR drug information OR ((information seeking OR seeking behavior OR seeking Behaviour OR seeking behaviors OR seeking Behaviours OR information behavior OR information Behaviour OR information behaviors OR information Behaviours) AND (patient OR patients OR consumer OR consumers OR general public OR layperson OR laypersons OR laypeople)))
Limiters - Publication Date: 20190101-20241231

Database: PROSPERO
Search date: 2/26/2024
Results: 49
("Artificial intelligence" OR "Generative AI" OR "generative intelligence" OR ChatGPT OR Copilot OR "bing chat" OR "bing AI" OR "Google Bard" OR "Bard AI" OR "Google Gemini" OR "Gemini AI") AND (Healthcare information OR Health information OR medical information OR information seeking OR seeking behavior OR Seeking Behaviour OR seeking behaviors OR Seeking Behaviours OR information behavior OR information Behaviour OR information behaviors OR information Behaviours OR drug information) WHERE CD FROM 01/01/2019 TO 29/02/2024

Database: PsycINFO & PsycArticles
Search date: 2/26/2024
Results: 88
(("AI" OR Artificial intelligence OR "Generative AI" OR generative intelligence OR ChatGPT OR Copilot OR bing chat OR "bing AI") OR (TI Bard OR AB Bard OR KW Bard OR SU Bard)) AND (patient OR patients OR consumers OR consumer OR general public OR layperson OR laypersons) AND (Healthcare information OR Health information OR medical information OR information seeking OR seeking behavior OR seeking behaviors OR information behavior OR information behaviors)
Limiters - Publication Year: 2019-2024

Database: PubMed
Search date: 2/28/2024
Results: 663
(("AI" OR "Artificial intelligence"[tiab] OR "Artificial Intelligence"[Mesh:NoExp] OR "Generative AI" OR "generative intelligence" OR ChatGPT OR Copilot OR "Bing chat" OR "Bing AI" OR Bard[tiab] OR "gemini") AND (Patients[Mesh:NoExp] OR patients[tiab] OR patient[tiab] OR Consumers[tiab] OR consumer[tiab] OR layperson OR laypersons OR laypeople OR "general public")) AND ("healthcare information" OR "health information" OR "Consumer Health Information"[Mesh:NoExp] OR "medical information" OR "drug information" OR "Information Seeking Behavior"[Mesh] OR "information seeking" OR "seeking behavior" OR "seeking behaviour" OR "seeking behaviors" OR "Seeking Behaviours" OR "information behavior" OR "information Behaviour" OR "information behaviors" OR "information Behaviours" OR "Inf Behav"[Journal:__jid9001390])
Filters: from 2019/1/1 - 2024/2/29

Database: Scopus
Search date: 2/28/2024
Results: 1050
(TITLE-ABS-KEY (ai OR "Artificial intelligence" OR "Generative AI" OR "generative intelligence" OR chatgpt OR bard OR copilot OR "bing chat" OR "bing ai" OR gemini) AND TITLE-ABS-KEY ("Healthcare information" OR "Health information" OR "medical information" OR "Drug information" OR "information seeking" OR "seeking behavior" OR "seeking behaviour" OR "seeking behaviors" OR "Seeking Behaviours" OR "information behavior" OR "information Behaviour" OR "information behaviors" OR "information Behaviours") AND TITLE-ABS-KEY (patients OR patient OR consumer OR consumers OR layperson OR laypersons OR laypeople OR "general public")) AND PUBYEAR > 2018 AND PUBYEAR < 2025

Database: Web of Science
Search date: 2/28/2024
Results: 634
TS=(AI OR "Artificial intelligence" OR "Generative AI" OR "generative intelligence" OR ChatGPT OR Bard OR Copilot OR "bing chat" OR "bing ai" OR gemini) AND ((TS=("Healthcare information" OR "Health information" OR "medical information" OR "drug information" OR information seeking OR seeking behavior OR Seeking Behaviour OR seeking behaviors OR Seeking Behaviours OR information behavior OR information Behaviour OR information behaviors OR information Behaviours) AND TS=(patients OR patient)) OR (TS=("Healthcare information" OR "Health information" OR "medical information" OR "drug information") AND TS=(Consumer OR Consumers OR layperson OR laypersons OR laypeople OR "general public")))
Timespan: 2019-01-01 to 2024-02-28 (Publication Date)

Supplementary Appendix 2. Quality appraisal using MMAT

Author(s) and year	Study design[b]	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5	Evaluation[c]
Al-Anezi (2024)[1a]	Qualitative	Yes	Yes	Yes	Yes	Yes	Excellent
Al-Anezi (2024)[2]	Qualitative	Yes	Yes	Can't tell	Yes	Yes	Good
Al-Anezi (2024)[3a]	Qualitative	Yes	Yes	Yes	Yes	Yes	Excellent
Al-Shboul et al. (2023)[4]	Qualitative	Yes	Yes	Yes	Yes	Yes	Excellent
Choudhury et al. (2024)[5]	Mixed	Yes	Yes	Yes	Yes	Yes	Excellent
Gordon et al. (2024) [6]	Quantitative-NR	Yes	Yes	Yes	Yes	Yes	Excellent
Karinshak et al. (2023)[7] – study 1	Quantitative-NR	Yes	Yes	Yes	Yes	Yes	Excellent
Karinshak et al. (2023)[7] – study 2	Quantitative-NR	Yes	Yes	Yes	Yes	Yes	Excellent
Kim et al. (2023)[8a]	Quantitative-NR	Yes	No	Yes	Can't tell	Yes	Fair
Lockie and Choi (2024)[9a]	Mixed	Yes	No	Yes	Yes	Yes	Good
Saeidnia et al. (2024)[10]	Mixed	Yes	No	Yes	Yes	Yes	Good
Schmidt et al. (2023)[11]	Quantitative-D	Yes	No	Yes	Yes	Yes	Good
Yun et al. (2023)[12]	Mixed	Yes	Yes	Yes	Yes	Yes	Excellent

Abbreviations: Quantitative-D = Quantitative descriptive. Quantitative-NR = Quantitative non-randomized.
[a]Publication year is based on online first release.
[b]Criteria questions depend on study design in MMAT[13]
[c]Evaluation of studies: Excellent (Yes = 5), Good (Yes = 4), Fair (Yes $\leq$ 3)

References

1. Al-Anezi FM. Exploring the use of ChatGPT as a virtual health coach for chronic disease management. *LEARNING HEALTH SYSTEMS*. Published online 2024. doi:10.1002/lrh2.10406
2. Al-Anezi F. Assessing the Effectiveness of ChatGPT in Delivering Mental Health Support: A Qualitative Study. *Journal of Multidisciplinary Healthcare*. 2024;17:461-471. doi:10.2147/JMDH.S447368
3. Al-Anezi F. Examining the role of ChatGPT in promoting health behaviors and lifestyle changes among cancer patients. *Nutrition and Health*. Published online 2024:02601060241244563. doi:10.1177/02601060241244563
4. Al Shboul MKI, Alwreikat A, Alotaibi FA. Investigating the Use of ChatGpt as a Novel Method for Seeking Health Information: A Qualitative Approach. *Science & Technology Libraries*. 2024;43(3):225-234. doi:10.1080/0194262X.2023.2250835
5. Choudhury A, Elkefi S, Tounsi A. Exploring factors influencing user perspective of ChatGPT as a technology that assists in healthcare decision making: A cross sectional survey study. *PLOS ONE*. 2024;19(3):e0296151. doi:10.1371/journal.pone.0296151
6. Gordon EB, Towbin AJ, Wingrove P, et al. Enhancing Patient Communication With Chat-GPT in Radiology: Evaluating the Efficacy and Readability of Answers to Common Imaging-Related Questions. *Journal of the American College of Radiology*. 2024;21(2):353-359. doi:10.1016/j.jacr.2023.09.011
7. Karinshak E, Liu SX, Park JS, Hancock JT. Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. *Proc ACM Hum-Comput Interact*. 2023;7(CSCW1):1-29. doi:10.1145/3579592
8. Kim SH, Tae JH, Chang IH, et al. Changes in patient perceptions regarding ChatGPT-written explanations on lifestyle modifications for preventing urolithiasis recurrence. *DIGITAL HEALTH*. 2023;9:20552076231203940. doi:10.1177/20552076231203940
9. Lockie E, Choi J. Evaluation of a chat GPT generated patient information leaflet about laparoscopic cholecystectomy. *ANZ Journal of Surgery*. 2024;94(3):353-355. doi:10.1111/ans.18834
10. Saeidnia HR, Kozak M, Lund BD, Hassanzadeh M. Evaluation of ChatGPT's responses to information needs and information seeking of dementia patients. *Scientific Reports*. 2024;14(1):10273. doi:10.1038/s41598-024-61068-5
11. Schmidt S, Zimmerer A, Cucos T, Feucht M, Navas L. Simplifying radiologic reports with natural language processing: a novel approach using ChatGPT in enhancing patient understanding of MRI results. *ARCHIVES OF ORTHOPAEDIC AND TRAUMA SURGERY*. Published online 2023. doi:10.1007/s00402-023-05113-4
12. Yun JY, Kim DJ, Lee N, Kim EK. A comprehensive evaluation of ChatGPT consultation quality for augmentation mammoplasty: A comparative analysis between plastic surgeons and laypersons. *International Journal of Medical Informatics*. 2023;179. doi:10.1016/j.ijmedinf.2023.105219
13. Hong QN, Pluye P, Fàbregues S, et al. Mixed methods appraisal tool (MMAT), version 2018. *Registration of copyright*. 2018;1148552.

Supplementary Appendix 3. Data extraction results

Author(s), publication year	**Country conducted**	**Study type; Aims; Theory used**	**Health topic**	**Design; data collection; year of data collection**	**Sample size and type of health consumers; recruitment site**	**GenAI chatbot; exposure; duration of exposure if direct**	**Key findings**
Al-Anezi (2024) [1][a]	Saudi Arabia	User experience study; Examine the use of ChatGPT as a virtual health coach for chronic disease management; None	Chronic disease	Qualitative; semi-structured interview; unspecified year	29 chronic disease patients; university hospital	ChatGPT 3.5; direct; used for 2 weeks	20 themes/factors which were categorized into opportunities and challenges in using ChatGPT as a virtual coach for chronic disease management. Opportunities include (1) Continuous or life-long learning; (2) scalability; (3) Cost-effectiveness; (4) Reminders; (5) Behavioral change support; (6) Adaptability; (7) Peer support; (8) Goal-setting; (9) Engagement; (10) Accessibility. Challenges include (1) Limited physical examination; (2) Lack of human connection; (3) Complexity of individual cases; (4) Privacy and security; (5) Legal and ethical challenges; (6) Language and cultural barriers; (7) Technical limitations; (8) Diagnostic limitations; (9) lack of reliability and trust; (10) Emergency situations.
Al-Anezi (2024) [2]	Saudi Arabia	User experience study; Understand the use and abilities of ChatGPT for mental health support and information seeking; None	Psychological health (mental health)	Qualitative; semi-structured interview; unspecified year	24 mental health patients; university hospital	ChatGPT 3.5; direct; used for 2 weeks	8 themes related to the mode of use were identified. All participants identified the LLM as a tool for education. Between 11 and 18 of 24 used it for emotional support, psychotherapeutic exercises, and self-assessment and monitoring. Between 3 and 8 of 24 identified it as a tool capable of providing information related to crisis intervention, cognitive behavioral therapy, referral and resources, goal setting, and

							motivation.
Al-Anezi (2024) [3][a]	Saudi Arabia	User experience study; Understand how cancer patients experience ChatGPT as a resource for health behavior and lifestyle change; None	Cancer	Qualitative; focus group; Nov 2022 to Apr 2023	72 cancer patients; university hospital	ChatGPT 3.5; direct; used for 2 weeks	Themes emerged that indicate ChatGPT assisted with health literacy and self-management. The self-management theme included feeling supported emotionally by ChatGPT. Concerns expressed include privacy, lack of personalization, and reliability (factual accuracy of the output) issues.
Al-Shboul et al.(2024) [4]	Jordan, Kuwait, United Arab Emirates	User experience study; Understand how participants interact with ChatGPT for health information seeking, including perceived benefits, drawbacks, usefulness, and effectiveness; None	General health	Qualitative; semi-structured interview; 2023	16 adults who had experience using ChatGPT for health information-seeking; social media (Facebook and X)	ChatGPT (version unspecified); Indirect; unknown	ChatGPT was found to be convenient and accessible. Concerns about dependability and trustworthiness were noted. Further personalization and tailoring of responses were identified as important. The need for emotional support and empathy from the chatbot was underscored.
Choudhury et al. (2024) [5]	United States	Consumer survey; Assess perceptions of ChatGPT for health-related information gathering; Unified Theory of Acceptance and se of technology (UTAUT)	General health	Mixed; online survey; Feb to Mar 2023	607 in general; 44 adults who used ChatGPT for health-related queries; survey panel (Centiment)	ChatGPT (version unspecified); Indirect	Qualitative findings show that engaging with ChatGPT for healthcare-related matters demonstrates a pronounced emphasis on safety and trust. There is a critical need for heightened accuracy, security, and ethical considerations, aligning with the sensitive nature of healthcare information and decision-making processes.
Gordon et al. (2024) [6]	United States	Evaluation study; In addition to assessing the accuracy, relevance, and readability of ChatGPT's responses to common imaging-related questions by patients and patient advocates provided feedback to assess the utility of the responses; None	Medical imaging (radiology)	Quantitative; survey; Mar 2023	2 patient advocates; unspecified site	ChatGPT 3.5; Indirect	ChatGPT demonstrates the potential to respond accurately, consistently, and relevantly to patients' imaging-related questions. However, imperfect accuracy and high complexity necessitate oversight before implementation. Prompts reduced response variability and yielded more targeted information, but they did not improve readability. ChatGPT has the potential to increase accessibility to health

							information and streamline the production of patient-facing educational materials; however, its current limitations require cautious implementation and further research.
Karinshak et al. (2023) [7] – study 1	Unspecified; likely United States	Evaluation study; Evaluate perceptions of ChatGPT-generated pro-vaccination messages; None	Vaccination	Quantitative; online survey; Likely before 2022	852 adults; survey panel (Amazon Mechanical Turk)	GPT-3; Indirect	GPT-3 vaccine messages were rated as more persuasive and more effective. Respondents also had a more positive attitude toward GPT-3 vaccine messages. Unvaccinated respondents rated all less favorably. Respondents who have greater education and are Democrat-leaning rated more favorably.
Karinshak et al. (2023) [7] – study 2	Unspecified; likely United States	Evaluation study; Evaluate the effect of messaging source favorability of pro-vaccination messaging; None	Vaccination	Quantitative; online survey; Likely before 2022	1,496 adults; survey panel (Amazon Mechanical Turk)	GPT-3; Indirect	GPT-3 messages were again rated as evoking more positive attitudes, higher strength, and more effective. When labeled as "AI" they were rated less favorably than "CDC" and no source. "Doctor" was not labeled more favorably than "AI." Significant effect between label source and strength. GPT-3 messages rated higher except when labeled as from "AI." That is, respondents preferred GPT-3 messages, but not when they were told they were LLM-created. Trustworthiness was a moderator, explaining why "AI" source label was preferred less than "CDC" and "Doctor."
Kim et al. (2023) [8]	South Korea	Evaluation study; To evaluate changes in patient perceptions regarding AI before and after receiving a ChatGPT-written explanatory note; None	Urolithiasis	Quantitative; survey; 2023	24 urolithiasis patients; likely university hospital	ChatGPT 3.5; Indirect	Significant differences in the summation of negative questionnaire scores between pre- and post-surveys of ChatGPT but not in the positive questionnaire scores. The mean difference of negative questionnaires was increased by 1.3 ± 2.1, indicating that negative emotions, such as worry or wariness relative to the AI, were

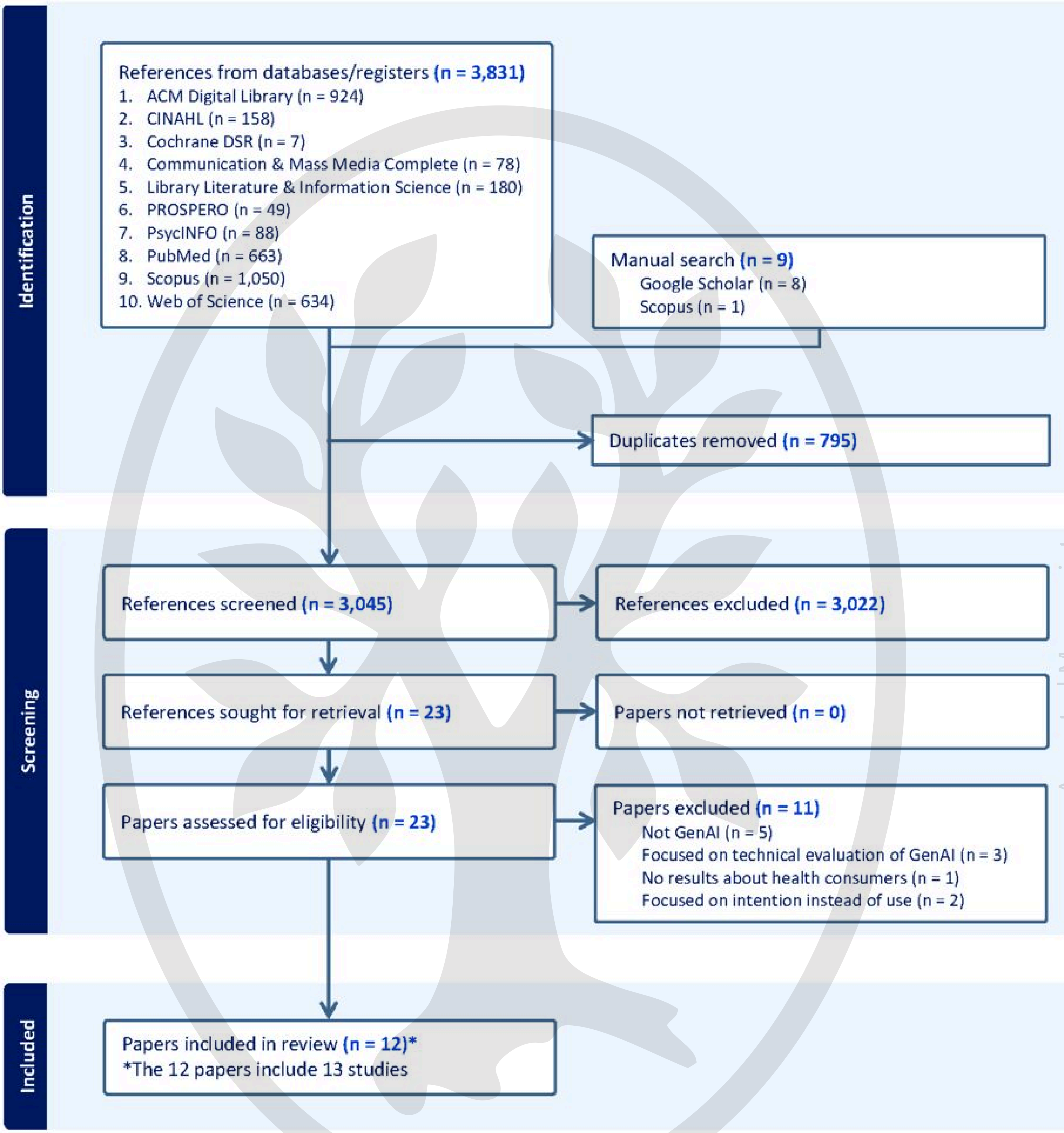
							augmented. Most (80%) agreed or strongly agreed that the generated explanation helped them understand the disease process while only 66% had confidence in the explanation. Linear regression identified education level as positively correlated with satisfaction. No correlations between demographic data and satisfaction questionnaire results ($p > 0.05$).
Lockie & Choi (2024) [9]	Australia	Evaluation study; Compare patients' evaluation of ChatGPT patient information leaflet to a surgeon-created leaflet regarding laparoscopic cholecystectomy; None	Surgical procedure (Laparoscopic cholecystectomy)	Mixed; survey; May-Jun 2023	28 patients undergoing elective laparoscopic cholecystectomy; private hospital	ChatGPT (version unspecified); Indirect	Patients and doctors rated the ChatGPT patient information leaflets higher, by mean score, than the surgeon-created leaflets. Notable qualitative comments from patients about the ChatGPT leaflet include: "well presented", "it is plain simple language easy to read", and "a bit wordy".
Saeidnia et al. (2024) [10]	Iran	Evaluation study; Understand clinicians' and informal caregivers' opinions of ChatGPT as a resource for information seeking about dementia; None	Psychological health (Dementia)	Mixed; structured interviews and survey; Apr 2023	15 informal caregivers of patients with dementia; social media (Instagram, Facebook and Telegram)	ChatGPT (version unspecified); Indirect	Interview findings show that informal caregivers were more positive about using ChatGPT to obtain non-specialized information about dementia compared to formal caregivers (i.e., clinicians). Survey results show that informal caregivers gave higher ratings (M = 3.77 out of 5) of ChatGPT's responsiveness on the items describing information needs than formal caregivers (M = 3.13 out of 5).
Schmidt et al. (2023) [11]	Germany	Evaluation study; Assess the comprehensibility, information density, and conclusion possibilities of simplified MRI findings of the knee joint; None	Medical imaging (MRI of knee joint)	Quantitative; survey; Dec 2022	20 orthopedic patients; university hospital	ChatGPT 3.5; Indirect	Patient evaluation ChatGPT-simplified MRI findings show consistent quality of reports, depending on information complexity. Simplicity of word choice and sentence structure was rated "Agree" on average, with significant differences between simple and complex findings and between moderate

							and complex findings. Patients reported being significantly better at knowing what the text was about and drawing the correct conclusions the more simplified the report of findings was.
Yun et al. (2023)[12]	Unspecified; likely South Korea	Evaluation study; Assess the answers provided by ChatGPT during hypothetical breast augmentation consultations across various categories and depths; None	Surgical procedure (mammoplasty)	Mixed; survey; unspecified year	5 adults; unspecified site	ChatGPT 4; Indirect	Laypeople's mean scores of ChatGPT consultations were significantly higher than surgeons based on reliability (3.61 vs 3.47), information quality (3.81 vs 3.40), overall quality rating (4.52 vs 3.83), and emotion (3.49 vs 3.05).

[a]Not fully published; publication year is based on its online-first release. The rest are based on the official publication year.

References

1 Al-Anezi FM. Exploring the use of ChatGPT as a virtual health coach for chronic disease management. *LEARNING HEALTH SYSTEMS*. Published Online First: 2024. doi: 10.1002/lrh2.10406
2 Al-Anezi F. Assessing the Effectiveness of ChatGPT in Delivering Mental Health Support: A Qualitative Study. *Journal of Multidisciplinary Healthcare*. 2024;17:461–71. doi: 10.2147/JMDH.S447368
3 Al-Anezi F. Examining the role of ChatGPT in promoting health behaviors and lifestyle changes among cancer patients. *Nutrition and Health*. 2024;02601060241244563. doi: 10.1177/02601060241244563
4 Al Shboul MKI, Alwreikat A, Alotaibi FA. Investigating the Use of ChatGpt as a Novel Method for Seeking Health Information: A Qualitative Approach. *Science & Technology Libraries*. 2024;43:225–34. doi: 10.1080/0194262X.2023.2250835
5 Choudhury A, Elkefi S, Tounsi A. Exploring factors influencing user perspective of ChatGPT as a technology that assists in healthcare decision making: A cross sectional survey study. *PLOS ONE*. 2024;19:e0296151. doi: 10.1371/journal.pone.0296151
6 Gordon EB, Towbin AJ, Wingrove P, *et al.* Enhancing Patient Communication With Chat-GPT in Radiology: Evaluating the Efficacy and Readability of Answers to Common Imaging-Related Questions. *Journal of the American College of Radiology*. 2024;21:353–9. doi: 10.1016/j.jacr.2023.09.011
7 Karinshak E, Liu SX, Park JS, *et al.* Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. *Proc ACM Hum-Comput Interact*. 2023;7:1–29. doi: 10.1145/3579592
8 Kim SH, Tae JH, Chang IH, *et al.* Changes in patient perceptions regarding ChatGPT-written explanations on lifestyle modifications for preventing urolithiasis recurrence. *DIGITAL HEALTH*. 2023;9:20552076231203940. doi: 10.1177/20552076231203940
9 Lockie E, Choi J. Evaluation of a chat GPT generated patient information leaflet about laparoscopic cholecystectomy. *ANZ Journal of Surgery*. 2024;94:353–5. doi: 10.1111/ans.18834
10 Saeidnia HR, Kozak M, Lund BD, *et al.* Evaluation of ChatGPT's responses to information needs and information seeking of dementia patients. *Scientific Reports*. 2024;14:10273. doi: 10.1038/s41598-024-61068-5
11 Schmidt S, Zimmerer A, Cucos T, *et al.* Simplifying radiologic reports with natural language processing: a novel approach using ChatGPT in enhancing patient understanding of MRI results. *ARCHIVES OF ORTHOPAEDIC AND TRAUMA SURGERY*. Published Online First: 2023. doi: 10.1007/s00402-023-05113-4
12 Yun JY, Kim DJ, Lee N, *et al.* A comprehensive evaluation of ChatGPT consultation quality for augmentation mammoplasty: A comparative analysis between plastic surgeons and laypersons. *International Journal of Medical Informatics*. 2023;179. doi: 10.1016/j.ijmedinf.2023.105219


Identification
References from databases/registers (n = 3,831)
1. ACM Digital Library (n = 924)
2. CINAHL (n = 158)
3. Cochrane DSR (n = 7)
4. Communication & Mass Media Complete (n = 78)
5. Library Literature & Information Science (n = 180)
6. PROSPERO (n = 49)
7. PsycINFO (n = 88)
8. PubMed (n = 663)
9. Scopus (n = 1,050)
10. Web of Science (n = 634)
Manual search (n = 9)
Google Scholar (n = 8)
Scopus (n = 1)
Duplicates removed (n = 795)
Screening
References screened (n = 3,045)
References excluded (n = 3,022)
References sought for retrieval (n = 23)
Papers not retrieved (n = 0)
Papers assessed for eligibility (n = 23)
Papers excluded (n = 11)
Not GenAI (n = 5)
Focused on technical evaluation of GenAI (n = 3)
No results about health consumers (n = 1)
Focused on intention instead of use (n = 2)
Included
Papers included in review (n = 12)*
*The 12 papers include 13 studies