

---

Review

# Applications, Challenges, and Future Directions of Large Language Models in Health Care Communication: Scoping Review

---

Jing Chang<sup>1\*</sup>, MS; Ruotong Peng<sup>2\*</sup>, PhD; Xi Chen<sup>1</sup>, PhD; Yishu Zhu<sup>1</sup>, MS; Ruting Miao<sup>3</sup>, MS; Zeng Cao<sup>4\*</sup>, MS; Hui Feng<sup>5\*</sup>, PhD

<sup>1</sup>Xiangya School of Nursing, Central South University, Changsha, China

<sup>2</sup>School of Nursing, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

<sup>3</sup>School of Nursing, Jiangxi Medical College, Nanchang University, Nanchang, China

<sup>4</sup>Xiangya Hospital, Central South University, , Changsha, China

<sup>5</sup>Xiangya School of Nursing, Central South University, Changsha, China

\*these authors contributed equally

## Corresponding Author:

Hui Feng, PhD

Xiangya School of Nursing, Central South University

Changsha

Changsha 410083

China

Phone: 1 073182650297

Email: [feng.hui@csu.edu.cn](mailto:feng.hui@csu.edu.cn)

---

## Abstract

**Background:** Effective health care communication is crucial in the medical field. However, effective communication in clinical practice still faces numerous obstacles, and large language models (LLMs) offer various possibilities for improving the quality of medical communication. To date, there are no published reviews on the use of LLMs in health care communication.

**Objective:** This review sought to summarize the applications and challenges of LLMs in health care communication and to identify directions for future research.

**Methods:** A comprehensive literature search was conducted in PubMed, Embase, Web of Science, and the Cochrane Library from January 2018 to November 2025. The search and selection process followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guideline and the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) checklist. Eligible studies used LLMs to facilitate health care communication among the public, patients, and clinicians. Following rigorous data extraction and cross-checking, we conducted a quantitative analysis of characteristics of the included literature. Furthermore, using communication accommodation theory as a framework, we identified application patterns of LLMs in health care communication and summarized current challenges and future directions.

**Results:** Ninety-six studies were included in this review, all published between 2023 and 2025, summarizing 4 patterns of LLM application in health care communication: transforming medical information (n=30), facilitating dynamic interaction (n=38), empowering communication capabilities (n=10), and optimizing clinical workflows (n=18). The role of LLMs in health care communication is undergoing a paradigm shift from “static information processing” to “dynamic intelligent interaction.” Although they show great promise for practical applications, current evaluation methods and dimensions exhibit significant heterogeneity. Furthermore, LLMs still face multiple challenges in their practical application in health care communication, including technical reliability issues, social trust and adoption, interaction and access barriers, and clinical integration challenges.

**Conclusions:** Unlike previous studies that merely touched upon the challenges and future directions, this scoping review uses communication accommodation theory to systematically map the application patterns and developmental landscape of LLM-mediated health care communication. Health care communication powered by LLMs holds significant innovation potential and is currently still in the early stages of rapid development. Future research should focus on optimizing model

performance, strengthening ethical governance frameworks, enhancing human-machine collaboration models, and ensuring responsible application of LLMs in health care through rigorous empirical validation.

**Trial Registration:** OSF Registries 10.17605/OSF.IO/YVXSP; <https://osf.io/yvxsp/overview>

*J Med Internet Res* 2026;28:e84726; doi: [10.2196/84726](https://doi.org/10.2196/84726)

**Keywords:** health care communication; communication accommodation theory; scoping review; digital health; large language models

## Introduction

A good clinician-patient relationship is the foundation of medical practice [1]. Effective health care communication not only facilitates interprofessional collaboration and high-quality care delivery but also positively impacts patient treatment adherence, health outcomes, and overall health care quality [2-5]. However, achieving effective communication in clinical practice remains hindered by numerous barriers. Research indicates that clinicians spend more than 20% of their working hours on communication activities, with economic inefficiencies costing approximately US \$4.9 billion annually [6]. This significant communication burden and efficiency pressure further limit opportunities for patient-centered interactions [5,6]. Additionally, complex medical terminology and language barriers hinder patients' comprehension of health care information, negatively impacting treatment decision-making and medication adherence [7,8].

Large language models (LLMs) offer a transformative paradigm for addressing these challenges. As artificial intelligence (AI) systems trained on massive text corpora, LLMs demonstrate exceptional performance in natural language understanding and generation tasks [9,10]. From text analysis and summary generation to clinical applications, LLMs demonstrate diverse capabilities for clinical support [11-13]. Simultaneously, through domain-specific fine-tuning, LLMs can maintain ongoing engagement with user queries, facilitate interactions, and generate controlled outputs [14]. Furthermore, existing research has highlighted the unique value of LLMs in health care communication by developing chatbots with customized behaviors [14]. Multiple studies indicate that LLMs are key tools for improving information transmission efficiency and alleviating the burden of clinical communication [15-18]. Therefore, leveraging LLMs to optimize information dissemination and communication methods will be a crucial approach to improving the quality of health care communication in the future [19].

In this study, health care communication is defined as the dynamic, interactive process within medical settings that facilitates accurate transfer of clinical information, emotional exchange, and collaborative decision-making through linguistic and nonlinguistic mediation [20-22]. To systematically interpret how LLMs intervene in this complex process, this review introduces the interaction-centered communication accommodation theory (CAT) as an analytical framework [23]. Proposed by Howard Giles in 1973, CAT emphasizes that individuals dynamically adjust speech, intonation, and discourse to manage social distance and

interpersonal relationships [24,25]. In clinical settings, CAT reveals how providers adapt their communication styles to foster trust and understanding [26,27]. CAT identifies 5 sociolinguistic strategies, including approximation, explicitness, interpersonal control, discourse management, and emotional expression [24]. It provides a systematic lens for analyzing how LLMs empower medical communication [28]. Research underscores that CAT uniquely suits large-scale text-based telemedicine among multiple frameworks that elucidate clinical communication [24].

As an innovative and transformative technology, LLM-based health care communication demonstrates immense potential to advance medical communication toward greater efficiency, precision, and personalization [29,30]. However, to date, only one commentary has briefly explored the role and challenges of LLMs in medical communication, highlighting the field's future potential [31]. At present, the specific use cases, challenges, and future directions of LLMs in current applications remain unclear. Therefore, this study aims to systematically map the current state of research in this field through a scoping review and identify known knowledge gaps. The key issues addressed in this review include (1) systematizing application patterns of LLMs in health care communication based on the CAT, (2) exploring current evaluation methods and dimensions in the health care communication domain, (3) identifying limitations and challenges in applying LLMs to health care communication, and (4) proposing recommendations for future research to inform the better development and application of LLM-based health care communication.

## Methods

### *Study Design, Protocol, and Registration*

This scoping review adheres to the methodological framework proposed by Arksey and O'Malley [32], and was conducted in accordance with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews; [Checklist 1](#)) and the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension; [Checklist 2](#)) to ensure methodological transparency and reproducibility [33,34]. Our full review protocol is published in the Open Science Framework registries [35].

### *Information Sources and Search*

To identify relevant English-language literature, the research team systematically searched the PubMed, Embase, Web of Science, and Cochrane Library databases, without using

additional information sources. The initial search was conducted on July 30, 2024, and the final search on November 20, 2025, using the same search strategy to identify newly published research. Each database was searched via its web interface, with searches conducted separately within each database. No research registries or other online sources (such as websites, conference proceedings, and journal directories) were searched. We did not contact researchers to seek additional sources. All search strategies were developed specifically for this scoping review and were not formally peer-reviewed by independent experts before implementation. A filter for publications released after 2018 was applied to all databases. The search strategy combined MeSH (Medical Subject Headings) terms with free-text keywords, with core concepts covering LLMs (eg, large language model OR ChatGPT OR large language model\* OR language neural network\* OR generative AI OR AI-Generated OR generative artificial intelligence OR ChatGPT OR Artificial Intelligence Chatbots OR MedPalm OR GPT OR pretrained model\* OR conversational AI OR deep learning language model\* OR language model\* OR language generation model\*) and communication (eg, Respon\* OR Repl\* OR Report\* OR Question\* OR Transform\* OR Summar\* OR Communicat\* OR Interpret\* OR Explan\* OR Inform\* OR Answer). The specific search terms were determined by the research team through iterative discussions, with search strategies and keywords adjusted based on results to ensure the retrieval of literature spanning both research domains. We adjusted the search strategy based on each database, and the complete search strategy is detailed in [Multimedia Appendix 1 \[36-131\]](#). Additionally, we manually screened the reference lists of included studies for additional eligible records. No forward citation searching was performed.

## Selection of Sources of Evidence

The retrieved literature was managed and deduplicated using EndNote 21 software (Clarivate). Two systematically trained researchers independently screened the titles and abstracts of each article based on inclusion and exclusion criteria. Similarly, the full text of articles included in the title and abstract screening was independently reviewed by 2 authors (JC and RP) for inclusion in the evaluation. Any discrepancies were resolved through discussion, involving a third author (HF) when necessary. Additionally, this study aims to provide a comprehensive overview of LLMs' applications in health care communication. Therefore, the included studies were not assessed for methodological quality to ensure the breadth of the literature review.

## Eligibility Criteria

The review applied the following inclusion criteria:

1. Peer-reviewed empirical studies applying LLMs to health care communication.
2. Published between 2018 and 2025. This cutoff reflects the introduction of Bidirectional Encoder Representations from Transformers, a novel language representation model widely regarded as the origin of contemporary LLMs [132].

The review applied the following exclusion criteria:

1. Research that focuses solely on LLMs as tools for static knowledge retrieval or question-answering (eg, evaluating only their accuracy in answering questions on a medical licensing exam). This review defines "health care communication" as a collaborative process aimed at achieving information exchange, emotional interaction, and shared decision-making through dynamic interaction. Accuracy assessments based on static question-answering focus solely on the quality of knowledge retrieval; they fail to reflect a model's ability to respond in real time to user feedback and emotional needs, nor can they evaluate the model's effectiveness in applying communicative adaptation strategies for dynamic adjustment (for details on excluded categories, refer to [Multimedia Appendix 1](#)).
2. Articles unrelated to health care communication, such as prediction or diagnosis of disease.
3. Study protocols, preprints, trial registrations, editorials, letters, and commentaries.

## Data Charting Process and Data Items

The data for each article were extracted through a pre-designed data extraction form by the research team. Two authors (JC and RP) independently extracted data from the identified studies using Microsoft Excel, including the author, year, country, study design, medical disciplines, targeted population, research objectives, type of model, evaluation methods, evaluation content, application patterns of LLMs in health care communication, challenges, and other relevant information (eg, barriers). Potential discrepancies in data extraction were discussed by the authors (JC and RP) and resolved. At least one additional author (HF) independently verified the accuracy of each literature record to validate the analysis results.

## Synthesis of Results

This study summarizes and analyzes the extracted data. The research team used descriptive statistical methods to summarize the general characteristics, evaluation methods, and evaluation content of the included studies. Additionally, this study used thematic analysis within the CAT framework to identify patterns and strategies for the application of LLMs in health care communication and to outline the existing challenges. This study strictly followed the 3-stage thematic synthesis method proposed by Thomas and Harden [133]. The 2 authors independently performed open coding of the extracted content from the included studies based on the meaning and content of the data. Subsequently, while remaining faithful to the original findings of the included studies, they distilled their core meanings, categorized the open-coded data, and developed descriptive themes. The descriptive themes were subsequently developed into analytical themes focusing on the patterns and challenges of LLM applications in health care communication. To ensure the rigor of the analysis, all initial coding and the thematic framework were independently reviewed and validated by a third researcher (HF). For any discrepancies

in understanding or coding, the research team held multiple meetings for in-depth discussion and, when necessary, redefined the coding manual until the discrepancies were resolved and consensus was reached. The final analysis results were presented through a combination of narrative descriptions and charts, aiming to provide a comprehensive reflection of the current state of this field.

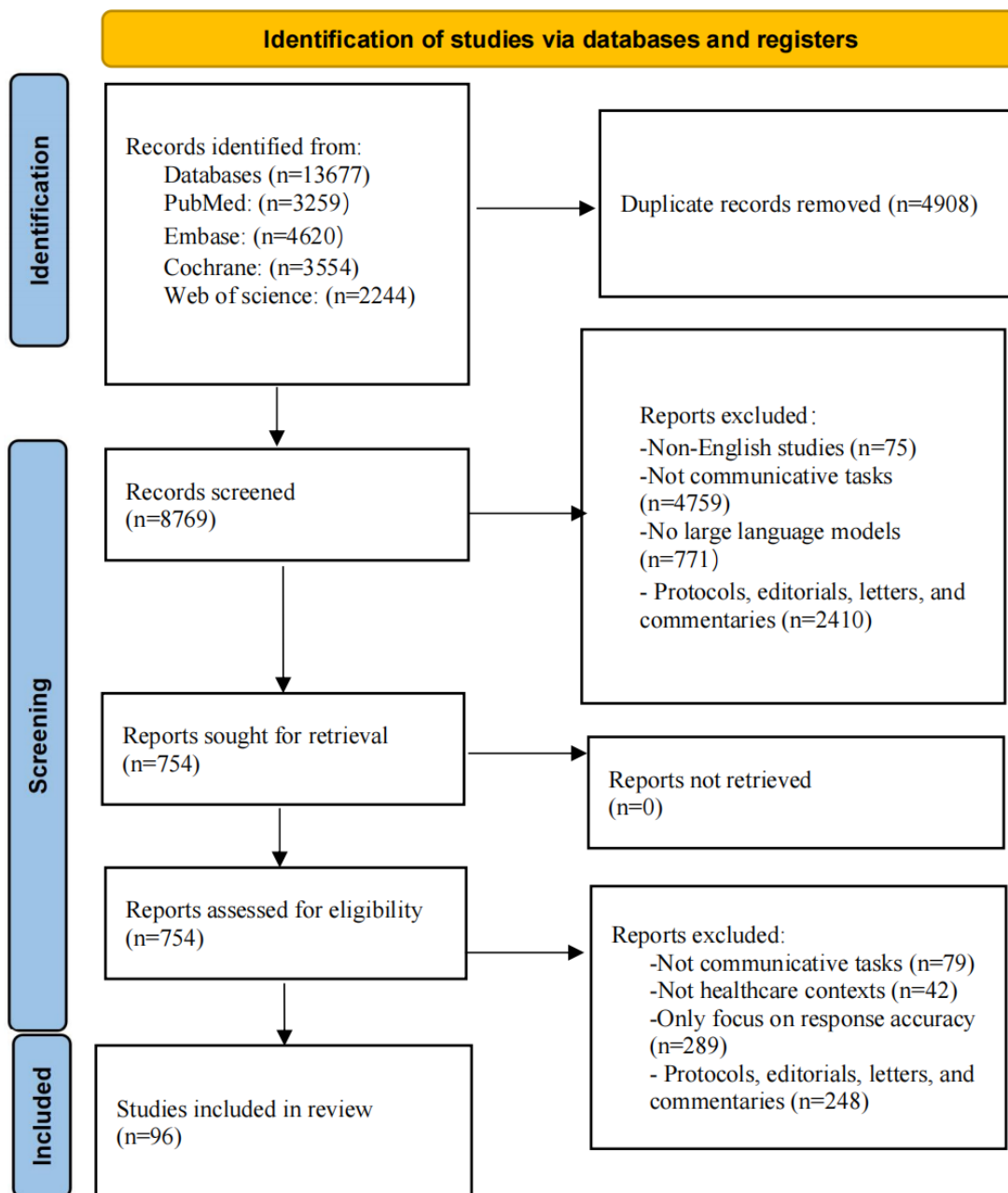
Meta-Analyses) flow diagram (Figure 1). A total of 13,677 articles were retrieved from the database, and 4908 duplicates were removed. The full text of 754 articles was screened after screening the titles and abstracts. Finally, a total of 96 studies [36-131] met the inclusion criteria.

## Results

### Selection of Sources of Evidence

The study screening process is summarized in the PRISMA (Preferred Reporting Items for Systematic Reviews and

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of article selection.



### Characteristics of Sources of Evidence

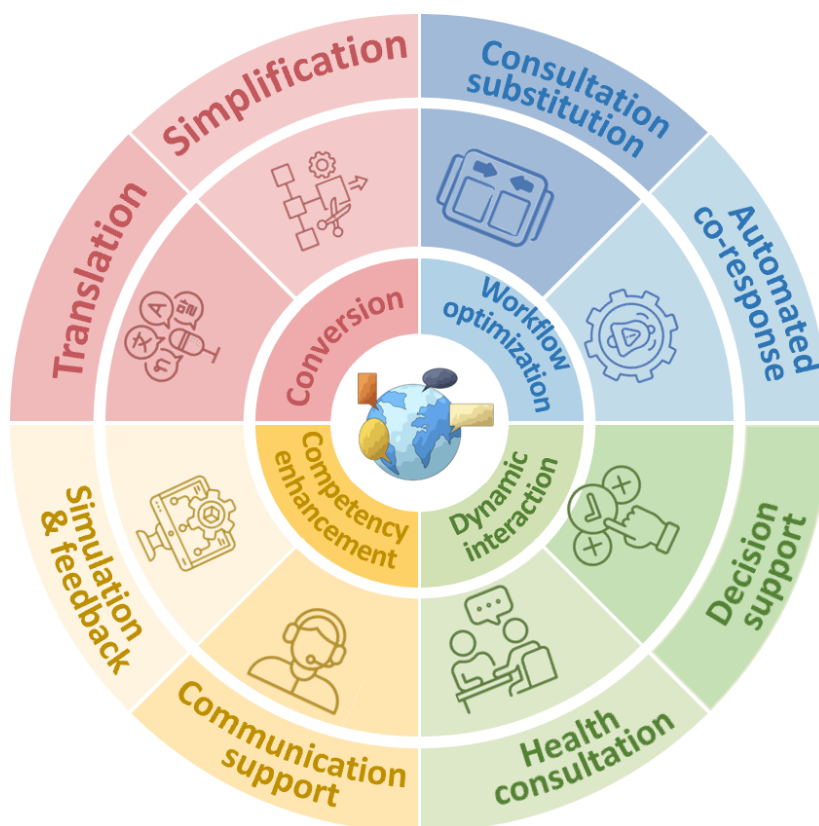
The included studies were published between 2023 and 2025 across 19 countries. The United States had the most studies (n=46) [37-41,44-47,51,53-55,58-60,62,64-66,68,69,73,78,80,83,86,88,94,99,100,102,110,113,116,119,122-131], followed by China (n=13), Turkey (n=10) [70-72,74-76,82,85,89,92], and Germany (n=7) [42,48,50,52,93,103,115]. Other countries included Australia (n=4) [43,101,107,112], South Korea (n=2) [108,117], India (n=2) [49,61], Poland [79], Italy [118], the United Kingdom [36], Ireland [87], Romania [67], the Netherlands [77], Japan [84], Brazil [90], Thailand [120], Sweden [104], Singapore [96], and Colombia [63] (each n=1). The included publications exhibited diverse methodological characteristics, comprising 33 descriptive studies [38,39,41,45,46,49,56,58-60,62,64-66,68,69,71,76,80,82,83,89,94,96,99,100,102,103,115,116,122,126,127], 24 comparative studies [43,47,48,52,61,70,72-75,77,79,81,84,85,87,88,91-93,95,97,98,117], 8 cross-sectional studies [51,54,55,63,78,86,90,125], 6 technology development studies [67,111-113,120,121], 5 randomized controlled trials [37,40,105,114,118], 6 quality improvement studies [44,123,124,128-130], 3 mixed methods studies [108,119,126], and 1 each of proof-of-concept studies [101], observational studies [131], cohort studies[50], qualitative studies [104], exploratory studies [42], pilot studies [36], feasibility studies [109], usability studies [107], multiphase studies

[110], multicenter quantitative studies [57], and survey studies [53]. In terms of the specific models used, 83 studies [36-38,40-42,44-53,55-66,68-97,99,100,102-110,112,115-117,119,122,124-131] examined ChatGPT (OpenAI) performance either on its own or in comparison with other models, 6 studies [39,54,98,101,111,123] did not specify the models used, and 4 studies [67,85,113,114] used models they had developed themselves. Research applications spanned multiple medical disciplines, with radiology being the most common field (n=14) [36,37,39-41,45,46,48-50,52-54,65]. For details on the country, study design, medical disciplines, target population, research objectives, model type, evaluation methods, and evaluation content for each study, please refer to [Multimedia Appendix 1](#).

### Synthesis of Results: Different Categories of Application Patterns of LLMs

The systematic search and screening process identified 96 studies [36-131] meeting the inclusion criteria. Based on the nature of communication tasks, these studies were categorized into four primary application patterns: (1) transforming medical information, (2) facilitating dynamic interaction, (3) empowering communication capabilities, and (4) optimizing clinical workflows (Figure 2). For clarity, Table 1 summarizes the strategies and tactics of CAT, listing their classic definitions and applications in this study.

Figure 2. Four application patterns of large language models in health care communication.



**Table 1.** CAT<sup>a</sup> strategies, definitions, and their application [24].

Strategy	Classic CAT definition	Tactics (this study) and application
Approximation	Adjusting one’s communicative style to be more similar (convergence), more distinct (divergence), or unchanged (maintenance) relative to the interlocutor.	<ul style="list-style-type: none"> <li>• Convergence: use language that resonates with users; simplify jargon.</li> <li>• Divergence: using technical jargon without a lay explanation.</li> <li>• Maintenance: maintain a consistent tone and style regardless of user changes.</li> </ul>
Interpretability	Ensuring messages are understandable by adjusting complexity, clarity, or explicitness.	<ul style="list-style-type: none"> <li>• Simplification: breaking down complex medical terms into lay language.</li> <li>• Clarification: providing additional explanations or examples to aid understanding.</li> </ul>
Discourse management	Regulating conversational flow, including turn-taking and topic control.	<ul style="list-style-type: none"> <li>• Topic initiation: guide the conversation toward relevant topics.</li> <li>• Topic shifting: redirect discussions to maintain focus or address new issues.</li> </ul>
Interpersonal control	Managing role dynamics and authority in interaction.	<ul style="list-style-type: none"> <li>• Assertiveness: delivering confident recommendations and guiding the consultation.</li> <li>• Responsiveness: promptly responding to patient questions and concerns.</li> </ul>
Emotional expression	Conveying empathy, support, and emotional alignment.	<ul style="list-style-type: none"> <li>• Empathy: perceiving the patient’s emotions, expressing empathy.</li> <li>• Support offer: encouraging the patient, providing emotional support.</li> </ul>

<sup>a</sup>CAT: communication accommodation theory.

The application of LLMs in health care communication has shown a significant growth trend and an evolution in functional focus between 2023 and 2025 (Figure 3). In Figure 3, the horizontal axis represents the year of publication, the vertical axis reflects different application patterns of LLMs in health care communication, and the size of the bubbles indicates the number of related studies. Among these, “Transforming medical information” emerged earliest, with the number of related studies steadily increasing from

3 [41,46,49] in 2023 to 17 [36-40,42-45,47-65] in 2025. The “Facilitating dynamic interaction” domain had no recorded applications in 2023, but rapidly increased to 6 in 2024 and further grew to 32 in 2025, becoming the most widely applied domain to date. Meanwhile, “Optimizing clinical workflow” grew from zero applications in 2023 to 10 in 2025. In contrast, “Empowering communication capabilities” started later (with only 2 studies [107,113] in 2024) but had grown to 8 [104-116] by 2025.

**Figure 3.** Evolution application patterns of large language models in health care communication. Note: The horizontal axis shows the distribution of publication years, while the vertical axis illustrates the application patterns of large language models in health care communication; the size of the bubbles indicates the number of studies.



## **Transforming Medical Information**

LLMs bridge the health literacy gap between physicians and patients by converting specialized medical texts into patient-friendly summaries through interpretability and approximation strategies (n=30). Most studies (n=23) focuses on the “deprofessionalization” of medical documentation. Specifically, LLMs were used to simplify radiology reports (n=13), pathology reports (n=5), and discharge summaries (n=5) [36-58]. Through simplification and clarification tactics, clinical terminology is adapted into plain-language information that patients can readily comprehend, reducing comprehension barriers for those with limited medical knowledge. Furthermore, specialized educational materials (eg, kidney stones, hand surgery, shoulder surgery, and palliative care; n=4) [59-62] and cross-language translations (n=3) were generated [63-65]. This approach mimics the expression patterns of nonspecialist audiences, not only eliminating language and comprehension barriers but also ensuring patients can participate equally in health care decision-making during dynamic clinical interactions, thereby promoting equitable dissemination of health information.

## **Facilitating Dynamic Interaction**

### **Dynamic Health Consultation**

LLMs enable dynamic personalized information exchange through multiturn dialogues, transforming health care information access from “static search” to “dynamic interaction” by leveraging the responsiveness substrategy within interpersonal control strategies (n=33). Thirty-three studies used LLMs to capture patient needs and provide expert-level advice by initiating topics within discourse management strategies, addressing disease-related inquiries [66-98]. For instance, He et al [97] investigated whether LLMs could provide patients with inflammatory bowel disease with appropriate advice comparable to that of gastroenterologists. Wu et al [98] evaluated the quality of preventive and therapeutic advice generated by language models for influenza-related queries in online health communities, focusing on their performance in delivering emotional support through emotional expression strategies.

### **Optimizing Shared Decision-Making**

LLMs bridge communication gaps between clinicians and patients by integrating approximation and discourse management strategies (n=5). Two studies used LLMs to provide clinicians with up-to-date clinical guidelines, conversation prompts, and key summaries via approximation strategies, while also helping patients understand disease information, treatment plans, and potential risks to facilitate shared decision-making [99,100]. Three studies examined the use of LLMs to clarify patients’ concerns through interactive dialogue, thereby enhancing the quality of the clinical informed consent process [100-102]. For instance, Allen et al [101] pioneered a 4-phase interactive model (precommunication, language model-guided interaction, clarification phase, and physician review) that uses discourse management strategies to optimize informed consent workflows. Furthermore, this model enables patients to clarify preoperative

concerns through repeatable, low-pressure dialogues using support-offer tactics, thereby improving the efficiency of informed consent [101]. The other 2 studies examined LLM support for informed consent in oxytocin-induced labor and total knee arthroplasty [102,103].

## **Empowering Communication Capabilities**

### **Scenario Simulation and Assessment Feedback**

LLMs support patient-centered communication by simulating real-world clinical scenarios, providing constructive feedback, and integrating biosignal technology (n=8). Seven studies used convergence tactics to simulate diverse patient roles, enabling trainees to engage in authentic role-playing across clinical contexts such as pharmacy, emergency medicine, obstetrics and gynecology, nursing education, and pain communication. This approach facilitates mastery of patient-centered communication techniques [104-110]. Some studies integrated virtual reality technology to simulate virtual wards, enabling concurrent training in verbal communication and physical environment management. LLMs further enhance communication capabilities through emotional expression strategies and interpersonal control strategies. Beyond simulation training, LLMs also provide constructive feedback on the quality and emotional resonance of clinical interactions, reinforcing skills such as active listening and empathy. Additionally, advanced systems such as EEG Emotion Copilot integrate biological signals and LLMs to assist clinicians in identifying patients’ emotions, thereby delivering personalized, emotionally intelligent treatment recommendations [111].

### **Communication Support**

For individuals with specific language or physical impairments, LLMs provide support by integrating approximation and interpretability strategies, thereby promoting communication equality (n=2). Adikari et al [112] integrated a custom-trained LLM into a conversational system, enabling real-time detection and correction of neologisms and semantic errors, as well as intelligent sentence completion suggestions during interruptions. This effectively leverages strategies to help people with aphasia express themselves in standard contexts, ensuring accurate transmission of meaning. For patients with amyotrophic lateral sclerosis, LLM-driven predictive text functionality significantly reduced the physical burden of eye-tracking operations, while clarification tactics enhanced communication fluency and efficiency in clinical settings [113].

## **Optimizing Clinical Workflows**

### **Clinical Consultation Substitution**

Seven studies explored leveraging discourse management strategies to empower LLMs, either supplementing or replacing health care professionals in specialized applications and triage scenarios, thereby streamlining clinical communication processes (n=7) [113-119]. The “SSPEC” chatbot developed by Wan et al [114] demonstrated its ability to

guide patients in recounting medical needs through topic initiation tactics, assist nurses in outpatient reception and triage, and address concerns about model accuracy through response monitoring and early warning mechanisms. In specialized clinical settings, 6 studies integrated interpretability and interpersonal control strategies to enhance the performance of language models in oral hygiene guidance, emergency department management, mental health coping, urology clinical consultations, dental education, and follow-up [115-120]. For example, Chung et al [116] used ChatGPT for preoperative consultations, effectively enhancing patients' disease awareness and optimizing clinical communication efficiency.

### Autonomous Collaborative Response

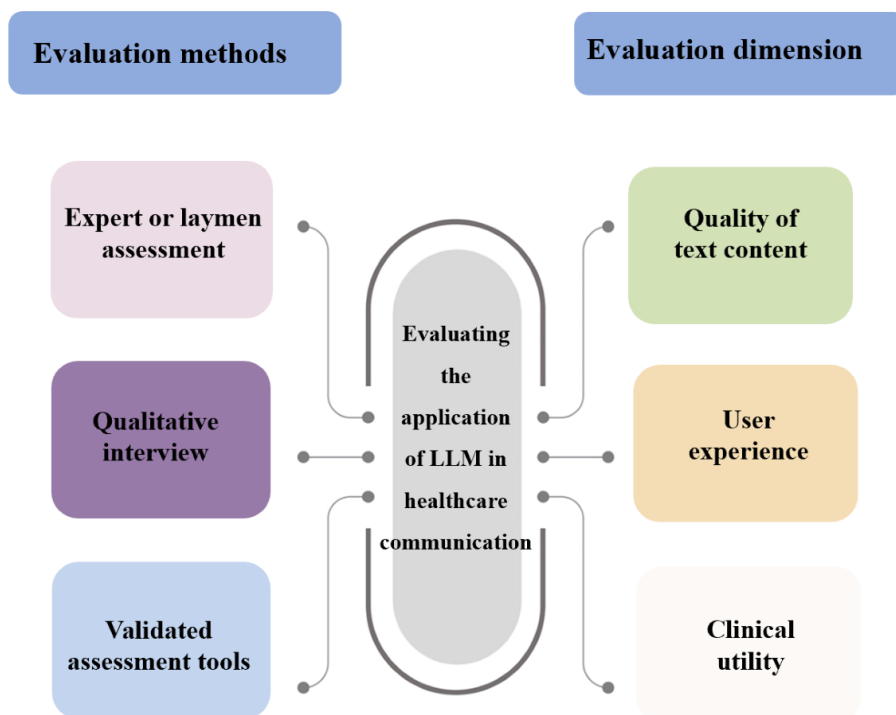
LLMs enhance communication efficiency in health care systems by leveraging interpretability and approximation strategies to automate information generation and enable cross-specialty collaboration (n=11). In prehospital emergency scenarios, LLMs, combined with speech recognition technology, rapidly generate diagnostic summaries to shorten treatment response times [121]. Addressing high-pressure handover environments in emergency departments, Genes et al [122] used LLMs to generate structured transfer summaries, thereby improving emergency department handover processes. Furthermore, this technology enhances

interdisciplinary comprehension by adding lay summaries to highly specialized records (eg, ophthalmology notes), ensuring nonspecialist physicians accurately interpret and implement specialist diagnoses during patient referrals [123]. LLMs also demonstrate utility in routine administrative tasks and electronic patient consultations. Specifically, 6 studies explored their capacity for clinical automation through discourse management strategies in automatically responding to patient portal messages and electronic health record inquiries [124-129]. Two additional studies used convergence tactics to ensure generated content aligns with professional norms and expression conventions, examining the "LLM draft-physician review" practice model's role in reducing health care provider burden and optimizing clinical workflows [130,131].

### Evaluation Methods and Dimensions of LLM Applications

Of the 96 studies included [36-131], 93 evaluated the practical applications of LLMs in health care communication [36-66,68-100,102-118,120-131]. Evaluation methods primarily included subjective assessments, qualitative interviews with domain experts and patients, and objective metric measurements using standardized assessment tools (Figure 4).

**Figure 4.** Mapping of evaluation methods and dimensions. LLM: large language model. Note: Validated assessment tools, for example, the Flesch-Kincaid Grade Level is a widely recognized readability formula that estimates the reading level of a text based on average sentence length and vocabulary complexity and generates a corresponding score [77].



The evaluation tools exhibited a diverse range of characteristics. Existing studies primarily used Likert scales, custom scoring systems, and questionnaires to evaluate

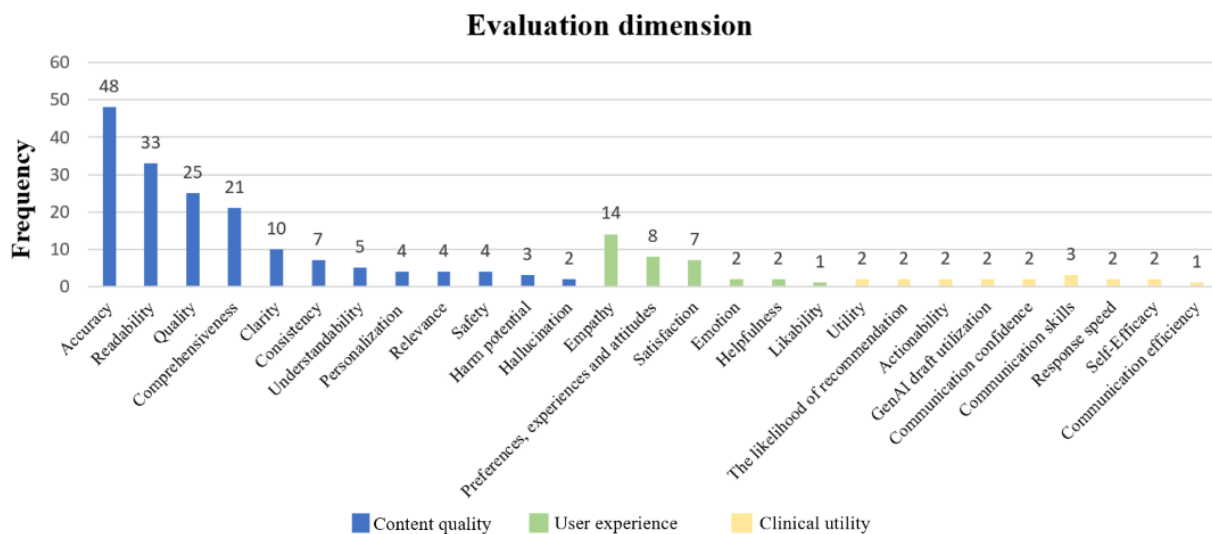
content, combined with qualitative interviews to capture user experience [88,89,110]. A small number of metrics use objective measurement tools; readability metrics include the

Flesch Reading Ease, Flesch-Kincaid Grade Level, Coleman-Liau Index, Simplified Nonsense Measure, Gunning-Fog Index, and the Flesch-Szigris formula [77,85,88]. Some researchers used the DISCERN tool (Deborah Charnock), the Global Quality Score, and the Misinformation Rating Scale to assess content quality [75,90]. Furthermore, measurement tools tailored to specific clinical scenarios, such as the Communication Confidence Self-Assessment Scale and the Mother-Infant Care Communication Assessment Scale, have been used to evaluate changes in users' communication skills [105,106].

The evaluation dimensions primarily include content quality, user experience, and clinical utility, with content quality being the core focus of the research (Figure 5). A total of 80 studies used the following criteria to assess content quality: accuracy, consistency, readability, clarity, comprehensiveness, overall quality, safety, and hallucinations [36-66,68-97,99,100,102,103,110,112,114-117,120-123,125-129]. Among these, accuracy (n=48) and readability (n=33) were the most frequently applied evaluation metrics. Twenty-nine studies evaluated users' experiences with LLM-based health care communication [53-55,61,78,79,81,84,86-88,93,98,104,105,108-110,114,117-119,123,125-130]. These evaluations

covered not only metrics such as satisfaction, likability, and practicality but also collected user experiences and subjective preferences through qualitative research [54,55,104,109]. Additionally, some studies assessed model performance from an emotional perspective (such as empathy) [78,79]. Regarding clinical utility, 27 studies examined the feasibility of integrating LLMs into clinical workflows and their impact on medical practice [36,42,47,57,61,75,81,85,87,97,104-109,113-115,117,118,122-124,126,128,131]. Evaluation metrics varied by specific clinical communication task, primarily focusing on feasibility, practical impact, and efficiency. For example, studies used metrics such as willingness to use, usability, usefulness, and applicability to assess feasibility [61,109]. The impact of LLMs as assistive tools on clinical practice was evaluated by assessing improvements in communication confidence, medical history collection, empathy, and communication skills [105,106]. Furthermore, efficiency metrics were primarily measured by quantifying relevant parameters within clinical workflows, including communication efficiency and AI draft usage rates [123,124,129]. For specific details, please refer to Multimedia Appendix 1.

Figure 5. Evaluation dimensions and frequencies. GenAI: generative artificial intelligence.



### Comparison of Performance Differences

A total of 19 studies evaluated performance differences among mainstream LLMs in health care communication [48, 51,61,63,70,71,73-75,77,85,89-93,95,99,119]. In tasks such as dental implant consultations, melanoma management, patient-physician communication regarding rare diseases, and the interpretation of pathology reports, ChatGPT consistently outperformed Bard (now Gemini; Google) in response accuracy; however, in rhinoplasty consultations, its performance was slightly inferior to Claude's performance [51,63,73, 77,92,93]. In obstetric care consultations, ChatGPT, Kimi (Moonshot AI), and ERNIE Bot (Baidu) showed similar

performance in terms of accuracy and completeness [95]. Although ChatGPT demonstrated high accuracy rates in multiple studies, its readability scores were generally low. In contrast, DeepSeek demonstrated advantages in optimizing the accessibility of medical information, with responses that outperformed GPT and Gemini in clarity, comprehensiveness, and readability [70,74,91,92]. In mental health interactions, Pi outperforms ChatGPT in empathy and user acceptance due to its human-like response style [119]. Regarding multilingual performance, ChatGPT performs best at interpreting pathology reports in Spanish, while Perplexity stands out on English tests [119].

Three studies compared the clinical performance of general-purpose models with domain-specific models [84,93,126]. In generating draft responses to patients, ChatGPT significantly outperformed specialized models in overall scores and responsiveness metrics. However, expert evaluations noted that responses from the specialized model CLAIR better aligned with physicians' professional language styles, whereas ChatGPT's answers were perceived as having a "robotic" tone [126]. In tasks involving doctor-patient communication regarding rare diseases, ChatGPT outperformed BioMistral. However, BioMistral 7B can run

locally within a medical setting, offering privacy advantages [93]. Additionally, Özcivelek and Özcan [85] noted that the domain-specific model Dental GPT demonstrated the highest factual accuracy in consultations regarding oral and maxillo-facial prosthetics, but performed worst in terms of readability.

### Existing Challenges

We have identified four existing challenges for LLMs in health care communication: (1) technical reliability issues, (2) social trust and adoption, (3) interaction and access barriers, and (4) clinical integration challenges (Figure 6).

Figure 6. Existing challenges.



### Technical Reliability Issues

Research indicates that, to improve readability for patients, LLMs may oversimplify medical terminology, potentially omitting critical details [58]. Furthermore, constrained by the "hallucination," models may produce associated inferences with ambiguous terminology definitions, erroneous mechanism descriptions, or insufficient evidence [36,82]. Furthermore, the effectiveness of LLM responses depends heavily on the quality and timeliness of training data. Most current models are trained on general internet corpora rather than rigorously vetted specialized medical datasets, and may generate information that is inconsistent with clinical guidelines or outdated [75]. Combined with inherent "black-box characteristics" and data update delays, this further undermines the interpretability and reliability of their outputs [97,98].

### Social Trust and Adoption

LLMs lack a holistic understanding of patients' clinical contexts, social backgrounds, and psychological states. Their responses may lack sufficient depth to support the comprehensive situational judgment required in medical practice, making it difficult to address complex clinical issues [87]. At the emotional-interaction level, LLMs cannot perceive nonverbal cues such as facial expressions and tonal shifts, making it difficult to gauge patients' emotional states and support needs [58]. Moreover, the directive tone and verbose expressions in their responses may provoke resistance among individuals experiencing emotional crises [119]. Furthermore, existing models are predominantly built on English-language datasets and Western value systems, exhibiting significant limitations in multicultural health care settings and potentially leading to biased outputs [128]. Furthermore, unclear legal

and ethical accountability, coupled with data privacy and security risks, collectively undermine societal trust in LLMs within serious medical contexts [58].

### Interaction and Access Barriers

The clinical efficacy of LLMs relies heavily on the quality of user prompts [43]. Existing research predominantly uses standardized question sets for testing; yet, real-world clinical inquiries often exhibit ambiguity, emotionality, or unstructured characteristics. Populations with lower health literacy may struggle to obtain accurate information through effective interaction. Additionally, regional disparities in model deployment may create access barriers for economically constrained areas and populations [98].

### Clinical Integration Challenges

Currently, clinical text summaries generated by LLMs often lack consistent formatting and standardization, making them difficult to integrate directly into existing clinical workflows or electronic health record systems [41]. Furthermore, these models tend to provide guideline-based "standardized" responses, struggling to tailor recommendations based on patient history, psychosocial factors, or individual preferences [70,97]. They also fail to dynamically optimize advice based on real-time assessments or rehabilitation progress. More critically, existing models lack mechanisms for medical risk assessment and emergency referral, and are unable to provide timely guidance during patient health crises, posing potential safety risks [119]. The lack of real-world clinical validation further hinders their ability to serve as true clinical substitutes in complex, dynamic doctor-patient interactions [97,121,125].

## Discussion

### Summary of Evidence

This review systematically explores the application patterns of LLMs in health care communication. Existing research has evaluated content quality, clinical utility, and user experience of LLMs primarily through objective and subjective metrics. Despite promising prospects, the field currently faces multiple challenges, which provide direction and focal points for future research. As technology advances and applications expand, LLMs will play an increasingly vital role in health care communication.

### Application Domains and Development Trends of LLMs

Relevant literature was published between 2023 and 2025, coinciding with the November 2022 release of ChatGPT. The geographic distribution of the included studies showed a bias toward developed countries, such as the United States. This may be attributed to the region's thriving IT industry, advanced health care infrastructure, and ample funding support [134]. ChatGPT emerged as the most frequently cited model in this review due to its practicality and ease of use for clinicians [135]. The studies also encompassed general-purpose models such as Gemini and DeepSeek, alongside a few custom-developed models, highlighting the breadth of models explored in the included research. By testing and comparing different models, researchers can systematically evaluate performance variations and identify model biases [136]. Application domains show diversification trends. As research advances and models optimize, customized professional models focused on clinical specialties are expected to be deployed across multiple scenarios.

The application of LLMs in health care communication is shifting from "static information processing" to "dynamic intelligent interaction." Transforming medical information is the earliest and most steadily growing area. These tasks rely primarily on structured reasoning and linguistic fluency rather than on complex diagnostic reasoning, giving them an early advantage in technological implementation [137]. As the technology continues to advance, the research focus has shifted toward addressing the more challenging demands of real-time feedback and multiturn dialogue. For instance, the sustained growth in the field of facilitating dynamic interaction indicates that LLMs are gradually evolving into "participants" capable of assisting doctor-patient interactions. Concurrently, the emergence of applications aimed at optimizing clinical workflows reflects a trend toward integrating LLM technology, specifically through the automation of communication tasks to alleviate the increasingly severe administrative burden and clinical pressure within health care systems. Although the field of communication enhancement had a relatively late start, its sustained growth demonstrates that the potential of LLMs to strengthen emotional connections between doctors and patients and assist in complex decision-making is gradually being realized. Future research should focus on translating

LLMs' efficacy into practical clinical settings, particularly on their real-world performance in multicenter, large-scale scenarios [125].

### Main Findings

LLMs empowered by 5 communication accommodation strategies are pioneering new approaches to health care communication. Through strategic language adaptation, they enable dynamic interaction with patients' needs, demonstrating the potential to bridge gaps in health care resources and improve the quality of care [24]. Current research predominantly focuses on low-level tasks such as text translation and basic consultations, reflecting that LLMs remain in the early stages of integrating into health care communication. Research indicates that simplification and clarification tactics are particularly crucial for addressing the inherent limitations of text-based online consultations [24]. Through these strategies, LLMs simplify medical texts and facilitate cross-language translation, thereby enhancing the transparency of communication between doctors and patients. One study indicates that multimodal LLMs are evolving toward context-aware capabilities [137]. This will help LLMs adjust their communication strategies in real time to achieve "communication adaptation," further enhancing the interpretability, clinical accuracy, and empathy of the text generated by the models.

The emergence of specialty-specific response chatbots highlights LLMs' vast potential to improve accessible medical consultations and empower patients' health management. By incorporating interpersonal control strategies, LLMs are driving a shift in medical interactions from static retrieval to interactive consultation. This strategy is essential for adhering to patient-centered care principles in virtual consultations [138]. In stigmatized domains such as sexually transmitted diseases and mental disorders, LLMs leverage anonymity and accessibility to create low-pressure communication environments [139]. This effectively reduces patients' need for impression management and fear of self-disclosure, encouraging authentic and in-depth expression [140]. As communication intermediaries, LLMs optimize the understanding and sharing of information through approximation strategies, thereby supporting shared decision-making between doctors and patients [103]. Research has shown that when communication methods align with patient needs, the effectiveness of consultations is significantly enhanced [138]. Future research should quantify the substantive impact of LLM interventions on decision conflict and clinician-patient trust. Given insufficient clinical validation, LLMs cannot replace clinical judgment. Still, they should serve as "communication copilots," improving medical communication through collaborative methods such as dynamic dialogue monitoring and refining medical history details [103]. Subsequent studies can focus on developing clinically adapted interactive tools to achieve optimal human-machine collaboration [141].

High-fidelity clinical scenario simulations generated by LLMs offer a low-cost, highly scalable solution for communication skills training. By dynamically simulating diverse

patient profiles, LLMs adapt their communication to align with users' language and needs, fostering deeper understanding and connection [24]. However, existing systems are still in their early stages of development, exhibiting limitations in processing nonverbal cues and in addressing privacy and security risks [142]. Consequently, virtual training should be conducted under human supervision, with continuous iteration based on professional feedback to achieve rapid improvement. Future efforts must focus on enhancing the reliability, safety, and scientific validity of virtual patients [143]. Furthermore, relying solely on automated assessment for communication skills may overlook students' psychological and emotional needs [143]. Therefore, establishing a "human-machine collaborative" assessment system is essential to balance teaching efficiency with humanistic care [142,144]. Additionally, the development of customized intervention programs for individuals with language or motor impairments that incorporate interpretability strategies holds profound significance for enhancing the communication autonomy of this marginalized patient population, thereby improving their social participation, vocational integration, and quality of life [112,113]. It is crucial to emphasize that patient-facing tools must be designed within reasonable parameters and built with appropriate safeguards to ensure safety [54].

LLMs demonstrate application value in optimizing clinical workflows, with their core lying in deep integration with clinical dialogue. Preconsultation effectively enhances patient disease awareness [145]. LLMs empowered by discourse management strategies can serve as foundational consulting tools to alleviate time pressures in clinical work and enhance the efficiency of in-person consultations [24]. Research has shown that effective discourse management strategies can maintain conversational coherence and address patient concerns, thereby fostering trust between physicians and patients in digital settings [138]. However, the absence of nonverbal cues can hinder effective communication, necessitating greater linguistic flexibility and highlighting the importance of integrating convergence tactics into communication systems [24]. Interdisciplinary summaries and automated responses generated through such strategies have demonstrated their potential to enhance collaborative efficiency. Although LLMs show potential in simulated treatment dialogues, there is currently a lack of empirical evaluation of their effectiveness in real clinical settings [146]. Therefore, LLMs should currently be positioned as clinical adjunct tools. Their successful deployment may be constrained by existing clinical workflows and health care systems' capacity to integrate novel communication tools alongside training resources [147,148]. To fully unlock the medical benefits of LLMs, forward-looking policy frameworks and industry standards must be established to ensure their effective integration with clinical practice.

### **Evaluation Methods and Dimensions**

There is significant heterogeneity in the evaluation of health care communication research using LLMs, making it difficult to accurately assess the models' task performance and

practical effectiveness and obscuring their potential in clinical practice.

Currently, expert evaluation remains the primary assessment method in this field, with only a few metrics evaluated using validated tools. Most studies focus on subjective metrics and user experience, lacking objective quantitative metrics and evaluation tools, which limits the comparability of research findings. In response, some scholars have proposed advancing the quantification and structuring of evaluation methods. For example, Huo et al [149] suggested developing and applying quantitative metrics to evaluate model outputs. Furthermore, the lack of a unified, standardized evaluation framework has led to significant heterogeneity in existing evaluation tools and dimensions, further hindering effective comparisons across studies [150]. Wei et al [151] provided insights for establishing evaluation guidelines for LLMs in medical response by integrating factors such as model versions and prompt design. Future work should prioritize developing standardized evaluation frameworks tailored to medical contexts and exploring hybrid assessment methods that combine human expert reviews, user feedback, and automated metrics.

Regarding evaluation content, existing research primarily focuses on content quality, clinical utility, and user experience [146]. However, metrics concerning ethical considerations, such as fairness and bias, remain underevaluated. Bedi et al [146] emphasize that incorporating bias into evaluation frameworks can be an effective way to mitigate harmful biases in LLMs. Traditional model evaluation has primarily focused on the accuracy of medical question-answering tasks; however, due to the lack of objective metrics, it is difficult to comprehensively assess the true effectiveness of LLMs in complex medical communication scenarios [68, 152]. Overall, current evaluation practices for LLM-based health care communication lack rigor. More controlled methods are needed to enhance scalability and scientific rigor, such as using validated tools, clearly defining evaluation dimensions, standardizing assessment criteria, and systematically examining changes in patient behavior or clinical outcomes [143]. Therefore, future evaluation systems must balance technical performance, clinical effectiveness, and ethical compliance to establish comprehensive and reliable metrics for the responsible application of LLMs in health care.

### **Comparison of Performance Differences**

Research indicates that performance differences among models stem from heterogeneity in their training datasets, algorithmic architectures, and underlying model capabilities [16,153]. Consistent with previous studies, ChatGPT outperforms most general-purpose models in response accuracy, but its limited readability limits its practical application [154]. However, the success of medical communication depends not only on the accurate transmission of information but also on whether that information can be translated into advice that patients can understand and act upon [155].

Across different clinical specialties, model performance exhibits heterogeneity. Compared to domain-specific fine-tuned models, general-purpose models may exhibit reduced reliability in highly specialized contexts due to a lack of specific training [156]. Taking Dental GPT as an example, this model demonstrates high accuracy and relatively low readability in the field of oral and maxillofacial prosthetics [85]. This aligns with previous research indicating that chatbots trained on domain-specific datasets outperform general-purpose language models [157]. This may be because the model was developed specifically for the dental field; the highly specialized training dataset ensures the model's precise grasp of complex medical facts, though it may also introduce comprehension barriers due to specialized terminology. The CLAIR series of models used by Liu et al [126] demonstrated greater empathy in generating draft patient responses, attributed to the model's fine-tuning on real clinical scenario data. This helps the model mimic clinicians' communication styles and care practices in actual practice, achieving a higher degree of clinical realism [126].

Additionally, linguistic differences can also impact model performance. Recent studies indicate that factors influencing LLM performance include not only model capacity but also linguistic diversity, contextual nuances, and the interaction of multimodal content [158,159]. The diversity of clinical specialties and languages underscores the urgent need to develop specialized, multilingual models [160]. Future research should not be limited to increasing model capacity but should also focus on balancing medical authority with public readability through domain-specific fine-tuning, thereby bridging the gap between professional depth and human communication [161].

## Challenges and Future Directions

Overall, we identified existing challenges and recommended ways to address them in the future. A summary of these recommendations is presented in [Multimedia Appendix 1](#).

Concerns regarding the clinical reliability of models arise from issues such as "hallucinations," "black-box nature," and information omissions. Research indicates that the quality of model responses is highly dependent on the training data [162]. Therefore, integrating knowledge graphs with multimodal data to construct standardized training corpora holds promise as an effective solution. Knowledge graphs ensure the accuracy and security of data sources [163]. The fusion of multimodal data facilitates comprehensive analysis of patient information, enhancing LLMs' capabilities in complex clinical consultations [164]. Furthermore, given the dynamic evolution of medical knowledge, incorporating retrieval-augmented generation frameworks not only dynamically integrates the latest clinical guidelines and medical evidence to improve the timeliness of model outputs but also enhances system transparency [165].

Limitations in emotional interaction and insufficient depth of expression, as well as legal-ethical concerns, represent significant bottlenecks affecting the social adoption and trust in LLMs. Future research may explore multimodal technologies integrating speech recognition, facial expression

analysis, and text comprehension with "human-machine collaboration" models to build context-adaptive, empathetic interaction systems [166]. The HAILEY system, developed by Sharma et al [167], provides empathy-based communication suggestions for physicians, precisely supporting medical interactions. Simultaneously, there is an urgent need to build culturally adapted regional corpora to enhance the representation of marginalized groups and eliminate systemic biases [168]. Additionally, multiple studies emphasize the importance of establishing a comprehensive legal framework [169]. Regarding data governance, strict safeguards for patients' rights to informed consent, data access, and data deletion must be implemented alongside rigorous access control and verification mechanisms [170]. Differential privacy technology, which ensures that individual data remains unidentifiable by introducing controlled noise, represents an effective solution [171].

To overcome interaction and access barriers, future research should focus on developing structured question-assistance frameworks and prompt optimization to guide users in providing structured, complete information, thereby reducing cognitive load during model interactions [172]. Digital inclusion remains paramount, as significant racial, gender, and educational disparities persist in internet access and digital literacy [173]. We therefore advocate for developing training programs to enhance clinicians' human-machine collaboration skills and implementing public digital literacy education to advance health equity.

LLMs face multiple limitations in clinical interactions. Model outputs should adhere to standardized templates and structured guidelines to ensure consistency and completeness [174]. With explicit authorization and strict privacy safeguards, granting secure model access to patient electronic health records may be a strategy for personalization [175]. Additionally, real-time, multitiered risk identification and response mechanisms should be established, integrating multimodal interaction capabilities to detect emotional shifts or psychological issues [119]. Ultimately, establishing channels for human intervention in scenarios involving complex decision-making or deep emotional support will systematically enhance the applicability and safety of models in clinical settings [166].

In summary, unlike previous studies that merely touched upon the challenges and future directions, this study marks the first systematic application of CAT to research on the use of LLMs in health care communication. It provides a comprehensive overview of the current state of applications and prospects in this field. The study lays the groundwork to unlock the potential of LLMs to optimize communication and to promote their responsible use and high-quality development in clinical practice.

## Limitations

Several limitations of the scoping review must be acknowledged. First, given the inherent conceptual breadth and interdisciplinary nature of LLMs, coupled with the rapid evolution of related concepts, it remains challenging to completely rule out the possibility of omissions despite the

comprehensive retrieval strategy used in this study. This study did not use quantitative measures such as Cohen kappa to assess coding consistency; therefore, it has limitations in reflecting statistical reliability among coders [176]. Additionally, excluding non-English literature may introduce selection bias. Second, most studies originate from high-income countries, such as the United States. This not only influences the scope of ethical discussions and the methods used to address specific issues but also raises questions about the applicability of these findings in low- and middle-income countries. Third, this study aims to provide a comprehensive overview of LLM applications in health care communication. To ensure the comprehensiveness of literature inclusion, we did not conduct a quality assessment of the included studies. Fourth, the rapidly evolving nature of LLMs means our findings primarily reflect the landscape as of the search date; subsequent new models and applications may alter current patterns.

## Conclusion

This review is the first to systematically summarize the application patterns of LLMs in health care communication by applying CAT. Currently, LLMs are still in the early stages of integration into clinical practice, and their widespread adoption continues to face challenges, including technical reliability, social trust and acceptance, barriers to interaction and access, and clinical integration. Future research should focus on optimizing model performance, strengthening ethical governance frameworks, and refining human-machine collaboration models, while ensuring safe application in the health care sector through rigorous empirical validation. This study highlights the potential of LLMs to optimize health care communication and is expected to promote their responsible application and high-quality development in medical practice.

---

## Acknowledgments

The authors declare that no generative artificial intelligence (AI) was used in the production of this manuscript.

---

## Funding

The research was supported and funded by the Key Research and Development Program of Hunan Province (2025JK2118).

---

## Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Database search strategies, types of studies excluded, characteristics of included studies, and existing challenges.

[\[DOCX File \(Microsoft Word File\), 54 KB-Multimedia Appendix 1\]](#)

---

## Checklist 1

PRISMA-ScR checklist.

[\[DOCX File \(Microsoft Word File\), 70 KB-Checklist 1\]](#)

---

## Checklist 2

PRISMA-S checklist.

[\[DOCX File \(Microsoft Word File\), 17 KB-Checklist 2\]](#)

---

## References

1. Singh M. Communication as a bridge to build a sound doctor-patient/parent relationship. *Indian J Pediatr*. Jan 2016;83(1):33-37. [doi: [10.1007/s12098-015-1853-9](https://doi.org/10.1007/s12098-015-1853-9)] [Medline: [26228061](https://pubmed.ncbi.nlm.nih.gov/26228061/)]
2. Medendorp NM, van den Heuvel LM, Han PKJ, Hillen MA, Smets EMA. Communication skills training for healthcare professionals in providing genetic counseling: a scoping literature review. *Patient Educ Couns*. Jan 2021;104(1):20-32. [doi: [10.1016/j.pec.2020.07.018](https://doi.org/10.1016/j.pec.2020.07.018)] [Medline: [32782179](https://pubmed.ncbi.nlm.nih.gov/32782179/)]
3. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open*. Jan 3, 2013;3(1):e001570. [doi: [10.1136/bmjopen-2012-001570](https://doi.org/10.1136/bmjopen-2012-001570)] [Medline: [23293244](https://pubmed.ncbi.nlm.nih.gov/23293244/)]
4. Riedl D, Schüßler G. The influence of doctor-patient communication on health outcomes: a systematic review. *Z Psychosom Med Psychother*. Jun 2017;63(2):131-150. [doi: [10.13109/zptm.2017.63.2.131](https://doi.org/10.13109/zptm.2017.63.2.131)] [Medline: [28585507](https://pubmed.ncbi.nlm.nih.gov/28585507/)]
5. Högländer J, Holmström IK, Lövenmark A, Van Dulmen S, Eide H, Sundler AJ. Registered nurse-patient communication research: an integrative review for future directions in nursing research. *J Adv Nurs*. Feb 2023;79(2):539-562. [doi: [10.1111/jan.15548](https://doi.org/10.1111/jan.15548)] [Medline: [36534429](https://pubmed.ncbi.nlm.nih.gov/36534429/)]
6. Vermeir P, Vandijck D, Degroote S, et al. Communication in healthcare: a narrative review of the literature and practical recommendations. *Int J Clin Pract*. Nov 2015;69(11):1257-1267. [doi: [10.1111/ijcp.12686](https://doi.org/10.1111/ijcp.12686)] [Medline: [26147310](https://pubmed.ncbi.nlm.nih.gov/26147310/)]

7. Meuter RFI, Gallois C, Segalowitz NS, Ryder AG, Hocking J. Overcoming language barriers in healthcare: a protocol for investigating safe and effective communication when patients or clinicians use a second language. *BMC Health Serv Res*. Sep 10, 2015;15(371):371. [doi: [10.1186/s12913-015-1024-8](https://doi.org/10.1186/s12913-015-1024-8)] [Medline: [26357948](https://pubmed.ncbi.nlm.nih.gov/26357948/)]
8. Schillinger D, Duran ND, McNamara DS, Crossley SA, Balyan R, Karter AJ. Precision communication: physicians' linguistic adaptation to patients' health literacy. *Sci Adv*. Dec 17, 2021;7(51):eabj2836. [doi: [10.1126/sciadv.abj2836](https://doi.org/10.1126/sciadv.abj2836)] [Medline: [34919437](https://pubmed.ncbi.nlm.nih.gov/34919437/)]
9. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large language models in medicine: the potentials and pitfalls: a narrative review. *Ann Intern Med*. Feb 2024;177(2):210-220. [doi: [10.7326/M23-2772](https://doi.org/10.7326/M23-2772)] [Medline: [38285984](https://pubmed.ncbi.nlm.nih.gov/38285984/)]
10. Min B, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: a survey. *ACM Comput Surv*. Feb 29, 2024;56(2):1-40. [doi: [10.1145/3605943](https://doi.org/10.1145/3605943)]
11. Bicknell BT, Butler D, Whalen S, et al. ChatGPT-4 omni performance in USMLE disciplines and clinical skills: comparative analysis. *JMIR Med Educ*. Nov 6, 2024;10:e63430. [doi: [10.2196/63430](https://doi.org/10.2196/63430)] [Medline: [39504445](https://pubmed.ncbi.nlm.nih.gov/39504445/)]
12. Kaczmarczyk R, Wilhelm TI, Martin R, Roos J. Evaluating multimodal AI in medical diagnostics. *NPJ Digit Med*. Aug 7, 2024;7(1):205. [doi: [10.1038/s41746-024-01208-3](https://doi.org/10.1038/s41746-024-01208-3)] [Medline: [39112822](https://pubmed.ncbi.nlm.nih.gov/39112822/)]
13. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. Mar 30, 2023;388(13):1233-1239. [doi: [10.1056/NEJMs2214184](https://doi.org/10.1056/NEJMs2214184)] [Medline: [36988602](https://pubmed.ncbi.nlm.nih.gov/36988602/)]
14. Wu T, He S, Liu J, et al. A brief overview of ChatGPT: the history, status quo and potential future development. *IEEE/CAA J Autom Sinica*. 2023;10(5):1122-1136. [doi: [10.1109/JAS.2023.123618](https://doi.org/10.1109/JAS.2023.123618)]
15. Yu E, Chu X, Zhang W, et al. Large language models in medicine: applications, challenges, and future directions. *Int J Med Sci*. 2025;22(11):2792-2801. [doi: [10.7150/ijms.111780](https://doi.org/10.7150/ijms.111780)] [Medline: [40520893](https://pubmed.ncbi.nlm.nih.gov/40520893/)]
16. Xu X, Chen Y, Miao J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. *J Educ Eval Health Prof*. 2024;21(6):6. [doi: [10.3352/jeehp.2024.21.6](https://doi.org/10.3352/jeehp.2024.21.6)] [Medline: [38486402](https://pubmed.ncbi.nlm.nih.gov/38486402/)]
17. Ning Y, Ong JCL, Cheng H, et al. How can artificial intelligence transform the training of medical students and physicians? *Lancet Digit Health*. Oct 2025;7(10):100900. [doi: [10.1016/j.landig.2025.100900](https://doi.org/10.1016/j.landig.2025.100900)] [Medline: [41047321](https://pubmed.ncbi.nlm.nih.gov/41047321/)]
18. Li S, Li Y, Zhou S, et al. A community-codesigned LLM-powered chatbot for primary care: a randomized controlled trial. *Nat Health*. 2026;1(2):238-250. [doi: [10.1038/s44360-025-00021-w](https://doi.org/10.1038/s44360-025-00021-w)] [Medline: [41659358](https://pubmed.ncbi.nlm.nih.gov/41659358/)]
19. O'Brien LM, Lawless M, Ambagtsheer RC. Health literacy focused communication training for primary healthcare providers working with older adults: a scoping review. *Geriatr Nurs*. 2025;66(Pt A):103601. [doi: [10.1016/j.gerinurse.2025.103601](https://doi.org/10.1016/j.gerinurse.2025.103601)] [Medline: [40829532](https://pubmed.ncbi.nlm.nih.gov/40829532/)]
20. Momand B, Dubrowski A. Addressing social context in health provider and senior communication training: what can we learn from communication accommodation theory? *Cureus*. Dec 23, 2020;12(12):e12247. [doi: [10.7759/cureus.12247](https://doi.org/10.7759/cureus.12247)] [Medline: [33505814](https://pubmed.ncbi.nlm.nih.gov/33505814/)]
21. Jechna Z. Physician's prosodic behavior and the regulation of patient anxiety: a quantitative study within the communication accommodation theory framework. *Patient Educ Couns*. Jan 2025;130(108426):108426. [doi: [10.1016/j.pec.2024.108426](https://doi.org/10.1016/j.pec.2024.108426)] [Medline: [39265453](https://pubmed.ncbi.nlm.nih.gov/39265453/)]
22. Niederdeppe J, Boyd AD, King AJ, Rimal RN. Strategies for effective public health communication in a complex information environment. *Annu Rev Public Health*. Apr 2025;46(1):411-431. [doi: [10.1146/annurev-publhealth-071723-120721](https://doi.org/10.1146/annurev-publhealth-071723-120721)] [Medline: [39656948](https://pubmed.ncbi.nlm.nih.gov/39656948/)]
23. Bylund CL, Peterson EB, Cameron KA. A practitioner's guide to interpersonal communication theory: an overview and exploration of selected theories. *Patient Educ Couns*. Jun 2012;87(3):261-267. [doi: [10.1016/j.pec.2011.10.006](https://doi.org/10.1016/j.pec.2011.10.006)] [Medline: [22112396](https://pubmed.ncbi.nlm.nih.gov/22112396/)]
24. Wang F, Wang J, Hu H, Shi W. When language heals: evaluating patient-centered communication in Chinese telemedicine through communication accommodation theory. *Digit Health*. 2025;11:20552076251411220. [doi: [10.1177/20552076251411220](https://doi.org/10.1177/20552076251411220)] [Medline: [41473837](https://pubmed.ncbi.nlm.nih.gov/41473837/)]
25. Hewett DG, Watson BM, Gallois C, Ward M, Leggett BA. Intergroup communication between hospital doctors: implications for quality of patient care. *Soc Sci Med*. Dec 2009;69(12):1732-1740. [doi: [10.1016/j.socscimed.2009.09.048](https://doi.org/10.1016/j.socscimed.2009.09.048)] [Medline: [19846246](https://pubmed.ncbi.nlm.nih.gov/19846246/)]
26. Fullwood D, Fallon E, Bolajoko O, et al. From unheard to empowered: utilizing communication accommodation theory to address Black men's with prostate cancer healthcare preferences. *Cancer Control*. 2024;31:10732748241293987. [doi: [10.1177/10732748241293987](https://doi.org/10.1177/10732748241293987)] [Medline: [39865591](https://pubmed.ncbi.nlm.nih.gov/39865591/)]
27. Hehl J, McDonald DD. Older adults' pain communication during ambulatory medical visits: an exploration of communication accommodation theory. *Pain Manag Nurs*. Jun 2014;15(2):466-473. [doi: [10.1016/j.pmn.2012.03.007](https://doi.org/10.1016/j.pmn.2012.03.007)] [Medline: [24882026](https://pubmed.ncbi.nlm.nih.gov/24882026/)]

28. Hemsley B, Balandin S, Worrall L. Nursing the patient with complex communication needs: time as a barrier and a facilitator to successful communication in hospital. *J Adv Nurs*. Jan 2012;68(1):116-126. [doi: [10.1111/j.1365-2648.2011.05722.x](https://doi.org/10.1111/j.1365-2648.2011.05722.x)] [Medline: [21831131](https://pubmed.ncbi.nlm.nih.gov/21831131/)]
29. Tao X, Zhou S, Ding K, et al. An LLM chatbot to facilitate primary-to-specialist care transitions: a randomized controlled trial. *Nat Med*. Mar 2026;32(3):934-942. [doi: [10.1038/s41591-025-04176-7](https://doi.org/10.1038/s41591-025-04176-7)] [Medline: [41555035](https://pubmed.ncbi.nlm.nih.gov/41555035/)]
30. Cui X, Gu Y, Fang H, Zhu T. Development and evaluation of LLM-based suicide intervention chatbot. *Front Psychiatry*. 2025;16(1634714):1634714. [doi: [10.3389/fpsyt.2025.1634714](https://doi.org/10.3389/fpsyt.2025.1634714)] [Medline: [40838251](https://pubmed.ncbi.nlm.nih.gov/40838251/)]
31. Raghu Subramanian C, Yang DA, Khanna R. Enhancing health care communication with large language models-the role, challenges, and future directions. *JAMA Netw Open*. Mar 4, 2024;7(3):e240347. [doi: [10.1001/jamanetworkopen.2024.0347](https://doi.org/10.1001/jamanetworkopen.2024.0347)] [Medline: [38466311](https://pubmed.ncbi.nlm.nih.gov/38466311/)]
32. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. Feb 2005;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
33. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. Oct 2, 2018;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
34. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev*. Jan 26, 2021;10(1):39. [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
35. Scoping review protocol. OSF. URL: <https://osf.io/7vf2u/files/a3p5z> [Accessed 2026-05-03]
36. Anand E, Ghersin I, Lingam G, et al. AI-generated patient-friendly MRI fistula summaries: a pilot randomised study. *J Imaging*. Sep 4, 2025;11(9):302. [doi: [10.3390/jimaging11090302](https://doi.org/10.3390/jimaging11090302)] [Medline: [41003353](https://pubmed.ncbi.nlm.nih.gov/41003353/)]
37. Berzolla E, Gosnell GG, Chen L, Vonck C, Alaia E, Meislin R. Artificial intelligence large language models improve patient comprehension of radiologist magnetic resonance imaging reports. *Arthroscopy*. Nov 2025;41(11):4607-4614. [doi: [10.1016/j.arthro.2025.04.033](https://doi.org/10.1016/j.arthro.2025.04.033)] [Medline: [40288466](https://pubmed.ncbi.nlm.nih.gov/40288466/)]
38. Bheemireddy S, Leslie SE, Durden JA, et al. The use of ChatGPT-4.0 to simplify breast pathology reports: a study on readability and accuracy. *Ann Surg Oncol*. Oct 2025;32(11):8400-8408. [doi: [10.1245/s10434-025-17860-2](https://doi.org/10.1245/s10434-025-17860-2)] [Medline: [40690168](https://pubmed.ncbi.nlm.nih.gov/40690168/)]
39. Butler JJ, Harrington MC, Tong Y, et al. From jargon to clarity: improving the readability of foot and ankle radiology reports with an artificial intelligence large language model. *Foot Ankle Surg*. Jun 2024;30(4):331-337. [doi: [10.1016/j.fas.2024.01.008](https://doi.org/10.1016/j.fas.2024.01.008)] [Medline: [38336501](https://pubmed.ncbi.nlm.nih.gov/38336501/)]
40. Chen AH, Rudin RS, Levine DM, Mehrotra A. Improving patient understanding of radiology reports using generative artificial intelligence: a vignette study of 2000 US adults. *J Am Med Inform Assoc*. Feb 1, 2026;33(2):326-335. [doi: [10.1093/jamia/ocaf187](https://doi.org/10.1093/jamia/ocaf187)] [Medline: [41241949](https://pubmed.ncbi.nlm.nih.gov/41241949/)]
41. Chung EM, Zhang SC, Nguyen AT, Atkins KM, Sandler HM, Kamrava M. Feasibility and acceptability of ChatGPT generated radiology report summaries for cancer patients. *Digit Health*. 2023;9(2):20552076231221620. [doi: [10.1177/20552076231221620](https://doi.org/10.1177/20552076231221620)] [Medline: [38130802](https://pubmed.ncbi.nlm.nih.gov/38130802/)]
42. Eisinger F, Holderried F, Mahling M, et al. What's going on with me and how can I better manage my health? The potential of GPT-4 to transform discharge letters into patient-centered letters to enhance patient safety: prospective, exploratory study. *J Med Internet Res*. Jan 21, 2025;27:e67143. [doi: [10.2196/67143](https://doi.org/10.2196/67143)] [Medline: [39836954](https://pubmed.ncbi.nlm.nih.gov/39836954/)]
43. Hains L, Kleinig O, Murugappa A, et al. Large language model discharge summary preparation using real-world electronic medical record data shows promise. *Intern Med J*. May 28, 2025;55(7):1188-1192. [doi: [10.1111/imj.70073](https://doi.org/10.1111/imj.70073)] [Medline: [40434141](https://pubmed.ncbi.nlm.nih.gov/40434141/)]
44. Khanmammadova N, Gevorkyan R, Afyouni AS, et al. Evaluation of prostate cancer pathology reports generated by ChatGPT to enhance patient comprehension. *Sci Rep*. Oct 2, 2025;15(1):34459. [doi: [10.1038/s41598-025-17639-1](https://doi.org/10.1038/s41598-025-17639-1)] [Medline: [41038938](https://pubmed.ncbi.nlm.nih.gov/41038938/)]
45. Kuckelman IJ, Wetley K, Yi PH, Ross AB. Translating musculoskeletal radiology reports into patient-friendly summaries using ChatGPT-4. *Skeletal Radiol*. Aug 2024;53(8):1621-1624. [doi: [10.1007/s00256-024-04599-2](https://doi.org/10.1007/s00256-024-04599-2)] [Medline: [38270616](https://pubmed.ncbi.nlm.nih.gov/38270616/)]
46. Li H, Moon JT, Iyer D, et al. Decoding radiology reports: potential application of OpenAI ChatGPT to enhance patient understanding of diagnostic reports. *Clin Imaging*. Sep 2023;101:137-141. [doi: [10.1016/j.clinimag.2023.06.008](https://doi.org/10.1016/j.clinimag.2023.06.008)] [Medline: [37336169](https://pubmed.ncbi.nlm.nih.gov/37336169/)]
47. Li Y, Li F, Hong N, et al. A comparative study of recent large language models on generating hospital discharge summaries for lung cancer patients. *J Biomed Inform*. Aug 2025;168:104867. [doi: [10.1016/j.jbi.2025.104867](https://doi.org/10.1016/j.jbi.2025.104867)] [Medline: [40544901](https://pubmed.ncbi.nlm.nih.gov/40544901/)]
48. Prucker P, Busch F, Dorfner F, et al. Performance of open-source and proprietary large language models in generating patient-friendly radiology chest CT reports. *Clin Imaging*. Sep 2025;125:110557. [doi: [10.1016/j.clinimag.2025.110557](https://doi.org/10.1016/j.clinimag.2025.110557)] [Medline: [40639135](https://pubmed.ncbi.nlm.nih.gov/40639135/)]

49. Sarangi PK, Lumbani A, Swarup MS, et al. Assessing ChatGPT's proficiency in simplifying radiological reports for healthcare professionals and patients. *Cureus*. Dec 2023;15(12):e50881. [doi: [10.7759/cureus.50881](https://doi.org/10.7759/cureus.50881)] [Medline: [38249202](https://pubmed.ncbi.nlm.nih.gov/38249202/)]
50. 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*. Feb 2024;144(2):611-618. [doi: [10.1007/s00402-023-05113-4](https://doi.org/10.1007/s00402-023-05113-4)] [Medline: [37950763](https://pubmed.ncbi.nlm.nih.gov/37950763/)]
51. Steimetz E, Minkowitz J, Gabutan EC, et al. Use of artificial intelligence chatbots in interpretation of pathology reports. *JAMA Netw Open*. May 1, 2024;7(5):e2412767. [doi: [10.1001/jamanetworkopen.2024.12767](https://doi.org/10.1001/jamanetworkopen.2024.12767)] [Medline: [38776080](https://pubmed.ncbi.nlm.nih.gov/38776080/)]
52. Stephan D, Bertsch AS, Schumacher S, et al. Improving patient communication by simplifying AI-generated dental radiology reports with ChatGPT: comparative study. *J Med Internet Res*. Jun 9, 2025;27:e73337. [doi: [10.2196/73337](https://doi.org/10.2196/73337)] [Medline: [40489773](https://pubmed.ncbi.nlm.nih.gov/40489773/)]
53. Sunshine A, Honce GH, Callen AL, et al. Evaluating the quality and understandability of radiology report summaries generated by ChatGPT: survey study. *JMIR Form Res*. Aug 27, 2025;9:e76097. [doi: [10.2196/76097](https://doi.org/10.2196/76097)] [Medline: [40864905](https://pubmed.ncbi.nlm.nih.gov/40864905/)]
54. Tang CC, Nagesh S, Fussell DA, et al. Generating colloquial radiology reports with large language models. *J Am Med Inform Assoc*. Nov 1, 2024;31(11):2660-2667. [doi: [10.1093/jamia/ocae223](https://doi.org/10.1093/jamia/ocae223)] [Medline: [39178375](https://pubmed.ncbi.nlm.nih.gov/39178375/)]
55. Williams CYK, Subramanian CR, Ali SS, et al. Physician- and large language model-generated hospital discharge summaries. *JAMA Intern Med*. Jul 1, 2025;185(7):818-825. [doi: [10.1001/jamainternmed.2025.0821](https://doi.org/10.1001/jamainternmed.2025.0821)] [Medline: [40323616](https://pubmed.ncbi.nlm.nih.gov/40323616/)]
56. Yang X, Xiao Y, Liu D, et al. Enhancing physician-patient communication in oncology using GPT-4 through simplified radiology reports: multicenter quantitative study. *J Med Internet Res*. Apr 17, 2025;27:e63786. [doi: [10.2196/63786](https://doi.org/10.2196/63786)] [Medline: [40245397](https://pubmed.ncbi.nlm.nih.gov/40245397/)]
57. Yang X, Xiao Y, Liu D, et al. Enhancing doctor-patient communication using large language models for pathology report interpretation. *BMC Med Inform Decis Mak*. Jan 23, 2025;25(1):36. [doi: [10.1186/s12911-024-02838-z](https://doi.org/10.1186/s12911-024-02838-z)] [Medline: [39849504](https://pubmed.ncbi.nlm.nih.gov/39849504/)]
58. Zaretsky J, Kim JM, Baskharoun S, et al. Generative artificial intelligence to transform inpatient discharge summaries to patient-friendly language and format. *JAMA Netw Open*. Mar 4, 2024;7(3):e240357. [doi: [10.1001/jamanetworkopen.2024.0357](https://doi.org/10.1001/jamanetworkopen.2024.0357)] [Medline: [38466307](https://pubmed.ncbi.nlm.nih.gov/38466307/)]
59. Vallurupalli M, Shah ND, Vyas RM. Optimizing readability of patient-facing hand surgery education materials using chat generative pretrained transformer (ChatGPT) 3.5. *J Hand Surg Am*. Oct 2024;49(10):986-991. [doi: [10.1016/j.jhsa.2024.05.007](https://doi.org/10.1016/j.jhsa.2024.05.007)] [Medline: [38970600](https://pubmed.ncbi.nlm.nih.gov/38970600/)]
60. Chandra K, Ghilzai U, Lawand J, Ghali A, Fiedler B, Ahmed AS. Improving readability of shoulder and elbow surgery online patient education material with Chat GPT (Chat Generative Pretrained Transformer) 4. *J Shoulder Elbow Surg*. Nov 2025;34(11):e1119-e1124. [doi: [10.1016/j.jse.2025.02.025](https://doi.org/10.1016/j.jse.2025.02.025)] [Medline: [40118438](https://pubmed.ncbi.nlm.nih.gov/40118438/)]
61. Gondode PG, Khanna P, Sharma P, Duggal S, Garg N. End-of-life care patient information leaflets-a comparative evaluation of artificial intelligence-generated content for readability, sentiment, accuracy, completeness, and suitability: ChatGPT vs Google Gemini. *Indian J Crit Care Med*. Jun 2024;28(6):561-568. [doi: [10.5005/jp-journals-10071-24725](https://doi.org/10.5005/jp-journals-10071-24725)] [Medline: [39130387](https://pubmed.ncbi.nlm.nih.gov/39130387/)]
62. Kianian R, Carter M, Finkelshtein I, Eleswarapu SV, Kachroo N. Application of artificial intelligence to patient-targeted health information on kidney stone disease. *J Ren Nutr*. Mar 2024;34(2):170-176. [doi: [10.1053/j.jrn.2023.10.002](https://doi.org/10.1053/j.jrn.2023.10.002)] [Medline: [37839591](https://pubmed.ncbi.nlm.nih.gov/37839591/)]
63. Juan-Guardela EJJ, Beltrán-España JA, Ravagli-Baquero MP, et al. Reliability and readability assessment of atrial fibrillation patient information delivered by artificial intelligence-based language models (ChatGPT, YouChat, Gemini, and Perplexity AI) in English and Spanish. *Clin Med Insights Cardiol*. 2025;19:11795468251383666. [doi: [10.1177/11795468251383666](https://doi.org/10.1177/11795468251383666)] [Medline: [41180819](https://pubmed.ncbi.nlm.nih.gov/41180819/)]
64. Dias R, Castan A, Gotoff K, et al. ChatGPT 3.5 better improves comprehensibility of English, than Spanish, generated responses to osteosarcoma questions. *J Surg Oncol*. Jun 2025;131(8):1692-1695. [doi: [10.1002/jso.28109](https://doi.org/10.1002/jso.28109)] [Medline: [39898783](https://pubmed.ncbi.nlm.nih.gov/39898783/)]
65. Gulati V, Roy SG, Moawad A, et al. Transcending language barriers: can ChatGPT be the key to enhancing multilingual accessibility in health care? *J Am Coll Radiol*. Dec 2024;21(12):1888-1895. [doi: [10.1016/j.jacr.2024.05.009](https://doi.org/10.1016/j.jacr.2024.05.009)] [Medline: [38880289](https://pubmed.ncbi.nlm.nih.gov/38880289/)]
66. Chandler RD, Warner S, Guillaume D, Wells J. Exploring ChatGPT's role in women's health self-education: a descriptive study comparing responses with public health guidance. *Nurs Outlook*. 2025;73(4):102468. [doi: [10.1016/j.outlook.2025.102468](https://doi.org/10.1016/j.outlook.2025.102468)] [Medline: [40614295](https://pubmed.ncbi.nlm.nih.gov/40614295/)]

67. Pap IA, Oniga S. eHealth assistant AI chatbot using a large language model to provide personalized answers through secure decentralized communication. *Sensors (Basel)*. Sep 23, 2024;24(18):6140. [doi: [10.3390/s24186140](https://doi.org/10.3390/s24186140)] [Medline: [39338885](https://pubmed.ncbi.nlm.nih.gov/39338885/)]
68. Akçay O, Öztürk Ö, Acar T, Gürsoy S. Accuracy and reliability of ChatGPT in answering patient questions about lung cancer and its surgery: an expert panel evaluation by thoracic surgeons. *J Cancer Educ*. Jul 4, 2025. [doi: [10.1007/s13187-025-02682-3](https://doi.org/10.1007/s13187-025-02682-3)] [Medline: [40615772](https://pubmed.ncbi.nlm.nih.gov/40615772/)]
69. Belge Bilgin G, Bilgin C, Childs DS, et al. Performance of ChatGPT-4 and Bard chatbots in responding to common patient questions on prostate cancer <sup>177</sup>Lu-PSMA-617 therapy. *Front Oncol*. 2024;14:1386718. [doi: [10.3389/fonc.2024.1386718](https://doi.org/10.3389/fonc.2024.1386718)] [Medline: [39070149](https://pubmed.ncbi.nlm.nih.gov/39070149/)]
70. Demir Cicek B, Cicek O. Evaluating the response of AI-based large language models to common patient concerns about endodontic root canal treatment: a comparative performance analysis. *J Clin Med*. Oct 22, 2025;14(21):7482. [doi: [10.3390/jcm14217482](https://doi.org/10.3390/jcm14217482)] [Medline: [41226878](https://pubmed.ncbi.nlm.nih.gov/41226878/)]
71. Doğan L, Yılmaz İE. The performance of ChatGPT-4 and Bing Chat in frequently asked questions about glaucoma. *Eur J Ophthalmol*. Jul 2025;35(4):1323-1328. [doi: [10.1177/11206721251321197](https://doi.org/10.1177/11206721251321197)] [Medline: [39973162](https://pubmed.ncbi.nlm.nih.gov/39973162/)]
72. Gencer B, Arzu U, Orhan SS, Dinçal T, Ekinçi M. Evaluation of ChatGPT responses about sexual activity after total hip arthroplasty: a comparative study with observers of different experience levels. *J Clin Med*. Apr 24, 2025;14(9):2942. [doi: [10.3390/jcm14092942](https://doi.org/10.3390/jcm14092942)] [Medline: [40363974](https://pubmed.ncbi.nlm.nih.gov/40363974/)]
73. Goshtasbi K, Best C, Powers B, et al. Comparative performance of the leading large language models in answering complex rhinoplasty consultation questions. *Facial Plast Surg Aesthet Med*. 2025;27(4):378-383. [doi: [10.1089/fpsam.2024.0206](https://doi.org/10.1089/fpsam.2024.0206)] [Medline: [39812574](https://pubmed.ncbi.nlm.nih.gov/39812574/)]
74. Gültekin O, Inoue J, Yılmaz B, et al. Evaluating DeepResearch and DeepThink in anterior cruciate ligament surgery patient education: ChatGPT-4o excels in comprehensiveness, DeepSeek R1 leads in clarity and readability of orthopaedic information. *Knee Surg Sports Traumatol Arthrosc*. Aug 2025;33(8):3025-3031. [doi: [10.1002/ksa.12711](https://doi.org/10.1002/ksa.12711)] [Medline: [40450565](https://pubmed.ncbi.nlm.nih.gov/40450565/)]
75. Guven Y, Ozdemir OT, Kavan MY. Performance of artificial intelligence chatbots in responding to patient queries related to traumatic dental injuries: a comparative study. *Dent Traumatol*. Jun 2025;41(3):338-347. [doi: [10.1111/edt.13020](https://doi.org/10.1111/edt.13020)] [Medline: [39578674](https://pubmed.ncbi.nlm.nih.gov/39578674/)]
76. Ibas M, Dursun S, Paksoy M, Ocal R, Karatas E. Accuracy and safety of ChatGPT-4o responses in rhinoplasty postoperative counseling: a panel-based study. *Acta Otolaryngol*. Sep 2025;145(9):851-856. [doi: [10.1080/00016489.2025.2541612](https://doi.org/10.1080/00016489.2025.2541612)] [Medline: [40772566](https://pubmed.ncbi.nlm.nih.gov/40772566/)]
77. Kamminga NCW, Kievits JEC, Plaisier PW, et al. Do large language model chatbots perform better than established patient information resources in answering patient questions? A comparative study on melanoma. *Br J Dermatol*. Jan 24, 2025;192(2):306-315. [doi: [10.1093/bjd/ljae377](https://doi.org/10.1093/bjd/ljae377)] [Medline: [39365602](https://pubmed.ncbi.nlm.nih.gov/39365602/)]
78. Kuo FH, Fierstein JL, Tudor BH, et al. Comparing ChatGPT and a single anesthesiologist's responses to common patient questions: an exploratory cross-sectional survey of a panel of anesthesiologists. *J Med Syst*. Aug 22, 2024;48(1):77. [doi: [10.1007/s10916-024-02100-z](https://doi.org/10.1007/s10916-024-02100-z)] [Medline: [39172169](https://pubmed.ncbi.nlm.nih.gov/39172169/)]
79. Lewandowski M, Łukowicz P, Świetlik D, Jendrysik B, Barańska-Rybak W. Comparing physician and artificial intelligence ChatGPT-4 responses to common patient questions regarding hidradenitis suppurativa: a single-blind study. *Postepy Dermatol Alergol*. 2025;42(5):460-466. [doi: [10.5114/ada.2025.151910](https://doi.org/10.5114/ada.2025.151910)] [Medline: [41262860](https://pubmed.ncbi.nlm.nih.gov/41262860/)]
80. Li AW, Adelstein JM, Li LT, Sinkler MA, Mistovich RJ. Assessing ChatGPT responses to frequently asked questions regarding pediatric supracondylar humerus fractures. *J Pediatr Orthop*. Jul 1, 2025;45(6):327-331. [doi: [10.1097/BPO.0000000000002923](https://doi.org/10.1097/BPO.0000000000002923)] [Medline: [39917871](https://pubmed.ncbi.nlm.nih.gov/39917871/)]
81. Guo S, Li G, Gou J, et al. The performance of ChatGPT-4.0 and ChatGPT-4omni on answering thyroid question: a multicenter study. *J Surg Res*. Sep 2025;313:500-508. [doi: [10.1016/j.jss.2025.06.066](https://doi.org/10.1016/j.jss.2025.06.066)] [Medline: [40706228](https://pubmed.ncbi.nlm.nih.gov/40706228/)]
82. Mete U, Özmen ÖA. Assessing the accuracy and reproducibility of ChatGPT for responding to patient inquiries about otosclerosis. *Eur Arch Otorhinolaryngol*. Mar 2025;282(3):1567-1575. [doi: [10.1007/s00405-024-09039-4](https://doi.org/10.1007/s00405-024-09039-4)] [Medline: [39461921](https://pubmed.ncbi.nlm.nih.gov/39461921/)]
83. Moothedan E, Jhumkhawala V, Burgoa S, Martinez L, Sacca L. Response to "Qualitatively Assessing ChatGPT Responses to Frequently Asked Questions Regarding Sexually Transmitted Diseases: Considerations". *Sex Transm Dis*. Apr 1, 2025;52(4):e11-e11. [doi: [10.1097/OLQ.0000000000002121](https://doi.org/10.1097/OLQ.0000000000002121)] [Medline: [39681087](https://pubmed.ncbi.nlm.nih.gov/39681087/)]
84. Motegi M, Shino M, Kuwabara M, et al. Comparison of physician and large language model chatbot responses to online ear, nose, and throat inquiries. *Sci Rep*. Jul 1, 2025;15(1):21346. [doi: [10.1038/s41598-025-06769-1](https://doi.org/10.1038/s41598-025-06769-1)] [Medline: [40596359](https://pubmed.ncbi.nlm.nih.gov/40596359/)]
85. Özcivelek T, Özcan B. Comparative evaluation of responses from DeepSeek-R1, ChatGPT-o1, ChatGPT-4, and dental GPT chatbots to patient inquiries about dental and maxillofacial prostheses. *BMC Oral Health*. May 31, 2025;25(1):871. [doi: [10.1186/s12903-025-06267-w](https://doi.org/10.1186/s12903-025-06267-w)] [Medline: [40450291](https://pubmed.ncbi.nlm.nih.gov/40450291/)]

86. Raghunathan R, Jacobs AR, Sant VR, et al. Can large language models address unmet patient information needs and reduce provider burnout in the management of thyroid disease? *Surgery*. Jan 2025;177:108859. [doi: [10.1016/j.surg.2024.06.075](https://doi.org/10.1016/j.surg.2024.06.075)] [Medline: [39424485](https://pubmed.ncbi.nlm.nih.gov/39424485/)]
87. Rotem R, Simon C, Rottenstreich M, O'reilly BA, Weintraub AY, O'Sullivan OE. ChatGPT in urogynecology: comparing large language model responses to human experts. *Acta Obstet Gynecol Scand*. Oct 23, 2025. [doi: [10.1111/aogs.70085](https://doi.org/10.1111/aogs.70085)] [Medline: [41128396](https://pubmed.ncbi.nlm.nih.gov/41128396/)]
88. Roy JM, Atallah E, Piper K, et al. Comparison of quality, empathy and readability of physician responses versus chatbot responses to common cerebrovascular neurosurgical questions on a social media platform. *Clin Neurol Neurosurg*. Aug 2025;255:108986. [doi: [10.1016/j.clineuro.2025.108986](https://doi.org/10.1016/j.clineuro.2025.108986)] [Medline: [40451125](https://pubmed.ncbi.nlm.nih.gov/40451125/)]
89. Şan H, Bayrakçı Ö, Çağdaş B, Serdengeçti M, Alagöz E. Reliability and readability analysis of ChatGPT-4 and Google Bard as a patient information source for the most commonly applied radionuclide treatments in cancer patients. *Rev Esp Med Nucl Imagen Mol (Engl Ed)*. 2024;43(4):500021. [doi: [10.1016/j.remnie.2024.500021](https://doi.org/10.1016/j.remnie.2024.500021)] [Medline: [38821410](https://pubmed.ncbi.nlm.nih.gov/38821410/)]
90. Scaff SPS, Reis FJJ, Ferreira GE, Jacob MF, Saragiotto BT. Assessing the performance of AI chatbots in answering patients' common questions about low back pain. *Ann Rheum Dis*. Jan 2025;84(1):143-149. [doi: [10.1136/ard-2024-226202](https://doi.org/10.1136/ard-2024-226202)] [Medline: [39874229](https://pubmed.ncbi.nlm.nih.gov/39874229/)]
91. Shao XR, Ruan T, Ju XG, Sun YH, Cui JC. Evaluating artificial intelligence chatbots' responses to gynecomastia inquiries: comparative study of information quality, readability, and guideline consistency. *Digit Health*. 2025;11:20552076251367645. [doi: [10.1177/20552076251367645](https://doi.org/10.1177/20552076251367645)] [Medline: [40881062](https://pubmed.ncbi.nlm.nih.gov/40881062/)]
92. Tuzlalı M, Baki N, Aral K, Aral CA, Bahçe E. Evaluating the performance of AI chatbots in responding to dental implant FAQs: a comparative study. *BMC Oral Health*. Oct 8, 2025;25(1):1548. [doi: [10.1186/s12903-025-06863-w](https://doi.org/10.1186/s12903-025-06863-w)] [Medline: [41063105](https://pubmed.ncbi.nlm.nih.gov/41063105/)]
93. Weber MT, Noll R, Marchl A, et al. MedBot vs RealDoc: efficacy of large language modeling in physician-patient communication for rare diseases. *J Am Med Inform Assoc*. May 1, 2025;32(5):775-783. [doi: [10.1093/jamia/ocaf034](https://doi.org/10.1093/jamia/ocaf034)] [Medline: [39998911](https://pubmed.ncbi.nlm.nih.gov/39998911/)]
94. White CA, Masturov YA, Haunschild E, Michaelson E, Shukla DR, Cagle PJ. Can ChatGPT reliably answer the most common patient questions regarding total shoulder arthroplasty? *J Shoulder Elbow Surg*. May 2025;34(5):e254-e264. [doi: [10.1016/j.jse.2024.08.025](https://doi.org/10.1016/j.jse.2024.08.025)] [Medline: [39419373](https://pubmed.ncbi.nlm.nih.gov/39419373/)]
95. Xue D, Liao S. Evaluation of three large language models' response performances to inquiries regarding post-abortion care in the context of Chinese language: a comparative analysis. *Risk Manag Healthc Policy*. 2025;18:2731-2741. [doi: [10.2147/RMHP.S531777](https://doi.org/10.2147/RMHP.S531777)] [Medline: [40862290](https://pubmed.ncbi.nlm.nih.gov/40862290/)]
96. Zhang S, Chua IJ, Liang ZC, Pang KC. Can ChatGPT provide personalised responses for frequently asked questions regarding anterior cruciate ligament reconstruction? *Knee*. Oct 2025;56:495-502. [doi: [10.1016/j.knee.2025.06.018](https://doi.org/10.1016/j.knee.2025.06.018)] [Medline: [40680504](https://pubmed.ncbi.nlm.nih.gov/40680504/)]
97. He Z, Liu Y, Wang Z, et al. Revolutionizing inflammatory bowel disease healthcare communication: a head-to-head comparison of gastroenterologist and ChatGPT responses. *Gastroenterol Rep (Oxf)*. 2025;13:goaf087. [doi: [10.1093/gastro/goaf087](https://doi.org/10.1093/gastro/goaf087)] [Medline: [41035903](https://pubmed.ncbi.nlm.nih.gov/41035903/)]
98. Wu H, Li M, Zhang L. Comparing physician and large language model responses to influenza patient questions in the online health community. *Int J Med Inform*. May 2025;197:105836. [doi: [10.1016/j.ijmedinf.2025.105836](https://doi.org/10.1016/j.ijmedinf.2025.105836)] [Medline: [39986122](https://pubmed.ncbi.nlm.nih.gov/39986122/)]
99. Muasher-Kerwin C, Hughes MC, Foster ML, Al Azher I, Alhoori H. Exploring large language models for summarizing and interpreting an online brain tumor support forum. *Digit Health*. 2025;11:20552076251337345. [doi: [10.1177/20552076251337345](https://doi.org/10.1177/20552076251337345)] [Medline: [40297356](https://pubmed.ncbi.nlm.nih.gov/40297356/)]
100. Bertges DJ, Beck AW, Schermerhorn M, et al. Testing ChatGPT's ability to provide patient and physician information on aortic aneurysm. *J Surg Res*. Mar 2025;307:129-138. [doi: [10.1016/j.jss.2025.01.015](https://doi.org/10.1016/j.jss.2025.01.015)] [Medline: [40014909](https://pubmed.ncbi.nlm.nih.gov/40014909/)]
101. Allen JW, Levy N, Wilkinson D. Empowering patient autonomy: the role of large language models (LLMs) in scaffolding informed consent in medical practice. *Bioethics*. Feb 2026;40(2):183-193. [doi: [10.1111/bioe.70030](https://doi.org/10.1111/bioe.70030)] [Medline: [40931965](https://pubmed.ncbi.nlm.nih.gov/40931965/)]
102. Grünebaum A, Dudenhausen J, Chervenak FA. Enhancing patient understanding in obstetrics: the role of generative AI in simplifying informed consent for labor induction with oxytocin. *J Perinat Med*. Jul 28, 2025;53(6):688-695. [doi: [10.1515/jpm-2024-0428](https://doi.org/10.1515/jpm-2024-0428)] [Medline: [39470098](https://pubmed.ncbi.nlm.nih.gov/39470098/)]
103. Kienzle A, Niemann M, Meller S, Gwinner C. ChatGPT may offer an adequate substitute for informed consent to patients prior to total knee arthroplasty-yet caution is needed. *J Pers Med*. Jan 5, 2024;14(1):69. [doi: [10.3390/jpm14010069](https://doi.org/10.3390/jpm14010069)] [Medline: [38248771](https://pubmed.ncbi.nlm.nih.gov/38248771/)]
104. Bakhaya S, Lehnbohm EC, de Carvalho Filho MA, Ma KY, Svensberg K. Development and evaluation of AI chatbot tool for written communication training in self-care: experiences of pharmacy students and faculty. *Curr Pharm Teach Learn*. Jan 2026;18(1):102503. [doi: [10.1016/j.cptl.2025.102503](https://doi.org/10.1016/j.cptl.2025.102503)] [Medline: [41061515](https://pubmed.ncbi.nlm.nih.gov/41061515/)]

105. Chen PJ. Effectiveness of integrating generative artificial intelligence with virtual reality for maternity communication simulation: a randomized controlled trial. *Clin Simul Nurs*. Aug 2025;105:101786. [doi: [10.1016/j.ecns.2025.101786](https://doi.org/10.1016/j.ecns.2025.101786)]
106. Chen PJ, Liou WK. ChatGPT-driven interactive virtual reality communication simulation in obstetric nursing: a mixed-methods study. *Nurse Educ Pract*. May 2025;85:104383. [doi: [10.1016/j.nepr.2025.104383](https://doi.org/10.1016/j.nepr.2025.104383)] [Medline: [40315616](https://pubmed.ncbi.nlm.nih.gov/40315616/)]
107. Gutiérrez Maquilón R, Uhl J, Schrom-Feiertag H, Tscheligi M. Integrating GPT-based AI into virtual patients to facilitate communication training among medical first responders: usability study of mixed reality simulation. *JMIR Form Res*. Dec 11, 2024;8:e58623. [doi: [10.2196/58623](https://doi.org/10.2196/58623)] [Medline: [39661979](https://pubmed.ncbi.nlm.nih.gov/39661979/)]
108. Kim J, Won J, Lee Y. Use of a generative pre-trained transformer-based virtual patient for health assessment and communication training in nursing education: a mixed-methods study. *Nurse Educ Pract*. Oct 2025;88:104536. [doi: [10.1016/j.nepr.2025.104536](https://doi.org/10.1016/j.nepr.2025.104536)] [Medline: [40902283](https://pubmed.ncbi.nlm.nih.gov/40902283/)]
109. Wang Z, Fan TT, Li ML, Zhu NJ, Wang XC. Feasibility study of using GPT for history-taking training in medical education: a randomized clinical trial. *BMC Med Educ*. Jul 10, 2025;25(1):1030. [doi: [10.1186/s12909-025-07614-9](https://doi.org/10.1186/s12909-025-07614-9)] [Medline: [40640776](https://pubmed.ncbi.nlm.nih.gov/40640776/)]
110. Weisman D, Sugarman A, Huang YM, Gelberg L, Ganz PA, Comulada WS. Development of a GPT-4-powered virtual simulated patient and communication training platform for medical students to practice discussing abnormal mammogram results with patients: multiphase study. *JMIR Form Res*. Apr 17, 2025;9:e65670. [doi: [10.2196/65670](https://doi.org/10.2196/65670)] [Medline: [40246299](https://pubmed.ncbi.nlm.nih.gov/40246299/)]
111. Chen H, Zeng W, Chen C, et al. EEG emotion copilot: optimizing lightweight LLMs for emotional EEG interpretation with assisted medical record generation. *Neural Netw*. Dec 2025;192:107848. [doi: [10.1016/j.neunet.2025.107848](https://doi.org/10.1016/j.neunet.2025.107848)] [Medline: [40683189](https://pubmed.ncbi.nlm.nih.gov/40683189/)]
112. Adikari A, Alahakoon D, Pallewela N, Pierce JE, Hernandez NJ, Rose ML. Reconstructing impaired language using generative AI for people with aphasia. *Sci Rep*. Nov 19, 2025;15(1):40877. [doi: [10.1038/s41598-025-24725-x](https://doi.org/10.1038/s41598-025-24725-x)] [Medline: [41258473](https://pubmed.ncbi.nlm.nih.gov/41258473/)]
113. Cai S, Venugopalan S, Seaver K, et al. Using large language models to accelerate communication for eye gaze typing users with ALS. *Nat Commun*. Nov 1, 2024;15(1):9449. [doi: [10.1038/s41467-024-53873-3](https://doi.org/10.1038/s41467-024-53873-3)] [Medline: [39487163](https://pubmed.ncbi.nlm.nih.gov/39487163/)]
114. Wan P, Huang Z, Tang W, et al. Outpatient reception via collaboration between nurses and a large language model: a randomized controlled trial. *Nat Med*. Oct 2024;30(10):2878-2885. [doi: [10.1038/s41591-024-03148-7](https://doi.org/10.1038/s41591-024-03148-7)] [Medline: [39009780](https://pubmed.ncbi.nlm.nih.gov/39009780/)]
115. Carl N, Haggemüller S, Wies C, et al. Evaluating interactions of patients with large language models for medical information. *BJU Int*. Jun 2025;135(6):1010-1017. [doi: [10.1111/bju.16676](https://doi.org/10.1111/bju.16676)] [Medline: [39967059](https://pubmed.ncbi.nlm.nih.gov/39967059/)]
116. Chung D, Dhillon H, Sidhom K, Bal D, Patel P, Jawanda G. (096) Real world utility of chatgpt in pre-vasectomy counselling in an office-based setting: a pilot study. *J Sex Med*. Aug 12, 2024;21(7). [doi: [10.1093/jsxmed/qdae167.094](https://doi.org/10.1093/jsxmed/qdae167.094)]
117. Park C, An MH, Hwang G, Park RW, An J. Clinical performance and communication skills of ChatGPT versus physicians in emergency medicine: simulated patient study. *JMIR Med Inform*. Jul 17, 2025;13:e68409. [doi: [10.2196/68409](https://doi.org/10.2196/68409)] [Medline: [40674718](https://pubmed.ncbi.nlm.nih.gov/40674718/)]
118. Santonocito S, Ciccì M, Ronsivalle V. Evaluation of the impact of AI-based chatbot on orthodontic patient education: a preliminary randomised controlled trial. *Clin Oral Investig*. Apr 30, 2025;29(5):278. [doi: [10.1007/s00784-025-06356-8](https://doi.org/10.1007/s00784-025-06356-8)] [Medline: [40304793](https://pubmed.ncbi.nlm.nih.gov/40304793/)]
119. Scholich T, Barr M, Wiltsey Stirman S, Raj S. A comparison of responses from human therapists and large language model-based chatbots to assess therapeutic communication: mixed methods study. *JMIR Ment Health*. May 21, 2025;12:e69709. [doi: [10.2196/69709](https://doi.org/10.2196/69709)] [Medline: [40397927](https://pubmed.ncbi.nlm.nih.gov/40397927/)]
120. Warin K, Taetrago U, Trachoo V, et al. AI chatbot for post-operative oral surgery information and support. *Stud Health Technol Inform*. Aug 7, 2025;329:1938-1939. [doi: [10.3233/SHTI251288](https://doi.org/10.3233/SHTI251288)] [Medline: [40776305](https://pubmed.ncbi.nlm.nih.gov/40776305/)]
121. Chen C, Hu Y, Cai W, et al. Deep learning-based in-ambulance speech recognition and generation of prehospital emergency diagnostic summaries using LLMs. *Int J Med Inform*. Nov 2025;203:106029. [doi: [10.1016/j.ijmedinf.2025.106029](https://doi.org/10.1016/j.ijmedinf.2025.106029)] [Medline: [40639122](https://pubmed.ncbi.nlm.nih.gov/40639122/)]
122. Genes N, Simon G, Koziatek C, et al. Generative artificial intelligence summaries to facilitate emergency department handoff. *Appl Clin Inform*. Aug 2025;16(4):1185-1191. [doi: [10.1055/a-2681-5008](https://doi.org/10.1055/a-2681-5008)] [Medline: [40795949](https://pubmed.ncbi.nlm.nih.gov/40795949/)]
123. Tailor PD, D'Souza HS, Castillejo Becerra CM, et al. Evaluation of AI summaries on interdisciplinary understanding of ophthalmology notes. *JAMA Ophthalmol*. May 1, 2025;143(5):410-419. [doi: [10.1001/jamaophthalmol.2025.0351](https://doi.org/10.1001/jamaophthalmol.2025.0351)] [Medline: [40178837](https://pubmed.ncbi.nlm.nih.gov/40178837/)]
124. Garcia P, Ma SP, Shah S, et al. Artificial intelligence-generated draft replies to patient inbox messages. *JAMA Netw Open*. Mar 4, 2024;7(3):e243201. [doi: [10.1001/jamanetworkopen.2024.3201](https://doi.org/10.1001/jamanetworkopen.2024.3201)] [Medline: [38506805](https://pubmed.ncbi.nlm.nih.gov/38506805/)]
125. Kaur A, Budko A, Liu K, Eaton E, Steitz BD, Johnson KB. Automating responses to patient portal messages using generative AI. *Appl Clin Inform*. May 2025;16(3):718-731. [doi: [10.1055/a-2565-9155](https://doi.org/10.1055/a-2565-9155)] [Medline: [40132987](https://pubmed.ncbi.nlm.nih.gov/40132987/)]
126. Liu S, McCoy AB, Wright AP, et al. Leveraging large language models for generating responses to patient messages—a subjective analysis. *J Am Med Inform Assoc*. May 20, 2024;31(6):1367-1379. [doi: [10.1093/jamia/ocae052](https://doi.org/10.1093/jamia/ocae052)] [Medline: [38497958](https://pubmed.ncbi.nlm.nih.gov/38497958/)]

127. Scott M, Muncey W, Seranio N, et al. Assessing artificial intelligence-generated responses to urology patient in-basket messages. *Urol Pract*. Sep 2024;11(5):793-798. [doi: [10.1097/UPJ.0000000000000637](https://doi.org/10.1097/UPJ.0000000000000637)] [Medline: [39162591](https://pubmed.ncbi.nlm.nih.gov/39162591/)]
128. Small WR, Wiesenfeld B, Brandfield-Harvey B, et al. Large language model-based responses to patients' in-basket messages. *JAMA Netw Open*. Jul 1, 2024;7(7):e2422399. [doi: [10.1001/jamanetworkopen.2024.22399](https://doi.org/10.1001/jamanetworkopen.2024.22399)] [Medline: [39012633](https://pubmed.ncbi.nlm.nih.gov/39012633/)]
129. Soroudi D, Gozali A, Knox JA, et al. Comparing provider and ChatGPT responses to breast reconstruction patient questions in the electronic health record. *Ann Plast Surg*. Nov 1, 2024;93(5):541-545. [doi: [10.1097/SAP.0000000000004090](https://doi.org/10.1097/SAP.0000000000004090)] [Medline: [39445873](https://pubmed.ncbi.nlm.nih.gov/39445873/)]
130. English E, Laughlin J, Sippel J, DeCamp M, Lin CT. Utility of artificial intelligence-generative draft replies to patient messages. *JAMA Netw Open*. Oct 1, 2024;7(10):e2438573. [doi: [10.1001/jamanetworkopen.2024.38573](https://doi.org/10.1001/jamanetworkopen.2024.38573)] [Medline: [39401041](https://pubmed.ncbi.nlm.nih.gov/39401041/)]
131. Mandal S, Wiesenfeld BM, Szerencsy AC, et al. Utilization of generative AI-drafted responses for managing patient-provider communication. *NPJ Digit Med*. Oct 2, 2025;8(1):591. [doi: [10.1038/s41746-025-01972-w](https://doi.org/10.1038/s41746-025-01972-w)] [Medline: [41038966](https://pubmed.ncbi.nlm.nih.gov/41038966/)]
132. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on May 24, 2019. URL: <https://doi.org/10.48550/arXiv.1810.04805> [Accessed 2026-03-20]
133. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol*. Jul 10, 2008;8(45):45. [doi: [10.1186/1471-2288-8-45](https://doi.org/10.1186/1471-2288-8-45)] [Medline: [18616818](https://pubmed.ncbi.nlm.nih.gov/18616818/)]
134. Sezgin E. Artificial intelligence in healthcare: complementing, not replacing, doctors and healthcare providers. *Digit Health*. 2023;9:20552076231186520. [doi: [10.1177/20552076231186520](https://doi.org/10.1177/20552076231186520)] [Medline: [37426593](https://pubmed.ncbi.nlm.nih.gov/37426593/)]
135. De Angelis L, Baglivo F, Arzilli G, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health*. 2023;11:1166120. [doi: [10.3389/fpubh.2023.1166120](https://doi.org/10.3389/fpubh.2023.1166120)] [Medline: [37181697](https://pubmed.ncbi.nlm.nih.gov/37181697/)]
136. Holmes G, Tang B, Gupta S, Venkatesh S, Christensen H, Whitton A. Applications of large language models in the field of suicide prevention: scoping review. *J Med Internet Res*. Jan 23, 2025;27:e63126. [doi: [10.2196/63126](https://doi.org/10.2196/63126)] [Medline: [39847414](https://pubmed.ncbi.nlm.nih.gov/39847414/)]
137. Al Zaabi A, Alshibli R, AlAmri A, AlRuheili I, Lutfi SL. Trends and trajectories in the rise of large language models in radiology: scoping review. *JMIR Med Inform*. Dec 9, 2025;13:e78041. [doi: [10.2196/78041](https://doi.org/10.2196/78041)] [Medline: [41364806](https://pubmed.ncbi.nlm.nih.gov/41364806/)]
138. Zolnieriek KBH, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. *Med Care*. Aug 2009;47(8):826-834. [doi: [10.1097/MLR.0b013e31819a5acc](https://doi.org/10.1097/MLR.0b013e31819a5acc)] [Medline: [19584762](https://pubmed.ncbi.nlm.nih.gov/19584762/)]
139. Shao X, Ruan T, Ju X, Sun Y, Cui J. Evaluating artificial intelligence chatbots' responses to gynecomastia inquiries: comparative study of information quality, readability, and guideline consistency. *Digit Health*. 2025;11:20552076251367645. [doi: [10.1177/20552076251367645](https://doi.org/10.1177/20552076251367645)] [Medline: [40881062](https://pubmed.ncbi.nlm.nih.gov/40881062/)]
140. Reis M, Reis F, Kunde W. Influence of believed AI involvement on the perception of digital medical advice. *Nat Med*. Nov 2024;30(11):3098-3100. [doi: [10.1038/s41591-024-03180-7](https://doi.org/10.1038/s41591-024-03180-7)] [Medline: [39054373](https://pubmed.ncbi.nlm.nih.gov/39054373/)]
141. Stade EC, Stirman SW, Ungar LH, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *Npj Ment Health Res*. Apr 2, 2024;3(1):12. [doi: [10.1038/s44184-024-00056-z](https://doi.org/10.1038/s44184-024-00056-z)] [Medline: [38609507](https://pubmed.ncbi.nlm.nih.gov/38609507/)]
142. Lin Y, Luo Z, Ye Z, et al. Applications, challenges, and prospects of generative artificial intelligence empowering medical education: scoping review. *JMIR Med Educ*. Oct 23, 2025;11:e71125. [doi: [10.2196/71125](https://doi.org/10.2196/71125)] [Medline: [41128430](https://pubmed.ncbi.nlm.nih.gov/41128430/)]
143. Zeng J, Qi W, Shen S, et al. Embracing the future of medical education with large language model-based virtual patients: scoping review. *J Med Internet Res*. Nov 13, 2025;27:e79091. [doi: [10.2196/79091](https://doi.org/10.2196/79091)] [Medline: [41232097](https://pubmed.ncbi.nlm.nih.gov/41232097/)]
144. Yu S, Lee SS, Hwang H. The ethics of using artificial intelligence in medical research. *KMJ*. Dec 2024;39(4):229-237. [doi: [10.7180/kmj.24.140](https://doi.org/10.7180/kmj.24.140)]
145. Qian H, Dong B, Yuan JJ, et al. Pre-consultation system based on the artificial intelligence has a better diagnostic performance than the physicians in the outpatient department of pediatrics. *Front Med (Lausanne)*. 2021;8:695185. [doi: [10.3389/fmed.2021.695185](https://doi.org/10.3389/fmed.2021.695185)] [Medline: [34820391](https://pubmed.ncbi.nlm.nih.gov/34820391/)]
146. Bedi S, Liu Y, Orr-Ewing L, et al. Testing and evaluation of health care applications of large language models: a systematic review. *JAMA*. Jan 28, 2025;333(4):319-328. [doi: [10.1001/jama.2024.21700](https://doi.org/10.1001/jama.2024.21700)] [Medline: [39405325](https://pubmed.ncbi.nlm.nih.gov/39405325/)]
147. Bayor AA, Li J, Yang IA, Varnfield M. Designing clinical decision support systems (CDSS)-a user-centered lens of the design characteristics, challenges, and implications: systematic review. *J Med Internet Res*. Jun 20, 2025;27:e63733. [doi: [10.2196/63733](https://doi.org/10.2196/63733)] [Medline: [40540451](https://pubmed.ncbi.nlm.nih.gov/40540451/)]
148. Jung K, Kashyap S, Avati A, et al. A framework for making predictive models useful in practice. *J Am Med Inform Assoc*. Jun 12, 2021;28(6):1149-1158. [doi: [10.1093/jamia/ocaa318](https://doi.org/10.1093/jamia/ocaa318)] [Medline: [33355350](https://pubmed.ncbi.nlm.nih.gov/33355350/)]
149. Huo B, Boyle A, Marfo N, et al. Large language models for chatbot health advice studies: a systematic review. *JAMA Netw Open*. Feb 3, 2025;8(2):e2457879. [doi: [10.1001/jamanetworkopen.2024.57879](https://doi.org/10.1001/jamanetworkopen.2024.57879)] [Medline: [39903463](https://pubmed.ncbi.nlm.nih.gov/39903463/)]

150. Shool S, Adimi S, Saboori Amlashi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Med Inform Decis Mak*. Mar 7, 2025;25(1):117. [doi: [10.1186/s12911-025-02954-4](https://doi.org/10.1186/s12911-025-02954-4)] [Medline: [40055694](https://pubmed.ncbi.nlm.nih.gov/40055694/)]
151. 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*. Mar 2024;151:104620. [doi: [10.1016/j.jbi.2024.104620](https://doi.org/10.1016/j.jbi.2024.104620)] [Medline: [38462064](https://pubmed.ncbi.nlm.nih.gov/38462064/)]
152. Doshi R, Amin K, Khosla P, Bajaj S, Chheang S, Forman HP. Utilizing large language models to simplify radiology reports: a comparative analysis of ChatGPT3.5, ChatGPT4.0, Google Bard, and Microsoft Bing. *Radiology and Imaging*. Preprint posted online on Jun 7, 2023. [doi: [10.1101/2023.06.04.23290786](https://doi.org/10.1101/2023.06.04.23290786)]
153. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res*. Jul 2020;99(7):769-774. [doi: [10.1177/0022034520915714](https://doi.org/10.1177/0022034520915714)] [Medline: [32315260](https://pubmed.ncbi.nlm.nih.gov/32315260/)]
154. Jiang C, Cheng W, Jiang X, Zhang J, Tang X. A comparative study of ChatGPT-4o and DeepSeek responses to mandibular angle osteotomy questions. *J Craniofac Surg*. Jul 31, 2025;36(7):e1113-e1116. [doi: [10.1097/SCS.00000000000011698](https://doi.org/10.1097/SCS.00000000000011698)] [Medline: [40742907](https://pubmed.ncbi.nlm.nih.gov/40742907/)]
155. Forsey J, Ng S, Rowland P, Freeman R, Li C, Woods NN. The basic science of patient-physician communication: a critical scoping review. *Acad Med*. Nov 1, 2021;96(11S):S109-S118. [doi: [10.1097/ACM.00000000000004323](https://doi.org/10.1097/ACM.00000000000004323)] [Medline: [34348382](https://pubmed.ncbi.nlm.nih.gov/34348382/)]
156. Zheng S, Zhao N, Wang J, et al. Comparison of a specialized large language model with GPT-4o for CT and MRI radiology report summarization. *Radiology*. Aug 2025;316(2):e243774. [doi: [10.1148/radiol.243774](https://doi.org/10.1148/radiol.243774)] [Medline: [40892451](https://pubmed.ncbi.nlm.nih.gov/40892451/)]
157. Lehman EP, Hernandez E, Mahajan D, et al. Do we still need clinical language models? Presented at: Conference on Health, Inference, and Learning (CHIL) 2023; Jun 22-24, 2023; Cambridge, MA. [doi: [10.48550/arXiv.2302.08091](https://doi.org/10.48550/arXiv.2302.08091)]
158. Zhang Q, Wu Z, Song J, Luo S, Chai Z. Comprehensiveness of large language models in patient queries on gingival and endodontic health. *Int Dent J*. Feb 2025;75(1):151-157. [doi: [10.1016/j.identj.2024.06.022](https://doi.org/10.1016/j.identj.2024.06.022)] [Medline: [39147663](https://pubmed.ncbi.nlm.nih.gov/39147663/)]
159. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst*. Mar 4, 2023;47(1):33. [doi: [10.1007/s10916-023-01925-4](https://doi.org/10.1007/s10916-023-01925-4)] [Medline: [36869927](https://pubmed.ncbi.nlm.nih.gov/36869927/)]
160. Luo PW, Liu JW, Xie X, et al. DeepSeek vs ChatGPT: a comparison study of their performance in answering prostate cancer radiotherapy questions in multiple languages. *Am J Clin Exp Urol*. 2025;13(2):176-185. [doi: [10.62347/UIAP7979](https://doi.org/10.62347/UIAP7979)] [Medline: [40400997](https://pubmed.ncbi.nlm.nih.gov/40400997/)]
161. Yıldırım A, Cicek O, Genç YS. Can AI-based ChatGPT models accurately analyze hand-wrist radiographs? a comparative study. *Diagnostics (Basel)*. Jun 14, 2025;15(12):1513. [doi: [10.3390/diagnostics15121513](https://doi.org/10.3390/diagnostics15121513)] [Medline: [40564836](https://pubmed.ncbi.nlm.nih.gov/40564836/)]
162. Chen YQ, Zhang SX. Superior resilience to poisoning and amenability to unlearning in quantum machine learning. *Nat Commun*. Mar 9, 2026;17(1):3716. [doi: [10.1038/s41467-026-70420-4](https://doi.org/10.1038/s41467-026-70420-4)] [Medline: [41803125](https://pubmed.ncbi.nlm.nih.gov/41803125/)]
163. Yang L, Chen H, Li Z, Ding X, Wu X. Give us the facts: enhancing large language models with knowledge graphs for fact-aware language modeling. *IEEE Trans Knowl Data Eng*. 2024;36(7):3091-3110. [doi: [10.1109/TKDE.2024.3360454](https://doi.org/10.1109/TKDE.2024.3360454)] [Medline: [25079929](https://pubmed.ncbi.nlm.nih.gov/25079929/)]
164. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. *Transl Psychiatry*. Oct 6, 2023;13(1):309. [doi: [10.1038/s41398-023-02592-2](https://doi.org/10.1038/s41398-023-02592-2)] [Medline: [37798296](https://pubmed.ncbi.nlm.nih.gov/37798296/)]
165. Quer G, Topol EJ. The potential for large language models to transform cardiovascular medicine. *Lancet Digit Health*. Oct 2024;6(10):e767-e771. [doi: [10.1016/S2589-7500\(24\)00151-1](https://doi.org/10.1016/S2589-7500(24)00151-1)] [Medline: [39214760](https://pubmed.ncbi.nlm.nih.gov/39214760/)]
166. Qian Y, Tang W, Xu T, Zhu D, Hua W. Emotional support powered by artificial intelligence in healthcare settings: a scoping review of technologies, contents, and outcomes. *Int J Nurs Stud*. Dec 2025;172:105243. [doi: [10.1016/j.ijnurstu.2025.105243](https://doi.org/10.1016/j.ijnurstu.2025.105243)] [Medline: [41092641](https://pubmed.ncbi.nlm.nih.gov/41092641/)]
167. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat Mach Intell*. 2023;5(1):46-57. [doi: [10.1038/s42256-022-00593-2](https://doi.org/10.1038/s42256-022-00593-2)]
168. Chen M, Li C, Sitaram S, Wang J, Xie X. CultureLLM: incorporating cultural differences into large language models. Presented at: *Advances in Neural Information Processing Systems 37*; Dec 10-15, 2024; Vancouver, BC, Canada. [doi: [10.52202/079017-2693](https://doi.org/10.52202/079017-2693)]
169. Shumway DO, Hartman HJ. Medical malpractice liability in large language model artificial intelligence: legal review and policy recommendations. *J Osteopath Med*. Jul 1, 2024;124(7):287-290. [doi: [10.1515/jom-2023-0229](https://doi.org/10.1515/jom-2023-0229)] [Medline: [38295300](https://pubmed.ncbi.nlm.nih.gov/38295300/)]
170. Ong JCL, Chang SYH, William W, et al. Ethical and regulatory challenges of large language models in medicine. *Lancet Digit Health*. Jun 2024;6(6):e428-e432. [doi: [10.1016/S2589-7500\(24\)00061-X](https://doi.org/10.1016/S2589-7500(24)00061-X)] [Medline: [38658283](https://pubmed.ncbi.nlm.nih.gov/38658283/)]
171. Alzoubi YI, Mishra A. Differential privacy and artificial intelligence: potentials, challenges, and future avenues. *EURASIP J Inf Sec*. 2025;2025(1). [doi: [10.1186/s13635-025-00203-9](https://doi.org/10.1186/s13635-025-00203-9)]

172. Wang M, Liu Y, Liang X, Li S, Huang Y, Zhang X, et al. LangGPT: rethinking structured reusable prompt design framework for llms from the programming language. arXiv. Preprint posted online on Jun 29, 2024. [doi: [10.48550/arXiv.2402.16929](https://doi.org/10.48550/arXiv.2402.16929)]
173. Perret S, Alon N, Carpenter-Song E, et al. Standardising the role of a digital navigator in behavioural health: a systematic review. *Lancet Digit Health*. Dec 2023;5(12):e925-e932. [doi: [10.1016/S2589-7500\(23\)00152-8](https://doi.org/10.1016/S2589-7500(23)00152-8)] [Medline: [38000876](https://pubmed.ncbi.nlm.nih.gov/38000876/)]
174. Wang Z, Gao J, Danek B, et al. Compliance and factuality of large language models for clinical research document generation. *J Am Med Inform Assoc*. Mar 1, 2026;33(3):563-572. [doi: [10.1093/jamia/ocaf174](https://doi.org/10.1093/jamia/ocaf174)] [Medline: [41144289](https://pubmed.ncbi.nlm.nih.gov/41144289/)]
175. Hao Y, Holmes J, Waddle MR, et al. Personalizing prostate cancer education for patients using an EHR-integrated LLM agent. *NPJ Digit Med*. Dec 18, 2025;8(1):770. [doi: [10.1038/s41746-025-02166-0](https://doi.org/10.1038/s41746-025-02166-0)] [Medline: [41413170](https://pubmed.ncbi.nlm.nih.gov/41413170/)]
176. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282. [Medline: [23092060](https://pubmed.ncbi.nlm.nih.gov/23092060/)]

## Abbreviations

**AI:** artificial intelligence

**CAT:** communication accommodation theory

**LLM:** large language model

**MeSH:** Medical Subject Headings

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**PRISMA-S:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension

**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

*Edited by Stefano Brini; peer-reviewed by Bell Eapen, Dillon Chrimes; submitted 25.Sep.2025; final revised version received 04.Apr.2026; accepted 06.Apr.2026; published 26.Jun.2026*

*Please cite as:*

*Chang J, Peng R, Chen X, Zhu Y, Miao R, Cao Z, Feng H*

*Applications, Challenges, and Future Directions of Large Language Models in Health Care Communication: Scoping Review*

*J Med Internet Res 2026;28:e84726*

*URL: <https://www.jmir.org/2026/1/e84726>*

*doi: [10.2196/84726](https://doi.org/10.2196/84726)*

© Jing Chang, Ruotong Peng, Xi Chen, Yishu Zhu, Ruting Miao, Zeng Cao, Hui Feng. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 26.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.