

Review

Applications of DeepSeek in Medicine: Bibliometric Analysis and Scoping Review

Haoran Zhang^{1,2*}, MS; Dawei Wang^{3*}, PhD; Yanliang Xu⁴, BSc; Shuming Han², MS; Guangxin Wang², MD, PhD

¹School of Clinical Medicine, Shandong Second Medical University, Weifang, Shandong, China

²Shandong Innovation Center of Intelligent Diagnostic Technology, Central Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

³Key Laboratory of Endocrine Glucose & Lipids Metabolism and Brain Aging, Ministry of Education; Department of Endocrinology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

⁴Library, Shandong Second Medical University, Weifang, Shandong, China

*these authors contributed equally

Corresponding Author:

Guangxin Wang, MD, PhD
Shandong Innovation Center of Intelligent Diagnostic Technology
Central Hospital Affiliated to Shandong First Medical University
105 Jiefang Road
Jinan, Shandong 250013
China
Phone: 86 531 55865152
Email: y22183@email.sdfmu.edu.cn

Abstract

Background: The integration of large language models (LLMs) into medicine has reshaped health care delivery, education, and research. Although proprietary models face challenges such as data privacy, regulation, and adaptability, DeepSeek, an open-source LLM, has emerged as a customizable and cost-effective alternative with significant potential for clinical and operational applications. However, the rapid expansion of research in this area necessitates a systematic mapping of its landscape, applications, and challenges.

Objective: This study combines bibliometric analysis with a scoping review to systematically map and characterize the literature on DeepSeek's medical applications. The aims were to (1) analyze publication trends, leading contributors, and research themes and (2) identify primary application domains, strengths, limitations, and future directions.

Methods: Following the framework by Arksey and O'Malley and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, a systematic search was conducted using PubMed, Web of Science, and Scopus from January 20, 2025, to November 30, 2025. Bibliometric analysis was then used to quantify publication trends, productivity, and research themes across 371 papers. The scoping review thematically synthesized the applications, strengths, and limitations of 353 original articles.

Results: The publication output showed a progressive increase, with China (n=163), Turkey (n=52), and the United States (n=48) as leading contributors. Keyword co-occurrence analysis formed 7 clusters; the 3 most frequent keywords were "large language model," "artificial intelligence," and "patient education." DeepSeek has shown promising yet preliminary performance across multiple domains, including patient education, clinical decision support, medical education, workflow optimization, and medical research. The evidence base remains predominantly low in quality, with 66.6% (235/353) of original articles classified as low-quality evidence, consisting largely of unvalidated benchmarking, simulated cases, and single-center retrospective analyses. Only 6.8% (24/353) of studies met the criteria to be considered high quality, and prospective randomized trials assessing patient-relevant outcomes were notably absent.

Conclusions: Publications on DeepSeek's medical applications increased progressively from January 2025 through November 2025, with China, Turkey, and the United States as the leading contributors. The scoping review found that DeepSeek has been evaluated across 5 domains (patient education, clinical decision support, medical education, workflow optimization, and research), with variable but often competitive performance relative to proprietary models. Strengths included readability, diagnostic accuracy in select specialties, cost-efficiency, and local deployability. Limitations included inconsistent cross-specialty performance, hallucinations, ethical concerns, data privacy issues, and regulatory gaps. The evidence base is

predominantly low-quality and simulation-based, with few prospective trials or randomized controlled trials. These findings indicate that DeepSeek's clinical readiness varies, and future research should address prospective validation, multimodal capabilities, bias mitigation, human oversight, and equitable access.

J Med Internet Res 2026;28:e93354; doi: [10.2196/93354](https://doi.org/10.2196/93354)

Keywords: DeepSeek; large language model; artificial intelligence in medicine; clinical decision support; medical education; scoping review; biomedical ethics; PRISMA

Introduction

The integration of artificial intelligence (AI), particularly large language models (LLMs), into medicine has prompted a paradigm shift in health care delivery, education, and research. LLMs, such as OpenAI's GPT series, have demonstrated considerable capabilities for processing complex medical data, supporting clinical decision-making, and improving patient communication. However, the widespread adoption of proprietary LLMs in clinical settings faces substantial challenges, including data privacy concerns, regulatory constraints, and limited adaptability to institutional requirements [1]. In this context, DeepSeek, an open-source LLM developed by Hangzhou DeepSeek Artificial Intelligence Basic Technology Research Co Ltd, has emerged as a promising alternative, distinguished by its customizability, cost-effectiveness, and alignment with data governance standards [1-3]. This model represents a significant advancement in AI, particularly for its sophisticated reasoning capabilities and its impact on AI research and applications.

DeepSeek's architecture, especially in reasoning-enhanced iterations such as DeepSeek-R1, incorporates innovative training approaches, including Group Relative Policy Optimization (GRPO). This rule-based reinforcement learning paradigm, which functions without task-specific supervised fine-tuning during the reasoning alignment phase and builds upon a pretrained base model, fosters emergent reasoning behaviors that are particularly valuable for complex medical reasoning tasks [4,5]. This open-weight nature enables local deployment, making it particularly attractive in health care settings where data security and privacy are paramount [1,6]. Since its release, DeepSeek and its associated intelligent agents have been implemented in multiple tertiary hospitals across China, resulting in measurable improvements in clinical and operational workflows, including patient follow-up, imaging analysis, and administrative automation [7-9]. Such real-world implementations underscore the potential for redefining AI-driven health care delivery.

The growing corpus of studies evaluating DeepSeek medical applications has revealed several strengths. In clinical diagnostics, DeepSeek-R1 achieved a diagnostic accuracy comparable to that of GPT-4 in complex clinicopathological cases [10]. In specialized areas, such as ophthalmology, it has exhibited diagnostic and management performance on par with OpenAI o1 while reducing token-related costs by approximately 15-fold [11]. Moreover, DeepSeek excels in Chinese-language medical contexts, outperforming ChatGPT at delivering prostate cancer radiotherapy information in Chinese and demonstrating superior results on

Chinese medical licensing examinations [12,13]. Beyond clinical decision support, DeepSeek shows promise in medical education, patient communication, and administrative tasks, with documented deployments across multiple Chinese tertiary hospitals supporting applications ranging from imaging interpretation to automated administrative workflows [9]. However, these promising benchmarking results warrant further examination in real-world clinical settings, which are now emerging primarily in China.

The rapid integration of DeepSeek into clinical practice, particularly within Chinese hospital systems [2,9], underscores the necessity for a thorough evaluation of its applications, limitations, and future directions. The existing literature lacks a comprehensive assessment of publication trends and emerging research fronts in this rapidly evolving domain. Evidence remains fragmented across medical specialties, and the heterogeneous methodologies and outcomes limit a holistic understanding of the model's clinical utility, safety profile, and readiness for broader implementation. Therefore, a comprehensive synthesis of available evidence is essential to guide health care institutions, policymakers, and developers in evaluating DeepSeek's realistic capabilities, optimal deployment strategies, and associated risks.

To address this gap and systematically map the research landscape, this study adopted an integrated methodological approach that combined bibliometric analysis with a scoping review. Bibliometric analysis quantitatively characterizes the field at the macro level, examining publication trends over time, core authors and institutions, high-frequency keywords, and journal distributions. This enables the objective identification of research hot spots and evolutionary trajectories [14,15]. Simultaneously, a scoping review is a systematic methodology designed to map key concepts, evidence types, and knowledge gaps within a broad or emerging field. Rather than synthesizing evidence for definitive conclusions, it uses qualitative or descriptive methods to identify existing research themes, methodological characteristics, and underexplored areas, thereby clarifying the overall research landscape [16]. Given that literature on DeepSeek in medicine is growing rapidly and includes highly heterogeneous publications, such as proof-of-concept studies, preclinical research, preliminary clinical trials, and technical descriptions, a scoping review is more suitable than a systematic review for this context, as it focuses on comprehensively mapping the domain without mandating formal quality appraisal. The combination of these two methods leveraged their complementary strengths: Bibliometric analysis provides an objective, structured quantitative overview, while the scoping review delivers a nuanced, contextualized conceptual map. This integrated analysis provided a more powerful

and multidimensional understanding of the field's scope, developmental dynamics, and future directions from both quantitative and qualitative perspectives.

Guided by this integrated approach, the study was structured as follows. First, a bibliometric analysis was conducted to examine relevant original articles and reviews, addressing the following questions: (1) What are the volume, growth trajectory, and geographic distribution of publications? (2) Which countries/regions, institutions, and authors are leading the research? and (3) What are the key research themes and their evolution? Second, a scoping review was performed to critically evaluate the literature content, focusing on the following questions: What are the primary medical application domains of DeepSeek, and how do trends vary across different health care fields? Finally, the discussion synthesizes findings from both methods to highlight implementation challenges, identify major research gaps, and suggest future directions for the effective integration of DeepSeek into global health care systems.

Methods

Overview

This study used an integrated approach that combined bibliometric analysis and a scoping review to provide complementary insights. The bibliometric method examined the current application of DeepSeek in medicine from multiple dimensions, analyzed researcher characteristics and journal distributions, and identified research hot spots and trends. The bibliometric analysis was conducted based on the framework proposed by Cobo et al [17], following the guidelines for reporting bibliometric reviews of biomedical literature (BIBLIO) [18]. This scoping review systematically extracted and synthesized the applications, challenges, and future research directions of DeepSeek in medicine. The study was conducted according to the framework by Arksey and O'Malley [19] and reported following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1) [20].

Databases, Search Strategy, and Screening Process

To ensure a comprehensive retrieval of the literature on the applications of DeepSeek in medicine, a systematic search was conducted on December 16, 2025, in PubMed, Web of Science Core Collection (WoSCC), and Scopus. The search strategy (Multimedia Appendix 1) used both controlled vocabularies (MeSH, Web of Science Categories, and SUBJAREA) and free-text terms tailored to each database to optimize retrieval.

Given that the public release of DeepSeek's reasoning model, DeepSeek-R1, on January 20, 2025 [21,22], marked the beginning of subsequent research into its applications, including in medicine, the search encompassed the period from January 20, 2025, to November 30, 2025.

To ensure comprehensive retrieval, the inclusion criteria were as follows: (1) studies investigating the application of DeepSeek in medicine, (2) document types limited to original articles and reviews for bibliometric analysis and original articles only for scoping review, (3) studies published in peer-reviewed academic journals, and (4) no language restrictions.

The exclusion criteria were as follows: (1) duplicate publications; (2) literature that proposed only speculative or hypothetical uses without substantive analysis or findings; (3) non-peer-reviewed journal items, including books, editorials, preprints, commentaries, conference abstracts, case reports, and retracted articles; and (4) studies with insufficient information for bibliometric analysis or whose full text was unavailable for in-depth content extraction during the scoping review.

After receiving professional training, two authors (HZ and DW) independently screened the titles and abstracts and excluded irrelevant studies based on the aforementioned criteria. The interrater agreement was almost perfect (Cohen $\kappa=0.93$). Any disagreements during screening were resolved through discussion or, when necessary, arbitration by a third reviewer (GW).

Bibliometric Analysis

The final bibliometric analysis included 371 papers. Full records of the selected publications were exported and stored in Excel 2021 (Microsoft Corp) and EndNote Desktop (Clarivate). Bibliographic metadata such as authors' names, affiliations, countries/regions, and keywords were standardized in a uniform format.

Excel 2021 was used to generate tables highlighting the top 10 authors, institutions, and countries/regions based on their publication output, whereas VOSviewer (version 1.6.19) was used for data visualization of bibliometric mapping, including keyword co-occurrence analysis. Keyword co-occurrence analysis examined the fundamental characteristics of keywords, such as their frequency and temporal evolution. This method helped identify research hot spots and track developmental trends within specialized fields. The three common types of visualizations used in the keyword co-occurrence analysis were the network, density, and overlay maps. In the network map, nodes represented keywords, and the connecting lines represented keyword co-occurrence relationships. The size of a node indicates its frequency, the thickness of a line represents the strength of co-occurrence, and the nodes are clustered together by color to reveal distinct research themes or subfields. The overlay map chronologically visualized the keyword trajectories by assigning chromatic codes corresponding to the computationally derived average publication years (APYs). The density map emphasizes the "research density" or concentration of keywords in the knowledge landscape. Areas with numerous closely located keywords appear as warm-colored regions, such as purple, indicating core well-developed research fronts. Cooler-colored areas such as blue or white represent sparser, potentially peripheral, or emerging topics. The centrality of keywords, which reflects their capacity to bridge

different parts of the research network, was derived using CiteSpace (version 7.0.0).

Scoping Review

This scoping review included a total of 353 publications. A data extraction form was created using Excel to extract in-depth content from the papers. This form included items such as paper title, research objectives, key findings, research design types, DeepSeek’s strengths, limitations and challenges, future recommendations, DeepSeek model version, quality tier, and application areas. It should be noted that, although quality assessment is not obligatory for scoping reviews, the methodological quality of all included studies was categorized into 3 tiers (high, moderate, and low) based on the criteria (Multimedia Appendix 2) in order to characterize the strength of the available evidence. Data extraction was conducted independently by 2 authors (HZ and DW). Both authors independently extracted data from all 353 included articles in duplicate using the data extraction form created in Excel. After independent extraction, the 2 authors compared their results. Disagreements were resolved through discussion or by consulting a third author (GW) when consensus could not be reached. The extracted data (Multimedia Appendix 3) were then critically analyzed and organized thematically

to address the research question, thereby mapping the key application areas of DeepSeek in medicine. The discussion section elaborates on the challenges, research gaps, and future work for the application of DeepSeek in the medical field.

Ethical Considerations

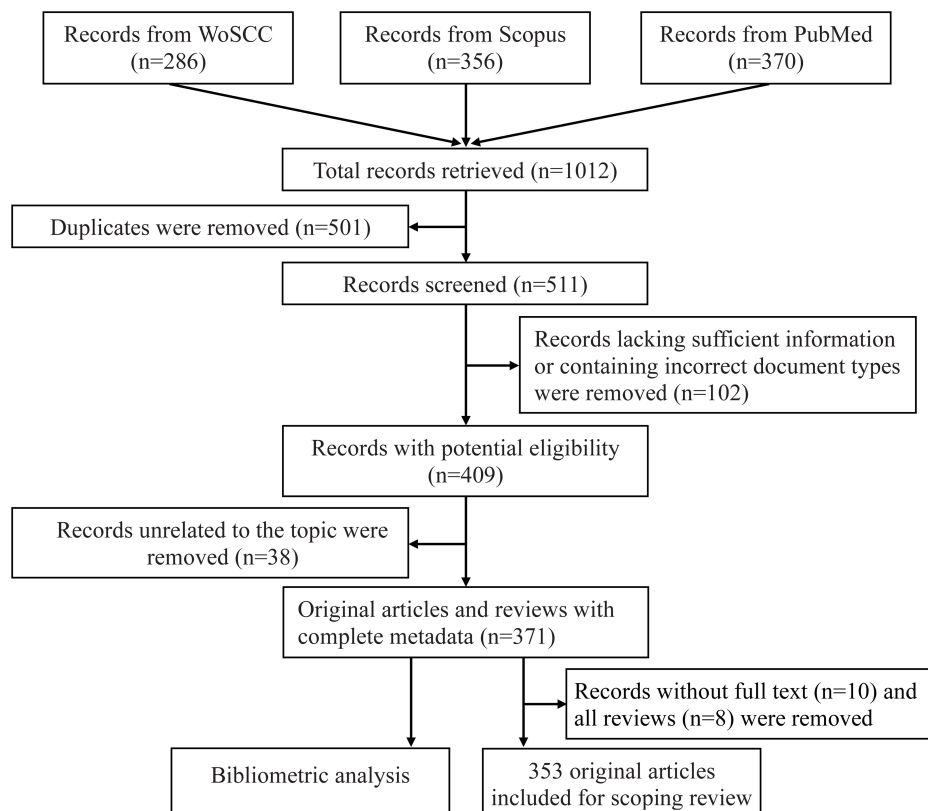
Since this study was a bibliometric and scoping review of previously published literature, ethical approval from an ethics committee is not required.

Results

Bibliometric Analysis of DeepSeek Applications in Medicine

A systematic search of PubMed, Scopus, and WoSCC yielded 371 publications on the application of DeepSeek in medicine for bibliometric analysis (Figure 1). Among these, the majority (363/371, 97.8%) were categorized as original articles, while the remaining (8/371, 2.2%) were reviews. In terms of publication languages, 358 papers were written in English, and 13 were written in Chinese.

Figure 1. The diagram depicting the paper selection process. WoSCC: Web of Science Core Collection.

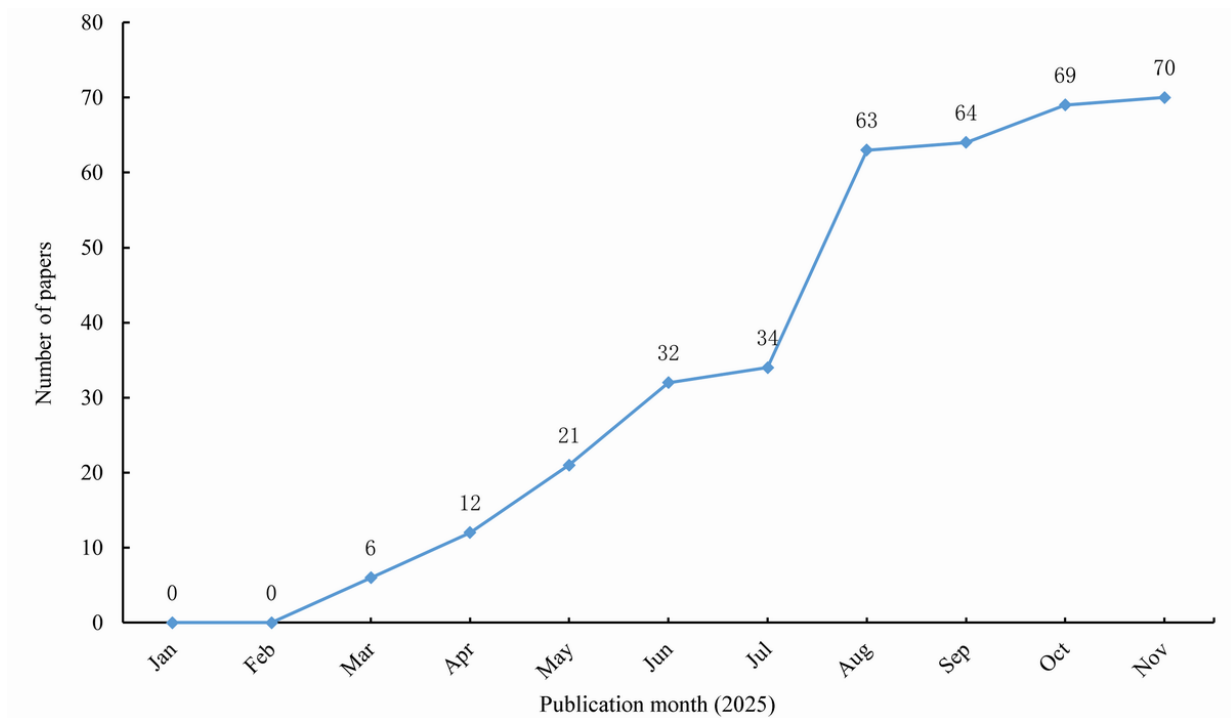


Monthly Publication Output

The monthly publication output increased progressively over time. From January to November 2025, the number of papers

rose from 0 to 70, with the highest output (70 papers) observed in November (Figure 2).

Figure 2. Monthly count of publications on DeepSeek's medical applications identified in this review.



Analysis of Source Journals

Of the 216 journals that published papers on the applications of DeepSeek in medicine, 12 published more than 5 papers each. The 10 most active journals collectively contributed 90 publications, accounting for 24.3% (90/371) of the total output. *Cureus* was the most productive journal with 19 publications, followed by *Scientific Reports* (n=10), *BMC Oral Health* (n=9), *International Journal of Medical Informatics* (n=9), *BMC Medical Education* (n=8), *Frontiers in Artificial Intelligence* (n=7), *Frontiers in Public Health*

(n=7), *JMIR Medical Informatics* (n=7), *Journal of Medical Internet Research* (n=7), and *Journal of Medical Systems* (n=7).

The Top 10 Authors, Institutions, and Nations/Regions Ranked by Publication Count

Table 1 presents the top 10 authors, institutions, and countries/regions ranked by their respective number of publications on the applications of DeepSeek in medicine.

Table 1. The top 10 authors, organizations, and countries ranked by the number of papers.

Rank	Authors ^a		Organizations ^a		Countries/Regions ^a	
	Name	Papers, n	Name	Papers, n	Name	Papers, n
1	Liu Y	6	Shanghai Jiao Tong University	16	China	163
2	Zhang J	5	Chinese Academy of Medical Sciences	10	Turkey	52
3	Li J	5	Sichuan University	10	United States	48
4	Wang J	5	Zhejiang University	9	Germany	24
5	Wang Y	5	Capital Medical University	9	India	23
6	Xu L	4	University of Health Sciences, Turkey	9	United Kingdom	20
7	Rozen WM	3	Southern Medical University	8	Italy	14
8	Cuomo R	3	Soochow University	7	Saudi Arabia	14
9	Marcaccini G	3	Sun Yat-sen University	7	Australia	9
10	Chen S	3	Tsinghua University	6	Canada	8

^aThese 3 categories are independent of each other.

Most Cited Papers on the Medical Applications of DeepSeek

Table 2 lists the 10 most-cited publications on the medical applications of DeepSeek: 9 were original articles, while 1 was a review [12,23-31].

Table 2. Top 10 most-cited publications on the medical applications of DeepSeek.

Rank	Authors	Publication date	Total citations, n	Research focus
1	Zhou et al [23]	June 2025	50	Comparative evaluation of DeepSeek and ChatGPT models
2	Deng et al [24]	May 2025	38	DeepSeek's advances, applications, and challenges across various domains, including health care
3	Kaygisiz and Teke [25]	April 2025	29	DeepSeek's diagnostic performance in oral pathologies
4	Rasool et al [26]	March 2025	28	DeepSeek's emotion-aware embedding fusion for responses
5	Yilmaz et al [27]	April 2025	16	Comparative performance of LLMs ^a on oral pathology multiple-choice questions
6	Marcaccini et al [28]	March 2025	16	DeepSeek and AI ^b in hand fracture management
7	Luo et al [12]	April 2025	16	DeepSeek versus ChatGPT in multilingual prostate cancer radiotherapy
8	Özcivelek and Özcan [29]	May 2025	15	Comparative evaluation of AI chatbots on dental and maxillofacial prostheses
9	Gültekin et al [30]	August 2025	14	Comparative evaluation of AI models for patient education
10	Seth et al [31]	March 2025	12	Evaluating DeepSeek and AI in hand surgery decisions

^aLLMs: large language models.

^bAI: artificial intelligence.

Keyword Co-Occurrence Analysis

A keyword co-occurrence analysis was performed to map predominant research hot spots. Synonyms were consolidated prior to analysis; specifically, "large language model(s)" was standardized as "large language model," and "generative artificial intelligence/AI" was standardized as "generative artificial intelligence." The top 10 keywords by frequency are listed in Table 3. Notably, "generative artificial intelligence" ranked seventh in frequency but third in centrality. From an initial set of 968 keywords, 41 occurring more than 4 times were included in the keyword co-occurrence analysis. These formed 7 well-defined clusters, visualized in the network map (Figure 3A).

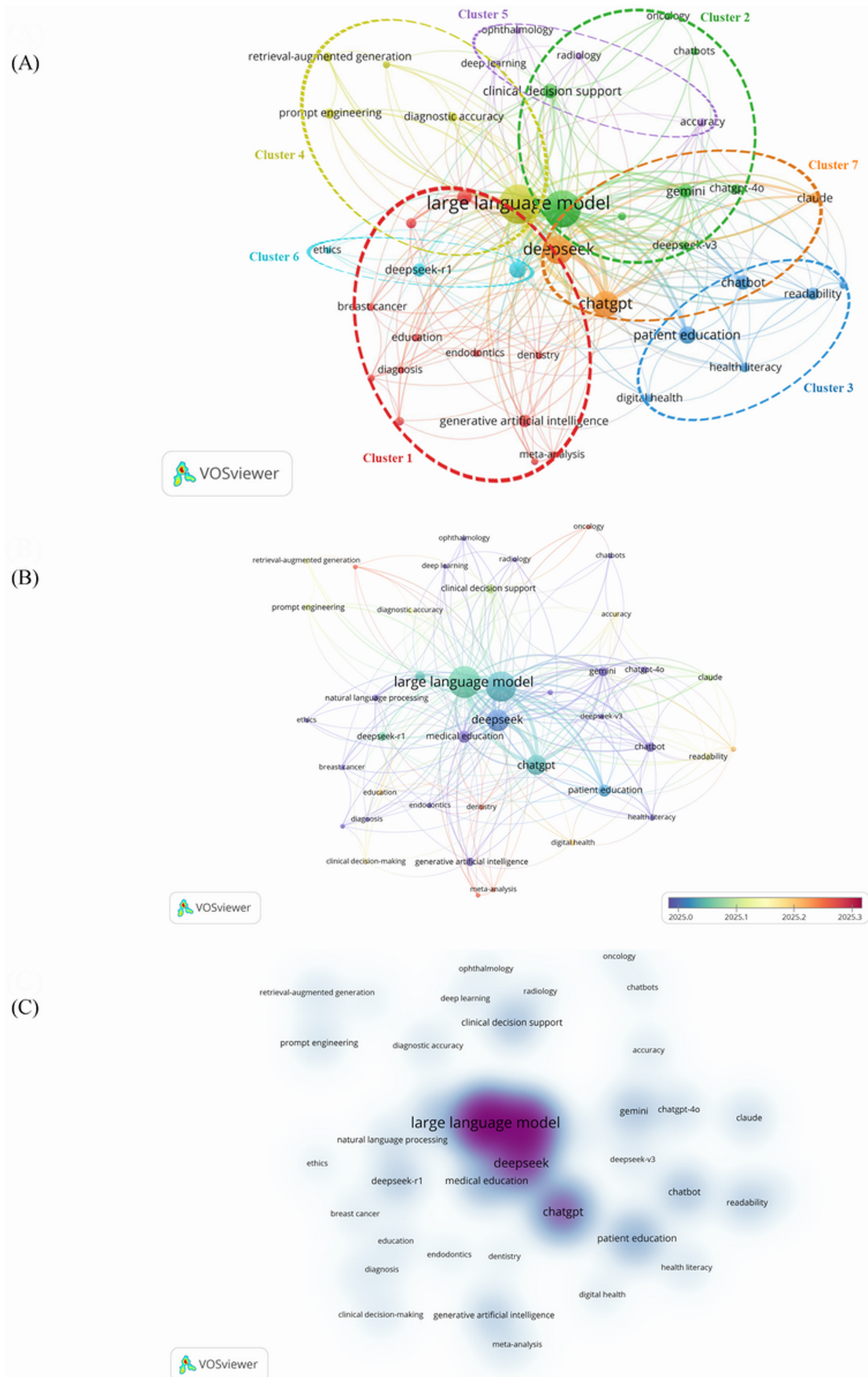
The temporal overlay map (Figure 3B) illustrates the evolution of research focus, with keywords colored by their APYs. Purple nodes represent earlier themes, while crimson indicates more recent activity. Early research concentrated primarily on medical education. The keywords "retrieval-augmented generation" and "oncology" showed the highest APY, reflecting a rising interest in these areas.

The density map (Figure 3C) displays keywords according to their average frequency of occurrence. Crimson regions correspond to the most frequently occurring keywords, followed by blue and then white areas, in descending order.

Table 3. The top 10 keywords regarding DeepSeek's applications in medicine.

Rank	Keywords	Frequency of occurrence, n	Centrality
1	Large language model	227	1.00
2	Artificial intelligence	197	0.55
3	Patient education	30	0.02
4	Medical education	28	0.01
5	Clinical decision support	19	0.01
6	Machine learning	19	0.05
7	Generative artificial intelligence	19	0.07
8	Natural language processing	9	0.01
9	Prompt engineering	8	0.00
10	Diagnostic accuracy	8	0.03

Figure 3. Keyword co-occurrence analysis visualization using VOSviewer [32,33]: (A) network visualization, with keywords grouped into 6 distinct thematic clusters; (B) overlay map colored by the average publication year of each keyword, ranging from purple (earlier) to crimson (recent); and (C) density map based on keyword co-occurrence frequency, where color intensity reflects occurrence rate: crimson (highest), blue (moderate), and white (lowest).



Summary of Extracted Data in the Scoping Review: Study Quality, Model Versions, Comparative Performance, and Documented Limitations

Of the 353 original articles, 24 (6.8%) met the criteria for high quality. These were primarily prospective evaluations and studies with external validation. A further 94 studies (94/353, 26.6%) were classified as moderate quality. The majority (235/353, 66.6%) were classified as low quality, reflecting the exploratory nature of the current evidence base, which is dominated by invalidated benchmarking using examination questions and single-center retrospective analyses.

Analysis of DeepSeek-specific versions revealed that DeepSeek-R1 was the most frequently studied (mentioned in 197 papers, 55.8% of the 353 articles), followed by DeepSeek-V3 (114/353, 32.3%) and unspecified versions of DeepSeek (61/353, 17.3%).

A total of 283 studies compared DeepSeek with other LLMs, primarily ChatGPT, in medical applications. Among these, 126 studies (126/283, 44.5%) reported positive results in which DeepSeek outperformed or showed significant advantages; 84 studies (84/283, 29.7%) reported neutral results with comparable performance, no statistically significant difference, or mixed strengths and limitations; and 73 studies (73/283, 25.8%) reported negative results in which DeepSeek underperformed relative to other models.

DeepSeek's primary weaknesses included inconsistent domain performance in 61 papers, incomplete answers in 47 papers, poor readability in 42 papers, and hallucinations in 38 papers. Ethical risks, though fewer in absolute count at 57 papers, were severe; specifically, non-maleficence was documented in 22 papers with potential patient harm, autonomy was documented in 15 papers with privacy and informed consent concerns, beneficence was documented in 8 papers with lack of empathy and impaired therapeutic relationship, and justice was documented in 12 papers highlighting bias and inequity. Other barriers reported in 55 papers further hindered clinical adoption.

Application Domains of DeepSeek in Medicine

Based on the scoping review of 353 full-text papers, the medical applications of DeepSeek can be summarized into the primary domains discussed in the following sections. Because a single study often evaluated DeepSeek in multiple domains, the sum of article counts across these domains exceeds 353.

DeepSeek in Patient Education and Communication

The applications of DeepSeek in patient education and communication were addressed in 105 articles. Among these, 91 were cross-sectional studies, 5 were descriptive studies, 4 were prospective studies including 1 randomized controlled trial (RCT), and the remaining 5 used other design types.

DeepSeek can generate patient-facing materials that are both readily comprehensible and clinically accurate. This capability has been empirically validated; for example, in generating patient education materials for spinal surgeries, DeepSeek-R1 achieved the lowest Flesch-Kincaid Grade Level scores, indicating content accessible to a broader audience including those with limited health literacy [23]. Similarly, in orthopedics, DeepSeek-R1 provided clearer and more easily understandable explanations of anterior cruciate ligament surgery than ChatGPT, which offered greater comprehensiveness but at a higher reading level [30]. This emphasis on linguistic accessibility is critical in patient-facing materials because improved readability enhances patient engagement, reduces anxiety, and supports informed decision-making [23,34]. Furthermore, DeepSeek has performed strongly in multilingual contexts, effectively generating patient education content in both Chinese and English, which is vital for serving diverse linguistic populations [12,35].

Although DeepSeek excels in readability, its responses sometimes lack comprehensive detail or sufficient citations of sources, and occasional inaccuracies or AI hallucinations have been noted [29,36,37]. Furthermore, some studies found that DeepSeek performed similarly to, or even less accurately than, ChatGPT when generating patient education materials [38,39].

DeepSeek in Clinical Decision Support and Treatment Planning

Of the 176 articles addressing DeepSeek in clinical decision support and treatment planning, 120 were cross-sectional studies, 22 were retrospective studies, 9 were prospective studies (including 2 RCTs), 2 were mixed-design studies, 14 were proof-of-concept studies, and the remaining 9 articles comprised expert consensus and other designs.

Regarding diagnostic accuracy, DeepSeek models have achieved notable results. In a dual-phase retrospective-prospective study classified as high methodological quality (n=300 liver lesions in the retrospective cohort and 126 liver lesions in the prospective cohort), DeepSeek-V3 demonstrated higher Liver Imaging Reporting and Data System (LI-RADS) classification accuracy than junior radiologists and achieved performance comparable with that of senior radiologists for hepatocellular carcinoma diagnosis [40]; however, this finding awaits replication in larger, multicenter settings. In a moderate-quality historical control study, DeepSeek-R1 demonstrated diagnostic accuracy comparable to that of GPT-4 in complex clinicopathologic cases [10]. In a low-quality cross-sectional study, Jiao et al [11] found that diagnostic accuracy in diagnosing corneal diseases varied significantly among LLMs ($P=.001$). GPT-4o achieved the highest accuracy (80%), while DeepSeek R1 achieved only 65%; both had accuracies that were significantly lower than that of human experts (92.5%; $P<.001$).

For treatment planning, DeepSeek-V3 demonstrated statistically superior accuracy compared with ChatGPT-o1 in head and neck cancer management [41], and DeepSeek-R1

outperformed OpenAI o1 in diagnostic accuracy and next-step decision-making in ophthalmology [42]. These models have demonstrated strengths in specialized domains, including hand fracture management [28], urinary incontinence management [43], and postprostatectomy urinary incontinence guidelines [44], although they have limitations in complex scenarios. Notably, DeepSeek's clinical reasoning capabilities are enhanced through its reinforcement learning framework, which enables emergent reasoning patterns, such as self-reflection and verification [5], contributing to its strong performance in clinical decision support tasks. However, although DeepSeek shows promising capabilities for clinical decision support, it cannot replace multidisciplinary tumor boards or human expertise, as it lacks contextual clinical judgment, physical examination capabilities, and the ability to negotiate complex trade-offs among specialists; instead, it streamlines clinical workflows by rapidly organizing patient data [41]. The integration of few-shot prompting has been shown to substantially enhance DeepSeek's accuracy in specialized tasks, such as Coronary Artery Disease Reporting and Data System (CAD-RADS) category assignment [42], suggesting that optimal prompt engineering is crucial for clinical implementation.

Overall, DeepSeek has emerged as a scalable tool to support treatment decisions, streamline workflows, and reduce diagnostic errors; however, integration requires careful validation and human oversight to mitigate risks.

DeepSeek in Medical Education and Benchmarking

Of 109 articles addressing the applications of DeepSeek in medical education and benchmarking, 93 were cross-sectional studies, 6 were retrospective studies, 5 were perspective studies, and 5 were descriptive studies.

On the Chinese National Medical Licensing Examination, DeepSeek-R1 achieved 92% accuracy, significantly outperforming ChatGPT-4o (87.2%) and demonstrating strength on low-difficulty questions [13]. Similarly, in the gastroenterology board examinations, both the base R1 model (77.1%) and search-augmented version (81.5%) surpassed the passing threshold and significantly outperformed the offline ChatGPT-3 (65.1%) and ChatGPT-4 (62.4%) models [45]. Cross-specialty comparisons revealed consistent patterns: In basic medical sciences, DeepSeek-R1 scored 78.33% alongside ChatGPT-4, whereas in clinical sciences, it scored 87.5%, demonstrating robust knowledge integration [46]. When evaluated against other reasoning-enhanced models on ophthalmology board-style questions, DeepSeek-R1 (72.5%) and its lighter variant R1-Lite (76.5%) performed competitively with OpenAI o1 Pro (83.4%), suggesting a balanced trade-off between performance and computational efficiency [47]. The model also demonstrated strong anatomical knowledge, achieving 89.2% accuracy on Turkish Dental Specialty Admission Exam anatomy questions, comparable with other major models, though below ChatGPT-4o's 98.6% [48]. These benchmark studies collectively indicate that DeepSeek provides a cost-effective, open-weight alternative for medical education, with utility in knowledge assessment

and examination preparation. However, performance gaps persist in specialized domains and image-based questions, highlighting areas for future development and the continued need for human oversight in comprehensive medical education frameworks.

DeepSeek for Clinical Workflow Optimization

A total of 63 articles described DeepSeek for clinical workflow optimization, including 26 cross-sectional studies, 2 descriptive studies, 17 retrospective studies, 4 prospective studies, 10 proof-of-concept studies, and 4 articles with other study designs.

The integration of DeepSeek models into health care systems offers significant potential to enhance operational efficiency and streamline clinical workflows, primarily by automating routine and time-consuming tasks. A prominent example is the locally deployed closed-loop system powered by DeepSeek for quality control of electronic nursing documentation. This system implements a comprehensive framework spanning the real-time, final, and vertical dimensions of quality assurance. The results include a dramatic reduction in documentation omission rates from 7.19% to just 1.79%; a decline in logical inconsistencies from 9.35% to 0.72%; and the complete elimination of timeliness errors, which previously stood at 8.63%. Concurrently, the quality control time per record decreased by 3.2-fold, reallocating nursing efforts toward direct patient care [6].

In dyslipidemia management, DeepSeek, alongside Claude-3 and GPT-4, optimized guideline-based workflows across 30 standardized cases, boosting accuracy from 72% for physicians to 91% with AI. Integration with human experts further raised simulated low-density lipoprotein cholesterol target attainment to 92%, demonstrating its utility in minimizing guideline deviations while enhancing workflow efficiency [49]. However, one moderate-quality study found that DeepSeek R1 achieved an accuracy of only 48.4% in a noncritical emergency department triage task, which is significantly lower than that of another LLM, Gemini 2.0 flash (73.8%) [50].

The large-scale deployment of DeepSeek across nearly 90 Chinese tertiary hospitals has reportedly increased patient follow-up efficiency 40-fold, marking a transformative impact on hospital administration and clinical workflow automation [9]. By managing labor-intensive tasks with high consistency and speed, DeepSeek enables a paradigm shift from reactive to proactive operational governance. This transition enabled health care professionals to focus their expertise on more complex clinical decision-making responsibilities.

Medical Research and Data Analysis

Medical research and data analysis were mentioned in 73 articles. Among these, 41 had a cross-sectional design, 6 were descriptive studies, 2 were perspective studies, 9 were proof-of-concept studies, 9 were retrospective studies, 1 had a mixed design, and the remaining 5 used other design types.

DeepSeek models have demonstrated significant utility in accelerating and refining medical research and data analysis workflows. DeepSeek facilitates the reading of medical literature, information extraction, and screening. Several studies have developed AI-powered screening tools using DeepSeek to identify relevant studies for systematic reviews, reporting high accuracy and a significant reduction in manual workload [51-53]. For example, the LitAutoScreener tool, which integrates DeepSeek, achieved high accuracy and significantly improved screening efficiency, reducing the processing time to seconds per article [51]. Similarly, other evaluations have confirmed that DeepSeek-based tools can reduce manual workload while maintaining high recall rates in literature screening for meta-analyses [53]. In fields such as aging research, DeepSeek-R1 is part of a multi-LLM ensemble that successfully extracts protocol details from clinical trial records, doubling the yield of conventional search methods and achieving expert-level accuracy for core data points [54]. Second, DeepSeek assists with generating and refining research topics and study designs. It helps researchers analyze cutting-edge trends, funding guidelines, and successful grant applications, thereby validating the novelty of the proposed research questions [3]. For instance, DeepSeek-R1 has been used to explore novel research ideas and generate systematic review topics in fields such as oral and maxillofacial surgery [55]. Similarly, in biomedical research, DeepSeek models show promise in extracting structured pre-analytical variability data from the scientific literature, facilitating standardized reporting and systematic evaluation [56]. Furthermore, DeepSeek serves as a valuable tool for peer review and for critiquing research proposals. Its capacity to generate high-quality evidence-based responses enables a preliminary assessment of a proposal's feasibility and soundness. This function is particularly beneficial in multidisciplinary contexts where the model's ability to synthesize information from diverse sources significantly enhances the evaluation process [57,58]. Third, DeepSeek demonstrated substantial potential as an assistant for drafting, editing, and refining the content of medical research papers. Its capabilities span various domains of medical research and practice, making it a versatile tool for enhancing the quality and efficiency of academic writing. The model's proficiency at generating structured, clear, and comprehensible content is particularly valuable in medical research, where precision and clarity are paramount [59].

Other Application Domains

In other application domains, 25 articles were identified, comprising 18 cross-sectional studies, 2 perspective articles, 2 descriptive studies, and 3 proof-of-concept studies.

Beyond the primary domains discussed, DeepSeek has been explored in several niche but critical areas, including treatment outcome prediction, drug development assistance, and suicide risk prediction. Instead of reactive question-answering, DeepSeek is integrated into predictive analytics platforms. It can proactively flag at-risk patients, suggest personalized screening intervals, and predict individual responses to therapies based on electronic health records and real-time data [60,61]. In nasopharyngeal carcinoma,

DeepSeek-V3-0324 demonstrated superior performance in treatment response evaluation compared with ChatGPT-4o-latest (96.5% vs 82.9%) and showed stronger agreement with expert annotations [62].

In drug discovery, DeepSeek aids with predicting drug-drug interactions and molecular property modeling, achieving superior performance in regression and classification tasks critical to drug discovery [63,64].

The model's chain-of-thought enabled analysis of factors associated with correct predictions, such as substance abuse and age-related comorbidities. This application underscores DeepSeek's potential for mental health risk assessment, though further validation is needed [65].

Discussion

Main Findings

This integrated bibliometric and scoping review provided a comprehensive early-stage mapping of the rapidly evolving research landscape concerning DeepSeek's applications in medicine. This field is characterized by explosive growth, global engagement, and exploration across a remarkably diverse spectrum of clinical and operational domains. The findings collectively underscore DeepSeek's emergence not merely as another LLM but as a potent, open-weight contender with specific capabilities that address critical needs in modern health care, including cost-effectiveness, linguistic accessibility, and scalability.

Bibliometric data showed a research frontier that has been intensively explored. The increased publication output regarding applications of DeepSeek in medicine is clear. Our results align with those of an analysis of the global research profile of another LLM, ChatGPT, conducted by Alessandri-Bonetti et al [66], who revealed explosive growth in publications during the first 7 months after its release. This pattern is also consistent with a broader LLM systematic review by Chen et al [67], which reported that, between January 2022 and September 2025, approximately 3.2 clinical LLM studies were published per day, with a linear increase of 7.04 studies per month following the release of ChatGPT. Notably, DeepSeek was not included in the analysis by Chen et al [67], underscoring the gap and the need for our focused review.

The geographical and institutional productivity led by China, followed by Turkey and the United States, reflects widespread international interest of DeepSeek's potential, with major academic medical centers driving early investigations. Papers on DeepSeek's applications in medicine have been published in various journals, ranging from well-known open-access journals such as *Cureus* and *Scientific Reports* to professional medical informatics and medical education journals such as the *Journal of Medical Internet Research*. This publication pattern indicates that the research reaches both broad scientific and specialized clinical audiences. Keyword co-occurrence analysis effectively identified the core themes of this research trend. The temporal overlay,

which revealed a shift from foundational medical education topics toward more specialized areas such as “retrieval-augmented generation” and “oncology,” illustrates the field’s rapid maturation and deepening focus. Synthesizing the scoping review findings, DeepSeek as a medical tool initially gained attention for its strength in democratizing medical information. For instance, in patient education, it can generate outputs with higher readability than its counterparts, such as ChatGPT.

Perhaps the most striking finding is that DeepSeek has demonstrated competitive and sometimes superior performance compared with existing proprietary models in clinical decision support tasks. The bibliometric analysis revealed that “clinical decision support” formed the largest cluster, while the scoping review further indicated that these studies primarily focused on three specific tasks: “aiding diagnosis,” “differential diagnosis,” and “treatment plan formulation.” The evidence that DeepSeek-V3 can match senior radiologists at specialized diagnostic classifications or that DeepSeek-R1 rivals GPT-4 and OpenAI o1 in diagnostic accuracy across ophthalmology and complex clinicopathological cases challenges the assumption that superior capability is the exclusive domain of closed, commercial models. This “performance parity” achieved through an open-weight architecture has profound implications. Specifically, it suggests a pathway toward breaking the monopoly of advanced AI in clinical support, potentially fostering innovation, reducing costs, and allowing for better adaptation to local health care contexts and linguistic needs.

The utility of this model in medical education and benchmarking further supports its position as a disruptive and cost-effective tool [68]. For institutions and learners worldwide, particularly in resource-constrained settings, DeepSeek offers a viable, high-quality alternative for exam preparation, simulation, and curriculum development, potentially lowering the barriers to accessing advanced medical training aids.

Beyond its direct clinical and educational applications, this review highlighted DeepSeek’s transformative potential across broader health care operations. Documented case studies have demonstrated reductions in documentation error rates in nursing and lower specimen return rates in gynecological examinations and enabled large-scale patient follow-up. By automating a vast array of low-complexity tasks, DeepSeek can free human resources to provide higher quality care and reduce systemic inefficiencies across the health care continuum.

Of the 353 papers included in this scoping review, only 6.8% (24/353) met the criteria for high quality, whereas the majority (235/353, 66.6%) were classified as low quality, consisting predominantly of invalidated benchmarking using examination questions, single-center convenience samples, and proof-of-concept studies. This distribution reflects a critical gap in the current literature: The rapid proliferation of DeepSeek in medicine has been accompanied by an abundance of exploratory studies with limited external validity. Although such benchmarking studies offer valuable insights

into the model’s technical capabilities and serve as initial performance indicators, they do not directly inform real-world diagnostic accuracy, patient safety, or clinical utility [69].

In head-to-head comparisons with other LLMs, DeepSeek demonstrated predominantly favorable or comparable performance: Positive outcomes (126/283, 44.5%) were more frequent than negative ones (73/283, 25.8%), and a substantial proportion of studies (84/283, 29.7%) showed no clear superiority of either model. However, because these results derived predominantly from low-quality (235/353, 66.6%) or moderate-quality evidence, with only 6.8% (24/353) meeting high methodological standards, performance claims should be considered preliminary and hypothesis-generating rather than definitive. Clinically, it excels in open-source accessibility, low cost, readability, Chinese language proficiency, and structured reasoning; nonetheless, limitations, including occasional inaccuracies, lower reliability in certain tasks, and the absence of prospective clinical trials, necessitate continued validation and human oversight.

Comparison With Prior Reviews on Other LLMs in Medicine

To contextualize the novel and distinct contributions of our work, we compared this review with existing reviews of other LLMs in medicine, such as ChatGPT, GPT-4, LLaMA, and Gemini. Several prior reviews have documented the rapid adoption of proprietary LLMs in health care, highlighting their utility in clinical reasoning, medical education, and patient communication [66,70,71]. However, most existing reviews have primarily focused on closed-source models, which are characterized by limited transparency, restricted capacity for local deployment, and substantial cost barriers. These limitations hinder their scalability and reduce their adaptability across diverse institutional settings. In contrast, this review specifically focused on DeepSeek, an open-weight LLM, and identified several distinctive features that differentiate it from the patterns reported in previous LLM reviews.

First, methodologically, we combined bibliometric analysis with a scoping review to provide both quantitative mapping of research trends and qualitative synthesis of applications and challenges of DeepSeek in medicine, a dual approach rarely applied in prior LLM reviews, which have tended to rely on either bibliometric or narrative synthesis alone [66,71].

Second, geographically, the research landscapes differ substantially. For ChatGPT, early publications were predominantly led by institutions in the United States and Europe, with a wide distribution across high-income countries [66,70]. In contrast, our analysis identified China as the dominant contributor to DeepSeek medical research (163 papers), followed by Turkey and the United States. This pattern aligns with DeepSeek’s country of origin and its rapid deployment across Chinese tertiary hospitals [2,9]. Notably, the early and substantial involvement of Turkish researchers (52 papers) in DeepSeek research is a distinctive feature not observed in early ChatGPT literature.

Third, previous reviews focused predominantly on proprietary models such as ChatGPT, GPT-4, LLaMA, and Gemini. In contrast, our study addressed a significant gap by examining an open-source alternative with distinct architectural advantages and greater deployment flexibility. In terms of real-world deployment, the deployment of DeepSeek across nearly 90 tertiary hospitals in China has resulted in measurable improvements in workflow efficiency and documentation quality. This scale of implementation has not been reported in similar reviews of other LLMs, which have largely focused on simulated or benchmarking studies [6,9]. In terms of application areas, prior work on ChatGPT and other proprietary LLMs identified medical education, clinical decision support, and patient communication as core areas [70,71]. Our keyword co-occurrence analysis confirmed that these are also central themes for DeepSeek. However, DeepSeek's open-weight architecture introduces distinctive features not emphasized in proprietary LLM reviews: on-premises deployability, data privacy, cost-effectiveness, and superior performance in Chinese-language medical tasks. These features represent unique contributions of DeepSeek to the medical LLM landscape and are not simply typical of any newly introduced LLM. Regarding performance and utility, our findings demonstrated that DeepSeek achieved competitive or superior performance compared with proprietary models in clinical diagnostics, medical licensing examinations, and patient education while substantially reducing costs, advantages that prior reviews have identified as critical unmet needs in AI integration [71, 72].

Challenges in the Applications of DeepSeek in Medicine

Guided by an ethical framework, the efficacy and safety of any medical intervention must be carefully calibrated in modern medical practice [73]. As aforementioned, DeepSeek demonstrates significant potential for enhancing medical workflows, medical education, and research. However, its application faces numerous challenges in terms of effectiveness and safety, including accuracy issues, data privacy concerns, ethical uncertainties, and diverse global regulations governing AI.

Accuracy and Variable Performance Across Medical Domains and Specialties

Although DeepSeek has demonstrated diagnostic accuracy comparable to that of specialist clinicians and proprietary models in certain areas [40,74-76], its overall efficacy remains inconsistent [42,77,78]. The model exhibits strong zero-shot and few-shot learning capabilities in general tasks; however, the rapid evolution of medical knowledge necessitates continuous pretraining on extensive volumes of high-quality, domain-specific data. In data-scarce specialties, particularly those lacking sufficient fine-tuning datasets, DeepSeek often fails to effectively acquire new features and patterns, leading to model hallucinations, defined as the generation of seemingly plausible but factually incorrect or unsupported information [36,79]. Such limitations

are particularly severe in domains involving rare diseases and complex, nonclassical clinical scenarios, where available pretraining data are often insufficient and clinically unvalidated [36,37,80,81]. Furthermore, as a fundamentally text-based model, DeepSeek exhibits inherent limitations in processing specialized nontextual medical data, such as medical images, complex laboratory metrics, and genomic data [74, 82-84]. These constraints collectively contribute to inconsistent model performance across specific medical domains and hinder its generalization.

Ethical and Safety Risks

The integration of DeepSeek into medical practice raises ethical challenges that implicate all 4 foundational principles of biomedical ethics, namely autonomy, nonmaleficence, beneficence, and justice, which were originally proposed by Beauchamp and Childress in 1979 [85].

Autonomy: Challenges to Patient Self-Determination and Informed Consent

The application of DeepSeek in medicine may undermine the principle of autonomy in medical ethics. As an open-source model, DeepSeek can be deployed on-premises in a hospital environment, which facilitates compliance with data privacy requirements [1,9,81]. However, its broader adoption is complicated by varying regulatory frameworks across regions, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) [3]. The Italian data protection authority, for instance, has restricted DeepSeek over concerns that its data handling methods fail to meet the strict privacy rules of the European Union [81]. Although techniques such as chain-of-thought have enhanced the interpretability of decision-making, the model's fundamental "black-box" nature persists, posing practical challenges to informed consent in clinical applications [60,86-90].

Nonmaleficence: Risks of Novel and Amplified Harms

The rapid, cost-effective integration of DeepSeek in Chinese hospitals underscores a central paradox in medicine: how to seize the opportunity for transformative innovation while mitigating the risks of undue haste and still upholding the principle of "first, do no harm" [2]. However, this model may provide overly definitive recommendations, potentially suggesting unnecessary tests or harmful treatments without adequate contextual warnings [91,92]. If clinicians over-rely on AI outputs, effectively delegating core cognitive tasks such as comprehensive analysis, differential diagnosis, and clinical judgment to the machine, it may lead to the erosion of clinical skills and their independent clinical reasoning. Furthermore, however data-driven its suggestions may be, DeepSeek may lack the nuanced and holistic understanding of a patient's psychosocial context that an experienced physician integrates. Collectively, these issues challenge the ethical principle of nonmaleficence.

Beneficence: The Challenge of Defining and Delivering “Good”

The principle of beneficence obligates health care providers to act in ways that promote patients' well-being and enhance clinical outcomes [93]. However, an emphasis on AI-driven efficiency may unintentionally marginalize the irreplaceable human dimensions of medicine, such as empathy, compassion, and the therapeutic physician-patient relationship. Although systems like DeepSeek are adept at optimizing measurable, data-informed endpoints, the concept of “good” in medical practice encompasses psychosocial, spiritual, and qualitative aspects of care that resist easy quantification [89,94]. Overreliance on algorithmic pathways designed to maximize metrics neglects the holistic components of beneficence [95]. Consequently, the physician's role as a compassionate interpreter of illness, which lies at the heart of medical beneficence, may be subordinate to the pursuit of algorithmic efficiency.

Justice: Amplifying Inequities in Algorithmic Health Care

The principle of justice concerns fair and equitable distribution of health care benefits and burdens. Despite the use of data preprocessing techniques and fairness-aware algorithms, DeepSeek can still perpetuate and potentially amplify societal or health care biases present in its historical medical training data, including the underdiagnosis of certain conditions within specific demographic groups, thereby harming marginalized populations [80,88,96]. Furthermore, because DeepSeek's training framework is primarily optimized for English and Chinese, it carries inherent lexical and cultural biases that may limit its applicability to global health care contexts [12,35,97]. Additionally, the benefits of advanced AI, such as DeepSeek, are likely to accrue disproportionately to well-resourced tertiary-care urban hospitals equipped with the necessary infrastructure and specialized personnel for local deployment. Such unequal access exacerbates existing health disparities across regions and socioeconomic groups.

Other Challenges

In addition to challenges such as accuracy, variable performance across medical domains and specialties, and medical ethics and safety issues, the application of DeepSeek in medicine faces other obstacles, including the redesign of clinical workflows, delineation of liability, regulatory lag, and trust and adoption. The deployment of DeepSeek challenges some clinicians' work habits and creates a demand for professionals who understand both clinical practice and AI. A shortage of talent limits its wider adoption. When errors in DeepSeek-assisted decision-making lead to medical incidents, how should legal responsibility be defined? Should it fall on the operating physician, the hospital that adopted the AI, or the model developers? Currently, global regulations in this field generally lag, and this uncertainty greatly dampens hospitals' willingness to implement such technologies. Trust remains another challenge; although DeepSeek is easy to use, concerns about risks affect its acceptance [87].

Future Work in the Applications of DeepSeek in Medicine

Based on the aforementioned challenges, future research and development should prioritize the directions highlighted in the following sections to advance the reliable, ethical, and equitable integration of DeepSeek into medical practice.

From Benchmarking to Clinical Validation: Prospective and Pragmatic Studies

The current evidence base is dominated by low-quality, simulation-based studies. Future work should move beyond examination-style benchmarks and retrospective analyses toward prospective, multicenter, and pragmatic clinical trials. Specifically, RCTs are urgently needed to compare DeepSeek-assisted care against standard practice using both proximal performance metrics, such as diagnostic accuracy, and patient-relevant outcomes, including treatment adherence, adverse events, and quality of life [98,99]. Such trials should also evaluate human-AI interaction models, for example, human-in-the-loop versus fully automated approaches, to determine the optimal balance between efficiency and safety [100,101]. Furthermore, real-world implementation science frameworks should be applied to assess scalability, usability, and unintended consequences across diverse health care settings.

Strengthening Governance, Explainability, and Safety

To address ethical and regulatory gaps, future work should co-develop clinically interpretable explainability methods tailored to DeepSeek's reasoning architecture. Techniques such as structured audit trails, uncertainty quantification, and natural language rationales can support informed consent and clinician oversight [89,102]. On the governance front, clear liability and accountability frameworks are required to delineate responsibilities among developers, health care institutions, and clinicians when AI-assisted errors occur [88,96]. Additionally, the “human-in-command” principle, which mandates that DeepSeek's recommendations serve as decision support rather than replacement for clinician judgment, should be embedded into clinical workflows and professional guidelines [98,103]. As articulated in the concept of AI-assisted medicine introduced by Wang et al [104], a discipline that uses AI technologies to assist with disease research, prevention, diagnosis, and treatment as well as to promote health maintenance, clinicians must retain ultimate decision-making authority and accountability [100,101]. This conceptual foundation reinforces that AI remains a tool to augment, not supplant, human expertise.

Mitigating Bias and Promoting Equitable Access

Despite DeepSeek's open-weight advantage, bias and inequity remain critical challenges. Future research should conduct systematic bias audits across demographic subgroups such as sex, socioeconomic status, and ethnicity using multi-institutional and multilingual datasets [105,106]. To

avoid perpetuating health care disparities, developers should expand medically validated support beyond English and Chinese to other major world languages while adapting outputs to local clinical guidelines and cultural contexts [12, 107].

Redefining Medical Education and Workforce Development

The rapid adoption of DeepSeek demands a parallel evolution in medical curricula. Future educational interventions should cultivate “AI literacy”: the ability to critically appraise AI-generated recommendations; recognize hallucinations and bias; and integrate AI outputs with compassionate, patient-centered communication [98,108]. Institutions should develop interdisciplinary training programs that bridge clinical practice and data science to build a workforce capable of deploying, auditing, and improving medical AI systems. Finally, professional societies should establish certification and continuing education standards for AI-augmented clinical practice.

Unexplored Domains and Long-Term Monitoring

Most current research focuses on diagnosis, medical education, and workflow efficiency, leaving prevention and long-term care underexplored. Future investigations should prioritize disease prevention, population health management, and long-term care [103,109]. Additionally, postdeployment surveillance systems should be established to monitor real-world performance, detect emergent harms, and enable continuous model improvement, closing the loop from evidence generation to sustained safe implementation [9,90].

Limitations of the Study

Several limitations of this study should be considered when interpreting the findings. First, the review covered literature published over a relatively short and recent timeframe. Consequently, the observed surge in publications may reflect early enthusiasm rather than sustained scientific progress. Second, although a language-agnostic search strategy was used, most included studies were published in English, with only a small number ($n=13$) in Chinese. This linguistic imbalance, coupled with the predominance of contributions from researchers based in China, indicates a notable geographical concentration of the available evidence. As a result, the findings may not be directly generalizable to health care systems operating within different regulatory, cultural, or infrastructural contexts. Third, the included studies exhibited substantial heterogeneity in methodologies, medical specialties, evaluation metrics, comparator models, and DeepSeek model versions—for example, R1 versus V3, which differ in parameter counts, training data, and reasoning depth. This variability precluded quantitative

synthesis of outcomes and hindered direct cross-study comparisons. Although we reported version-specific findings where available, direct comparisons of performance should be interpreted with caution. Future research should adopt standardized version reporting and benchmark against fixed model checkpoints to enhance comparability and reproducibility. Finally, much of the evidence is derived from benchmarking studies, simulated cases, or retrospective analyses, with a formal quality appraisal showing that 66.6% (235/353) of included original articles were of low quality and only 6.8% (24/353) met the criteria to be considered high quality. Prospective clinical trials or RCTs assessing DeepSeek’s impact on tangible patient health outcomes in real-world clinical settings remain notably scarce. Consequently, the overall quality of the evidence base is inherently preliminary, and the reviewed corpus carries a high risk of bias. The reported strengths of DeepSeek should be interpreted with caution, as these findings predominantly derive from low-quality, controlled, nongeneralizable settings.

Conclusion

This integrated bibliometric and scoping review synthesized the available evidence on DeepSeek’s applications in medicine. The bibliometric analysis revealed a progressive increase in publication output from January 2025 through November 2025, with China, Turkey, and the United States as the leading contributors. Keyword co-occurrence analysis formed 7 clusters; the 3 most frequent keywords were “large language model,” “artificial intelligence,” and “patient education.”

The scoping review found that DeepSeek has been evaluated across 5 primary application domains: patient education and communication, clinical decision support and treatment planning, medical education and benchmarking, clinical workflow optimization, and medical research and data analysis. In these domains, DeepSeek demonstrated variable but often competitive performance compared with proprietary models, with documented strengths in readability of patient education materials, diagnostic accuracy in select specialties, cost-efficiency, and local deployability. Nevertheless, it should be noted that most included studies were of moderate or low quality, and the evidence base is predominantly composed of benchmarking and simulation studies, with a notable scarcity of prospective clinical trials or RCTs assessing patient-relevant outcomes. Additionally, the review identified consistent limitations, including variable performance across medical specialties, model hallucinations, ethical concerns, data privacy challenges, and regulatory gaps. Future integration will require robust prospective clinical validation, expansion of multimodal capabilities, bias mitigation strategies, human-in-the-loop governance frameworks, and equitable access strategies.

Acknowledgments

The authors would like to acknowledge Editage for English language editing [110].

Funding

This work was supported by the Science and Technology Project of Jinan Health Commission (grant 2020-3-02).

Data Availability

All data generated or analyzed during this study are included in this published article and its multimedia appendices.

Authors' Contributions

Conceptualization: GW

Data curation: HZ, DW

Formal analysis: YX, SH

Funding acquisition: GW

Methodology: HZ, DW

Resources: HZ, DW, YX, SH

Software: HZ, DW, GW

Supervision: GW

Visualization: HZ, SH

Writing – original draft: HZ, DW, YX, SH

Writing – review & editing: GW

All the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[\[DOC File \(Microsoft Word File\), 49 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Quality assessment criteria for studies included in the scoping review.

[\[DOC File \(Microsoft Word File\), 34 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

The extracted data for scoping review.

[\[XLSX File \(Microsoft Excel File\), 246 KB-Multimedia Appendix 3\]](#)

Checklist 1

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.

[\[DOCX File \(Microsoft Word File\), 74 KB-Checklist 1\]](#)

References

1. Sandmann S, Hegselmann S, Fujarski M, et al. Benchmark evaluation of DeepSeek large language models in clinical decision-making. *Nat Med*. Aug 2025;31(8):2546-2549. [doi: [10.1038/s41591-025-03727-2](https://doi.org/10.1038/s41591-025-03727-2)] [Medline: [40267970](https://pubmed.ncbi.nlm.nih.gov/40267970/)]
2. Zeng D, Qin Y, Sheng B, Wong TY. DeepSeek's "low-cost" adoption across China's hospital systems: too fast, too soon? *JAMA*. Jun 3, 2025;333(21):1866-1869. [doi: [10.1001/jama.2025.6571](https://doi.org/10.1001/jama.2025.6571)] [Medline: [40293869](https://pubmed.ncbi.nlm.nih.gov/40293869/)]
3. MohanaSundaram A, Sathanantham ST, Ivanov A, Mofatteh M. DeepSeek's readiness for medical research and practice: prospects, bottlenecks, and global regulatory constraints. *Ann Biomed Eng*. Jul 2025;53(7):1754-1756. [doi: [10.1007/s10439-025-03738-7](https://doi.org/10.1007/s10439-025-03738-7)] [Medline: [40272697](https://pubmed.ncbi.nlm.nih.gov/40272697/)]
4. Jin I, Tangsrivimol JA, Darzi E, et al. DeepSeek vs. ChatGPT: prospects and challenges. *Front Artif Intell*. 2025;8:1576992. [doi: [10.3389/frai.2025.1576992](https://doi.org/10.3389/frai.2025.1576992)] [Medline: [40612384](https://pubmed.ncbi.nlm.nih.gov/40612384/)]
5. Guo D, Yang D, Zhang H, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature New Biol*. Sep 2025;645(8081):633-638. [doi: [10.1038/s41586-025-09422-z](https://doi.org/10.1038/s41586-025-09422-z)] [Medline: [40962978](https://pubmed.ncbi.nlm.nih.gov/40962978/)]
6. Lv J, Xu Y, Jiang M, et al. A DeepSeek-powered locally deployed closed-loop system for enhancing quality control in electronic nursing documentation: development and clinical validation. *J Am Med Inform Assoc*. Oct 1, 2025;32(10):1526-1532. [doi: [10.1093/jamia/ocaf109](https://doi.org/10.1093/jamia/ocaf109)] [Medline: [40668938](https://pubmed.ncbi.nlm.nih.gov/40668938/)]
7. Wang Y, Tan W, Cheng S, et al. Large language model agent for managing patients with suspected hypertension. *Hypertension*. Jan 2026;83(1):212-224. [doi: [10.1161/HYPERTENSIONAHA.125.25305](https://doi.org/10.1161/HYPERTENSIONAHA.125.25305)] [Medline: [41064862](https://pubmed.ncbi.nlm.nih.gov/41064862/)]

8. Miao Y, Wen J, Luo Y, Li J. MedARC: Adaptive multi-agent refinement and collaboration for enhanced medical reasoning in large language models. *Int J Med Inform.* Feb 2026;206:106136. [doi: [10.1016/j.ijmedinf.2025.106136](https://doi.org/10.1016/j.ijmedinf.2025.106136)] [Medline: [41109093](https://pubmed.ncbi.nlm.nih.gov/41109093/)]
9. Chen J, Miao C. DeepSeek deployed in 90 Chinese tertiary hospitals: how artificial intelligence is transforming clinical practice. *J Med Syst.* Apr 24, 2025;49(1):53. [doi: [10.1007/s10916-025-02181-4](https://doi.org/10.1007/s10916-025-02181-4)] [Medline: [40272650](https://pubmed.ncbi.nlm.nih.gov/40272650/)]
10. Chan L, Xu X, Lv K. DeepSeek-R1 and GPT-4 are comparable in a complex diagnostic challenge: a historical control study. *Int J Surg.* 2025;111(6):4056-4059. [doi: [10.1097/JS9.0000000000002386](https://doi.org/10.1097/JS9.0000000000002386)]
11. Jiao C, Rosas E, Asadigandomani H, et al. Diagnostic performance of publicly available large language models in corneal diseases: a comparison with human specialists. *Diagnostics (Basel).* May 13, 2025;15(10):1221. [doi: [10.3390/diagnostics15101221](https://doi.org/10.3390/diagnostics15101221)] [Medline: [40428214](https://pubmed.ncbi.nlm.nih.gov/40428214/)]
12. 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/)]
13. Wu J, Wang Z, Qin Y. Performance of DeepSeek-R1 and ChatGPT-4o on the Chinese National Medical Licensing Examination: a comparative study. *J Med Syst.* Jun 3, 2025;49(1):74. [doi: [10.1007/s10916-025-02213-z](https://doi.org/10.1007/s10916-025-02213-z)] [Medline: [40459679](https://pubmed.ncbi.nlm.nih.gov/40459679/)]
14. Huang Y, Wan Y, Chen J, Qin M, Wang J, Liang H. Knowledge mapping of biomarkers in amyotrophic lateral sclerosis: a comprehensive bibliometric and visual analysis. *Neurodegener Dis Manag.* Apr 2026;16(2):191-207. [doi: [10.1080/17582024.2025.2554525](https://doi.org/10.1080/17582024.2025.2554525)] [Medline: [40905501](https://pubmed.ncbi.nlm.nih.gov/40905501/)]
15. Chen X, Yang Y, Yun D, et al. Current status and solutions for AI ethics in ophthalmology: a bibliometric analysis. *NPJ Digit Med.* Oct 2, 2025;8(1):594. [doi: [10.1038/s41746-025-01976-6](https://doi.org/10.1038/s41746-025-01976-6)]
16. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci.* Sep 20, 2010;5:69. [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]
17. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. Science mapping software tools: review, analysis, and cooperative study among tools. *J Am Soc Inf Sci.* Jul 2011;62(7):1382-1402. URL: <http://doi.wiley.com/10.1002/asi.v62.7> [doi: [10.1002/asi.21525](https://doi.org/10.1002/asi.21525)]
18. Montazeri A, Mohammadi S, M Hesari P, Ghaemi M, Riazi H, Sheikhi-Mobarakeh Z. Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. *Syst Rev.* Dec 15, 2023;12(1):239. [doi: [10.1186/s13643-023-02410-2](https://doi.org/10.1186/s13643-023-02410-2)] [Medline: [38102710](https://pubmed.ncbi.nlm.nih.gov/38102710/)]
19. 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)]
20. 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/)]
21. Gibney E. China's cheap, open AI model DeepSeek thrills scientists. *Nature New Biol.* Feb 6, 2025;638(8049):13-14. [doi: [10.1038/d41586-025-00229-6](https://doi.org/10.1038/d41586-025-00229-6)]
22. Conroy G, Mallapaty S. How China created AI model DeepSeek and shocked the world. *Nature New Biol.* Feb 13, 2025;638(8050):300-301. [doi: [10.1038/d41586-025-00259-0](https://doi.org/10.1038/d41586-025-00259-0)]
23. Zhou M, Pan Y, Zhang Y, Song X, Zhou Y. Evaluating AI-generated patient education materials for spinal surgeries: comparative analysis of readability and DISCERN quality across ChatGPT and DeepSeek models. *Int J Med Inform.* Jun 2025;198:105871. [doi: [10.1016/j.ijmedinf.2025.105871](https://doi.org/10.1016/j.ijmedinf.2025.105871)] [Medline: [40107040](https://pubmed.ncbi.nlm.nih.gov/40107040/)]
24. Deng Z, Ma W, Han QL, et al. Exploring DeepSeek: a survey on advances, applications, challenges and future directions. *IEEE/CAA J Autom Sinica.* May 2025;12(5):872-893. [doi: [10.1109/JAS.2025.125498](https://doi.org/10.1109/JAS.2025.125498)]
25. Kaygisiz ÖF, Teke MT. Can DeepSeek and ChatGPT be used in the diagnosis of oral pathologies? *BMC Oral Health.* Apr 25, 2025;25(1):638. [doi: [10.1186/s12903-025-06034-x](https://doi.org/10.1186/s12903-025-06034-x)] [Medline: [40281436](https://pubmed.ncbi.nlm.nih.gov/40281436/)]
26. Rasool A, Shahzad MI, Aslam H, Chan V, Arshad MA. Emotion-aware embedding fusion in large language models (Flan-T5, Llama 2, DeepSeek-R1, and ChatGPT 4) for intelligent response generation. *AI.* Mar 13, 2025;6(3):56. [doi: [10.3390/ai6030056](https://doi.org/10.3390/ai6030056)]
27. Yilmaz BE, Gokkurt Yilmaz BN, Ozbey F. Artificial intelligence performance in answering multiple-choice oral pathology questions: a comparative analysis. *BMC Oral Health.* Apr 15, 2025;25(1):573. [doi: [10.1186/s12903-025-05926-2](https://doi.org/10.1186/s12903-025-05926-2)] [Medline: [40234873](https://pubmed.ncbi.nlm.nih.gov/40234873/)]
28. Marcaccini G, Seth I, Xie Y, et al. Breaking bones, breaking barriers: ChatGPT, DeepSeek, and Gemini in hand fracture management. *J Clin Med.* Mar 14, 2025;14(6):1983. [doi: [10.3390/jcm14061983](https://doi.org/10.3390/jcm14061983)] [Medline: [40142791](https://pubmed.ncbi.nlm.nih.gov/40142791/)]
29. 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/)]

30. Gültekin O, Inoue J, Yilmaz 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/)]
31. Seth I, Marcaccini G, Lim K, et al. Management of Dupuytren's disease: a multi-centric comparative analysis between experienced hand surgeons versus artificial intelligence. *Diagnostics (Basel)*. Feb 28, 2025;15(5):587. [doi: [10.3390/diagnostics15050587](https://doi.org/10.3390/diagnostics15050587)] [Medline: [40075834](https://pubmed.ncbi.nlm.nih.gov/40075834/)]
32. VOSviewer. URL: <https://www.vosviewer.com/> [Accessed 2026-06-10]
33. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2010;84:523-538. [doi: [10.1007/s11192-009-0146-3](https://doi.org/10.1007/s11192-009-0146-3)] [Medline: [20585380](https://pubmed.ncbi.nlm.nih.gov/20585380/)]
34. Lau JYS, Gerald Sng GR, Cao R, Chen J. A comparative study of ChatGPT and DeepSeek in spinal cord injury patient education: can artificial intelligence "speak" spinal cord injury? *J Spinal Cord Med*. May 2026;49(3):618-623. [doi: [10.1080/10790268.2025.2554013](https://doi.org/10.1080/10790268.2025.2554013)] [Medline: [40938207](https://pubmed.ncbi.nlm.nih.gov/40938207/)]
35. Liu Y, Yu F, Zhang X, et al. Assessing the role of large language models between ChatGPT and DeepSeek in asthma education for bilingual individuals: comparative study. *JMIR Med Inform*. Aug 13, 2025;13:e65365. [doi: [10.2196/65365](https://doi.org/10.2196/65365)] [Medline: [40802989](https://pubmed.ncbi.nlm.nih.gov/40802989/)]
36. Uldin H, Saran S, Gandikota G, et al. A comparison of performance of DeepSeek-R1 model-generated responses to musculoskeletal radiology queries against ChatGPT-4 and ChatGPT-4o - a feasibility study. *Clin Imaging*. Jul 2025;123:110506. [doi: [10.1016/j.clinimag.2025.110506](https://doi.org/10.1016/j.clinimag.2025.110506)] [Medline: [40381536](https://pubmed.ncbi.nlm.nih.gov/40381536/)]
37. Wu H, Yao S, Bao H, Guo Y, Xu C, Ma J. ChatGPT-4.0 and DeepSeek-R1 does not yet provide clinically supported answers for knee osteoarthritis. *Knee*. Oct 2025;56:386-396. [doi: [10.1016/j.knee.2025.06.007](https://doi.org/10.1016/j.knee.2025.06.007)] [Medline: [40618549](https://pubmed.ncbi.nlm.nih.gov/40618549/)]
38. Alluri AA, Khan Z, Krithika V, et al. Assessing the suitability of ChatGPT and DeepSeek AI for patient education on common rheumatological disorders. *Cureus*. Aug 2025;17(8):e90600. [doi: [10.7759/cureus.90600](https://doi.org/10.7759/cureus.90600)] [Medline: [40984935](https://pubmed.ncbi.nlm.nih.gov/40984935/)]
39. Gurbuz S, Bahar H, Yavuz U, Keskin A, Karslioglu B, Solak Y. Comparative efficacy of ChatGPT and DeepSeek in addressing patient queries on gonarthrosis and total knee arthroplasty. *Arthroplast Today*. Jun 2025;33:101730. [doi: [10.1016/j.artd.2025.101730](https://doi.org/10.1016/j.artd.2025.101730)] [Medline: [40521295](https://pubmed.ncbi.nlm.nih.gov/40521295/)]
40. Zhang J, Liu J, Guo M, Zhang X, Xiao W, Chen F. DeepSeek-assisted LI-RADS classification: AI-driven precision in hepatocellular carcinoma diagnosis. *Int J Surg*. 2025;111(9):5970-5979. [doi: [10.1097/JS9.0000000000002763](https://doi.org/10.1097/JS9.0000000000002763)]
41. Vural Camalan B, Doluoglu S, Taraf NH, Gunay MM, Ozlugedik S. ChatGPT versus DeepSeek in head and neck cancer staging and treatment planning: guideline-based study. *Eur Arch Otorhinolaryngol*. Sep 2025;282(9):4815-4824. [doi: [10.1007/s00405-025-09524-4](https://doi.org/10.1007/s00405-025-09524-4)] [Medline: [40523995](https://pubmed.ncbi.nlm.nih.gov/40523995/)]
42. Mikhail D, Farah A, Milad J, et al. DeepSeek-R1 vs OpenAI o1 for ophthalmic diagnoses and management plans. *JAMA Ophthalmol*. Oct 1, 2025;143(10):834-842. [doi: [10.1001/jamaophthalmol.2025.2918](https://doi.org/10.1001/jamaophthalmol.2025.2918)] [Medline: [40906471](https://pubmed.ncbi.nlm.nih.gov/40906471/)]
43. Cao H, Hao C, Zhang T, et al. Battle of the artificial intelligence: a comprehensive comparative analysis of DeepSeek and ChatGPT for urinary incontinence-related questions. *Front Public Health*. 2025;13:1605908. [doi: [10.3389/fpubh.2025.1605908](https://doi.org/10.3389/fpubh.2025.1605908)] [Medline: [40771241](https://pubmed.ncbi.nlm.nih.gov/40771241/)]
44. Pinto VBP, Ataídes RJC, do Nascimento LAP, et al. Performance of ChatGPT and DeepSeek in the management of postprostatectomy urinary incontinence. *Int Braz J Urol*. 2025;51(6):e20250325. [doi: [10.1590/S1677-5538.IBJU.2025.0325](https://doi.org/10.1590/S1677-5538.IBJU.2025.0325)] [Medline: [40857549](https://pubmed.ncbi.nlm.nih.gov/40857549/)]
45. Ibrahim AF, Danpanichkul P, Hayek A, et al. Artificial intelligence in gastroenterology education: DeepSeek passes the gastroenterology board examination and outperforms legacy ChatGPT models. *Am J Gastroenterol*. Apr 1, 2026;121(4):1041-1043. [doi: [10.14309/ajg.0000000000003552](https://doi.org/10.14309/ajg.0000000000003552)] [Medline: [40392256](https://pubmed.ncbi.nlm.nih.gov/40392256/)]
46. Meo SA, Abukhalaf FA, ElToukhy RA, Sattar K. Exploring the role of DeepSeek-R1, ChatGPT-4, and Google Gemini in medical education: how valid and reliable are they? *Pak J Med Sci*. Jul 2025;41(7):1887-1892. [doi: [10.12669/pjms.41.7.12183](https://doi.org/10.12669/pjms.41.7.12183)] [Medline: [40735572](https://pubmed.ncbi.nlm.nih.gov/40735572/)]
47. Shean R, Shah T, Pandiarajan A, et al. A comparative analysis of DeepSeek R1, DeepSeek-R1-Lite, OpenAi o1 Pro, and Grok 3 performance on ophthalmology board-style questions. *Sci Rep*. Jul 2, 2025;15(1):23101. [doi: [10.1038/s41598-025-08601-2](https://doi.org/10.1038/s41598-025-08601-2)] [Medline: [40595291](https://pubmed.ncbi.nlm.nih.gov/40595291/)]
48. Tassoker M. Who knows anatomy best? A comparative study of ChatGPT-4o, DeepSeek, Gemini, and Claude. *Clin Anat*. Jan 2026;39(1):25-29. [doi: [10.1002/ca.70012](https://doi.org/10.1002/ca.70012)] [Medline: [40708277](https://pubmed.ncbi.nlm.nih.gov/40708277/)]
49. Ucdal M, Yurtsever K, Yildiz P, Akalin A, Mert KU, Guven GS. Comparison of artificial intelligence models and human experts in managing dyslipidemia: assessment of adherence to clinical guidelines. *Cureus*. Aug 2025;17(8):e91363. [doi: [10.7759/cureus.91363](https://doi.org/10.7759/cureus.91363)] [Medline: [40904968](https://pubmed.ncbi.nlm.nih.gov/40904968/)]
50. Lee S, Jung S, Park JH, Cho H, Moon S, Ahn S. Performance of ChatGPT, Gemini and DeepSeek for non-critical triage support using real-world conversations in emergency department. *BMC Emerg Med*. Sep 1, 2025;25(1):176. [doi: [10.1186/s12873-025-01337-2](https://doi.org/10.1186/s12873-025-01337-2)] [Medline: [40890624](https://pubmed.ncbi.nlm.nih.gov/40890624/)]

51. Tao Y, Li X, Yisha Z, Yang S, Zhan S, Sun F. LitAutoScreener: development and validation of an automated literature screening tool in evidence-based medicine driven by large language models. *Health Data Sci.* 2025;5:0322. [doi: [10.34133/hds.0322](https://doi.org/10.34133/hds.0322)] [Medline: [40904687](https://pubmed.ncbi.nlm.nih.gov/40904687/)]
52. Ruan M, Fan J, Liu M, Meng Z, Zhang X, Zhang C. Artificial intelligence for the science of evidence synthesis: how good are AI-powered tools for automatic literature screening? *BMC Med Res Methodol.* Aug 25, 2025;25(1):199. [doi: [10.1186/s12874-025-02644-9](https://doi.org/10.1186/s12874-025-02644-9)] [Medline: [40855531](https://pubmed.ncbi.nlm.nih.gov/40855531/)]
53. Cai X, Geng Y, Du Y, et al. Utilizing large language models to select literature for meta-analysis shows workload reduction while maintaining a similar recall level as manual curation. *BMC Med Res Methodol.* Apr 28, 2025;25(1):116. [doi: [10.1186/s12874-025-02569-3](https://doi.org/10.1186/s12874-025-02569-3)] [Medline: [40295957](https://pubmed.ncbi.nlm.nih.gov/40295957/)]
54. Young RJ, Matthews AM, Poston B. Benchmarking multiple large language models for automated clinical trial data extraction in aging research. *Algorithms.* 2025;18(5):296. [doi: [10.3390/a18050296](https://doi.org/10.3390/a18050296)]
55. Grillo R, Llanos AH, Costa C, Melhem-Elias F. Comparison of large language models in oral and maxillofacial surgery. *Br J Oral Maxillofac Surg.* Jan 2026;64(1):43-49. [doi: [10.1016/j.bjoms.2025.08.015](https://doi.org/10.1016/j.bjoms.2025.08.015)] [Medline: [41076417](https://pubmed.ncbi.nlm.nih.gov/41076417/)]
56. Scholz V, Bichtemann S, Bott OJ, Illig T, Haag S. AI for extracting pre-analytical variability data from biomedical literature: feasibility and validation. *Stud Health Technol Inform.* Sep 3, 2025;331:52-62. [doi: [10.3233/SHTI251379](https://doi.org/10.3233/SHTI251379)] [Medline: [40899527](https://pubmed.ncbi.nlm.nih.gov/40899527/)]
57. Wu X, Cai G, Guo B, et al. A multi-dimensional performance evaluation of large language models in dental implantology: comparison of ChatGPT, DeepSeek, Grok, Gemini and Qwen across diverse clinical scenarios. *BMC Oral Health.* Jul 28, 2025;25(1):1272. [doi: [10.1186/s12903-025-06619-6](https://doi.org/10.1186/s12903-025-06619-6)] [Medline: [40721763](https://pubmed.ncbi.nlm.nih.gov/40721763/)]
58. Li Y, Dong J, Liu D, et al. Systematic benchmarking of large language models in programmed cell death-oriented gastric cancer research: a comparative analysis of DeepSeek-V3, DeepSeek-R1, and Claude 3.5. *Discov Onc.* Jul 1, 2025;16(1):1227. [doi: [10.1007/s12672-025-02911-7](https://doi.org/10.1007/s12672-025-02911-7)]
59. Kayaalp ME, Gültekin O, Akçaalan S, Kahraman HÇ, Topçu HN, Kavrul Kayaalp G. Artificial intelligence in medical and biological research: promise and perils of ChatGPT and DeepSeek in advancing healthcare. *Turk J Biol.* 2025;49(5):585-599. [doi: [10.55730/1300-0152.2765](https://doi.org/10.55730/1300-0152.2765)] [Medline: [41246235](https://pubmed.ncbi.nlm.nih.gov/41246235/)]
60. Abuabara A, do Nascimento T, Trentini SM, et al. Evaluating the accuracy of generative artificial intelligence models in dental age estimation based on the Demirjian's method. *Front Dent Med.* 2025;6:1634006. [doi: [10.3389/fdmed.2025.1634006](https://doi.org/10.3389/fdmed.2025.1634006)] [Medline: [40800006](https://pubmed.ncbi.nlm.nih.gov/40800006/)]
61. AlShahwan N, Fetyani IM, Beyari MB, et al. Comparative performance analysis of AI engines in answering American Board of Surgery in-training examination questions: a multi-subspecialty evaluation. *Surg Innov.* Dec 2025;32(6):502-506. [doi: [10.1177/15533506251361664](https://doi.org/10.1177/15533506251361664)] [Medline: [40664612](https://pubmed.ncbi.nlm.nih.gov/40664612/)]
62. Yang Y, Yang F, Xiao S, et al. Application of large language models in TN staging and treatment response evaluation for patients with nasopharyngeal carcinoma: a comparative performance analysis of ChatGPT-4o-Latest and DeepSeek-V3-0324. *J Magn Reson Imaging.* Dec 2025;62(6):1793-1801. [doi: [10.1002/jmri.70140](https://doi.org/10.1002/jmri.70140)] [Medline: [41045017](https://pubmed.ncbi.nlm.nih.gov/41045017/)]
63. Yan MY, Qin JA, Yan D. Performance evaluation and application value of large language models in the prediction of drug-drug interactions. *Yaoxue Xuebao.* 2025;60(7):2122-2131. [doi: [10.16438/j.0513-4870.2025-0590](https://doi.org/10.16438/j.0513-4870.2025-0590)]
64. Xie L, Jin Y, Xu L, Chang S, Xu X. Fusing domain knowledge with a fine-tuned large language model for enhanced molecular property prediction. *J Chem Theory Comput.* Jul 22, 2025;21(14):6743-6758. [doi: [10.1021/acs.jctc.5c00605](https://doi.org/10.1021/acs.jctc.5c00605)] [Medline: [40631446](https://pubmed.ncbi.nlm.nih.gov/40631446/)]
65. McCoy TH, Perlis RH. Reasoning language models for more transparent prediction of suicide risk. *BMJ Ment Health.* May 11, 2025;28(1):e301654. [doi: [10.1136/bmjment-2025-301654](https://doi.org/10.1136/bmjment-2025-301654)] [Medline: [40350181](https://pubmed.ncbi.nlm.nih.gov/40350181/)]
66. Alessandri-Bonetti M, Liu HY, Giorgino R, Nguyen VT, Egro FM. The first months of life of ChatGPT and its impact in healthcare: a bibliometric analysis of the current literature. *Ann Biomed Eng.* May 2024;52(5):1107-1110. [doi: [10.1007/s10439-023-03325-8](https://doi.org/10.1007/s10439-023-03325-8)] [Medline: [37482572](https://pubmed.ncbi.nlm.nih.gov/37482572/)]
67. Chen SF, Alyakin A, Seas A, et al. LLM-assisted systematic review of large language models in clinical medicine. *Nat Med.* Mar 2026;32(3):1152-1159. [doi: [10.1038/s41591-026-04229-5](https://doi.org/10.1038/s41591-026-04229-5)] [Medline: [41776077](https://pubmed.ncbi.nlm.nih.gov/41776077/)]
68. Anusitviwat C, Suwannaphisit S, Bvonpanttarananon J, Tangtrakulwanich B. Comparing ChatGPT and DeepSeek for assessment of multiple-choice questions in orthopedic medical education: cross-sectional study. *JMIR Form Res.* Dec 19, 2025;9:e75607. [doi: [10.2196/75607](https://doi.org/10.2196/75607)] [Medline: [41418321](https://pubmed.ncbi.nlm.nih.gov/41418321/)]
69. 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/)]
70. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* Aug 2023;29(8):1930-1940. [doi: [10.1038/s41591-023-02448-8](https://doi.org/10.1038/s41591-023-02448-8)] [Medline: [37460753](https://pubmed.ncbi.nlm.nih.gov/37460753/)]
71. He K, Mao R, Lin Q, et al. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. *Information Fusion.* Jun 2025;118:102963. [doi: [10.1016/j.inffus.2025.102963](https://doi.org/10.1016/j.inffus.2025.102963)]

72. Liu F, Zhou H, Gu B, et al. Application of large language models in medicine. *Nat Rev Bioeng*. Jul 2025;3(6):445-464. [doi: [10.1038/s44222-025-00279-5](https://doi.org/10.1038/s44222-025-00279-5)]
73. Unger JP, Morales I, De Paepe P, Roland M. Integrating clinical and public health knowledge in support of joint medical practice. *BMC Health Serv Res*. Dec 9, 2020;20(Suppl 2):1073. [doi: [10.1186/s12913-020-05886-z](https://doi.org/10.1186/s12913-020-05886-z)] [Medline: [33292211](https://pubmed.ncbi.nlm.nih.gov/33292211/)]
74. Hassanein FEA, El Barbary A, Hussein RR, et al. Diagnostic performance of ChatGPT-4o and DeepSeek-3 differential diagnosis of complex oral lesions: a multimodal imaging and case difficulty analysis. *Oral Dis*. Dec 2025;31(12):3361-3371. [doi: [10.1111/odi.70007](https://doi.org/10.1111/odi.70007)] [Medline: [40589366](https://pubmed.ncbi.nlm.nih.gov/40589366/)]
75. Goyal A, Sulaiman SA, Alaarag A, et al. Comparison of ChatGPT and DeepSeek large language models in the diagnosis of pericarditis. *World J Cardiol*. Aug 26, 2025;17(8):110489. [doi: [10.4330/wjc.v17.i8.110489](https://doi.org/10.4330/wjc.v17.i8.110489)] [Medline: [40949931](https://pubmed.ncbi.nlm.nih.gov/40949931/)]
76. He Q, Tan Z, Niu W, et al. From algorithms to operating room: can large language models master China's attending anesthesiology exam? A cross-sectional evaluation. *Int J Surg*. Jan 1, 2026;112(1):190-201. [doi: [10.1097/JS9.0000000000003406](https://doi.org/10.1097/JS9.0000000000003406)] [Medline: [40905848](https://pubmed.ncbi.nlm.nih.gov/40905848/)]
77. Karataş G, Karataş ME. Artificial intelligence in pediatric ophthalmology: a comparative study of ChatGPT-4.0 and DeepSeek-R1 performance. *Strabismus*. Mar 2026;34(1):61-67. [doi: [10.1080/09273972.2025.2536782](https://doi.org/10.1080/09273972.2025.2536782)] [Medline: [40726359](https://pubmed.ncbi.nlm.nih.gov/40726359/)]
78. Smith A, Liebrez M, Bhugra D, Grana J, Schleifer R, Buadze A. Are clinical improvements in large language models a reality? Longitudinal comparisons of ChatGPT models and DeepSeek-R1 for psychiatric assessments and interventions. *Int J Soc Psychiatry*. Feb 2026;72(1):91-102. [doi: [10.1177/00207640251358071](https://doi.org/10.1177/00207640251358071)] [Medline: [40741928](https://pubmed.ncbi.nlm.nih.gov/40741928/)]
79. Harada Y, Kawamura R, Yokose M, Singh H, Shimizu T. Atypical presentations at risk for diagnostic errors in internal medicine: a scoping review. *J Gen Intern Med*. May 2026;41(7):1937-1956. [doi: [10.1007/s11606-025-09901-z](https://doi.org/10.1007/s11606-025-09901-z)] [Medline: [41085962](https://pubmed.ncbi.nlm.nih.gov/41085962/)]
80. Brohi S, Mastoi Q ul ain, Jhanjhi NZ, Pillai TR. A research landscape of agentic AI and large language models: applications, challenges and future directions. *Algorithms*. 2025;18(8):499. [doi: [10.3390/a18080499](https://doi.org/10.3390/a18080499)]
81. Temsah A, Alhasan K, Altamimi I, et al. DeepSeek in healthcare: revealing opportunities and steering challenges of a new open-source artificial intelligence frontier. *Cureus*. Feb 2025;17(2):e79221. [doi: [10.7759/cureus.79221](https://doi.org/10.7759/cureus.79221)] [Medline: [39974299](https://pubmed.ncbi.nlm.nih.gov/39974299/)]
82. Liu Q, Xin Y, Wu C, et al. Diagnostic value of combining ultrafast cine MRI and morphological measurements on gastroesophageal reflux disease. *Abdom Radiol*. 2025;50(10):4495-4506. [doi: [10.1007/s00261-025-04890-3](https://doi.org/10.1007/s00261-025-04890-3)]
83. Diniz-Freitas M, Diz-Dios P. DeepSeek: another step forward in the diagnosis of oral lesions. *J Dent Sci*. Jul 2025;20(3):1904-1907. [doi: [10.1016/j.jds.2025.02.023](https://doi.org/10.1016/j.jds.2025.02.023)] [Medline: [40654453](https://pubmed.ncbi.nlm.nih.gov/40654453/)]
84. ElSayed A, Updegrove GF. Limitations of broadly trained LLMs in interpreting orthopedic Walch glenoid classifications. *Front Artif Intell*. 2025;8:1644093. [doi: [10.3389/frai.2025.1644093](https://doi.org/10.3389/frai.2025.1644093)] [Medline: [40951327](https://pubmed.ncbi.nlm.nih.gov/40951327/)]
85. Beauchamp T, Childress J. *Principles of Biomedical Ethics*: marking its fortieth anniversary. *Am J Bioeth*. Nov 2019;19(11):9-12. [doi: [10.1080/15265161.2019.1665402](https://doi.org/10.1080/15265161.2019.1665402)] [Medline: [31647760](https://pubmed.ncbi.nlm.nih.gov/31647760/)]
86. Wang M, Shen Y, Zhao B, Zhou X, Sun L, Liu X. Enhancing LLM-based clinical reasoning in anesthesiology via graph-augmented retrieval and explainable generation. *Health Inf Sci Syst*. Dec 2025;13(1):62. [doi: [10.1007/s13755-025-00379-x](https://doi.org/10.1007/s13755-025-00379-x)] [Medline: [41041605](https://pubmed.ncbi.nlm.nih.gov/41041605/)]
87. Choudhury A, Shahsavari Y, Shamszade H. User intent to use DeepSeek for health care purposes and their trust in the large language model: multinational survey study. *JMIR Hum Factors*. May 26, 2025;12:e72867. [doi: [10.2196/72867](https://doi.org/10.2196/72867)] [Medline: [40418796](https://pubmed.ncbi.nlm.nih.gov/40418796/)]
88. Wang Z, Zhou H, Song T. A bibliometric analysis of large language model-based AI chatbots in surgery. *Annals of Medicine & Surgery*. 2025;87(7):4127-4138. [doi: [10.1097/MS9.0000000000003234](https://doi.org/10.1097/MS9.0000000000003234)]
89. Moëll B, Sand Aronsson F, Akbar S. Medical reasoning in LLMs: an in-depth analysis of DeepSeek R1. *Front Artif Intell*. 2025;8:1616145. [doi: [10.3389/frai.2025.1616145](https://doi.org/10.3389/frai.2025.1616145)] [Medline: [40607450](https://pubmed.ncbi.nlm.nih.gov/40607450/)]
90. Cao Y, Wang J, Li Y, Zhang Y, Zhong G, Song P. Expert consensus on the deployment of DeepSeek in medical institutions. *Chinese Medical Ethics*. 2025;38(5):674-678. [doi: [10.12026/j.issn.1001-8565.2025.05.19](https://doi.org/10.12026/j.issn.1001-8565.2025.05.19)]
91. Si Y, Meng Y, Chen X, et al. Quality safety and disparity of an AI chatbot in managing chronic diseases: simulated patient experiments. *NPJ Digit Med*. Sep 25, 2025;8(1):574. [doi: [10.1038/s41746-025-01956-w](https://doi.org/10.1038/s41746-025-01956-w)] [Medline: [40999038](https://pubmed.ncbi.nlm.nih.gov/40999038/)]
92. Dong C, Qiu X, Deng J, et al. Comparative evaluation of large language models in delivering guideline-compliant recommendations for topical NSAID use in musculoskeletal pain: a multidimensional analysis. *Clin Rheumatol*. Nov 2025;44(11):4703-4710. [doi: [10.1007/s10067-025-07640-4](https://doi.org/10.1007/s10067-025-07640-4)] [Medline: [40952435](https://pubmed.ncbi.nlm.nih.gov/40952435/)]
93. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? *NPJ Digit Med*. 2020;3:4. [doi: [10.1038/s41746-019-0206-x](https://doi.org/10.1038/s41746-019-0206-x)] [Medline: [31970289](https://pubmed.ncbi.nlm.nih.gov/31970289/)]

94. Watts E, Patel H, Kostov A, Kim J, Elkbuli A. The role of compassionate care in medicine: toward improving patients' quality of care and satisfaction. *J Surg Res*. Sep 2023;289:1-7. [doi: [10.1016/j.jss.2023.03.024](https://doi.org/10.1016/j.jss.2023.03.024)] [Medline: [37068438](https://pubmed.ncbi.nlm.nih.gov/37068438/)]
95. Thomas RL, Uminsky D. Reliance on metrics is a fundamental challenge for AI. *Patterns (N Y)*. May 13, 2022;3(5):100476. [doi: [10.1016/j.patter.2022.100476](https://doi.org/10.1016/j.patter.2022.100476)] [Medline: [35607624](https://pubmed.ncbi.nlm.nih.gov/35607624/)]
96. Su H, Sun Y, Li R, et al. Large language models in medical diagnostics: scoping review with bibliometric analysis. *J Med Internet Res*. Jun 9, 2025;27:e72062. [doi: [10.2196/72062](https://doi.org/10.2196/72062)] [Medline: [40489764](https://pubmed.ncbi.nlm.nih.gov/40489764/)]
97. Zhou H, Wang Z, Wang R, et al. DeepSeek versus GPT: evaluation of large language model chatbots' responses on orofacial clefts. *J Craniofac Surg*. Sep 1, 2025;36(6):2197-2201. [doi: [10.1097/SCS.0000000000011399](https://doi.org/10.1097/SCS.0000000000011399)] [Medline: [40245329](https://pubmed.ncbi.nlm.nih.gov/40245329/)]
98. Zhou J, Cheng Y, He S, Chen Y, Chen H. Large language models for transforming healthcare: a perspective on DeepSeek-R1. *MedComm – Future Medicine*. Jun 2025;4(2):e70021. URL: <https://onlinelibrary.wiley.com/toc/27696456/4/2> [doi: [10.1002/mef2.70021](https://doi.org/10.1002/mef2.70021)]
99. Wang Q, Chen Z, Zhang H, et al. Large language models could be applied in personalized out-of-hospital management for breast cancer: a prospective randomized single blind study. *Sci Rep*. Sep 29, 2025;15(1):33589. [doi: [10.1038/s41598-025-18759-4](https://doi.org/10.1038/s41598-025-18759-4)]
100. Sahni NR, Carrus B. Artificial intelligence in U.S. health care delivery. *N Engl J Med*. Oct 12, 2023;389(15):1442-1443. [doi: [10.1056/NEJMc2310288](https://doi.org/10.1056/NEJMc2310288)]
101. Finkenbergl J. NASS 2023 presidential address: artificial intelligence and its effect on the art of medicine and the physician- patient relationship. *Spine J*. Feb 2024;24(2):191-194. [doi: [10.1016/j.spinee.2023.11.001](https://doi.org/10.1016/j.spinee.2023.11.001)] [Medline: [37944759](https://pubmed.ncbi.nlm.nih.gov/37944759/)]
102. Hui L, Khosa F. Artificial intelligence in action: racial and gender disparities in academic radiology. *Cureus*. Sep 2025;17(9):e92382. [doi: [10.7759/cureus.92382](https://doi.org/10.7759/cureus.92382)] [Medline: [41103889](https://pubmed.ncbi.nlm.nih.gov/41103889/)]
103. Dai X, Xi Y, Hu Y, et al. LLM evaluation for thyroid nodule assessment: comparing ACR-TIRADS, C-TIRADS, and clinician-AI trust gap. *Front Endocrinol*. 2025;16:1667809. [doi: [10.3389/fendo.2025.1667809](https://doi.org/10.3389/fendo.2025.1667809)]
104. Wang G, Meng X, Zhang F. Past, present, and future of global research on artificial intelligence applications in dermatology: a bibliometric analysis. *Medicine (Baltimore)*. 2023;102(45):e35993. [doi: [10.1097/MD.0000000000035993](https://doi.org/10.1097/MD.0000000000035993)]
105. Huang Y, Yang G, Shen Y, et al. Application of large language models in complex clinical cases: cross-sectional evaluation study. *JMIR Med Inform*. Aug 14, 2025;13:e73941. [doi: [10.2196/73941](https://doi.org/10.2196/73941)] [Medline: [41055081](https://pubmed.ncbi.nlm.nih.gov/41055081/)]
106. Sallam M, Alasfoor IM, Khalid SW, et al. Chinese generative AI models (DeepSeek and Qwen) rival ChatGPT-4 in ophthalmology queries with excellent performance in Arabic and English. *Narra J*. Apr 2025;5(1):e2371. [doi: [10.52225/narra.v5i1.2371](https://doi.org/10.52225/narra.v5i1.2371)] [Medline: [40352182](https://pubmed.ncbi.nlm.nih.gov/40352182/)]
107. Lim ECN, Cheng NCL, Lim CED. The art of medical synthesis: where Chinese medical wisdom intersects with artificial intelligence. *Journal of Traditional Chinese Medical Sciences*. Jan 2026;13(1):51-59. [doi: [10.1016/j.jtcms.2025.08.001](https://doi.org/10.1016/j.jtcms.2025.08.001)]
108. Patil NG, Kou NL, Baptista-Hon DT, Monteiro O. Artificial intelligence in medical education: a practical guide for educators. *MedComm – Future Medicine*. Jun 2025;4(2):e70018. URL: <https://onlinelibrary.wiley.com/toc/27696456/4/2> [doi: [10.1002/mef2.70018](https://doi.org/10.1002/mef2.70018)]
109. Yan W, Liu J, Liang W. DeepSeek empowers general medicine: potential application and prospect. *Chinese General Practice*. Jun 2025;28(17):2065-2069. [doi: [10.12114/j.issn.1007-9572.2025.0023](https://doi.org/10.12114/j.issn.1007-9572.2025.0023)]
110. Editage. URL: <https://www.editage.com/> [Accessed 2026-06-07]

Abbreviations

AI: artificial intelligence

APY: average publication year

BIBLIO: bibliometric reviews of biomedical literature

CAD-RADS: Coronary Artery Disease Reporting and Data System

GDPR: General Data Protection Regulation

GRPO: Group Relative Policy Optimization

HIPAA: Health Insurance Portability and Accountability Act

LI-RADS: Liver Imaging Reporting and Data System

LLM: large language model

MeSH: medical subject headings

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

RCT: randomized controlled trial

WoSCC: Web of Science Core Collection

Edited by Andrew Coristine; peer-reviewed by Fenglin Liu, Sully Chen, Zabir Al Nazi; submitted 11.Feb.2026; final revised version received 19.May.2026; accepted 20.May.2026; published 15.Jun.2026

Please cite as:

Zhang H, Wang D, Xu Y, Han S, Wang G

Applications of DeepSeek in Medicine: Bibliometric Analysis and Scoping Review

J Med Internet Res 2026;28:e93354

URL: <https://www.jmir.org/2026/1/e93354>

doi: [10.2196/93354](https://doi.org/10.2196/93354)

© Haoran Zhang, Dawei Wang, Yanliang Xu, Shuming Han, Guangxin Wang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 15.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.