Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Generative Artificial Intelligence (GenAI) Adoption in Healthcare: A Systematic Scoping Review

Amal Altalhi[1*], Itamar E Shabtai[1], Tajinder Barring[1]
[1*]Center of Information Systems and Technology, Claremont Graduate University, 91711, Claremont, CA-USA
***Corresponding Author:** *Amal Altalhi, amal.altalhi@cgu.edu*

ARTICLE INFO
Received:22 Apr 2025
Revised:20 May 2025
Accepted:27 May 2025

ABSTRACT

Importance: Generative artificial intelligence (AI) has revolutionized healthcare domains by offering solutions to complex issues, better disease detection and treatment, and patient care experience and related activities. However, despite its immense promises, the technology exhibits some problems that compromise the exploitation of its full potential.

Objectives: To provide a comprehensive overview of how generative artificial intelligence technologies have been implemented in healthcare

Evidence Review: PubMed, Web of Science, and Google Scholar were the electronic databases searched for studies published between 2022 and 2024 discussing GenAI in healthcare. An array of MeSH terms and keywords was used to query the databases and retrieve studies for inclusion based on eligibility criteria.

Findings: Of the 2178 studies retrieved from the search process, 67 were considered for final inclusion. Clinically, the GenAI application cuts across radiology, cardiology, cancer screening, and management of comorbidities. Some notable benefits of GenAI healthcare adoption include healthcare decision support, personalized care delivery, enhanced disease detection and management through improved diagnosis and imaging, advanced medical research support, and support for administrative (documentation) tasks. However, the main pitfalls included model hallucination, data privacy and security, and information accuracy and accountability (bias).

Conclusions and Relevance: Our scoping review identifies various applications of GenAI, with a comprehensive overview of promises or benefits and possible issues compromising the efficacy of the models. This article contributes to Gen AI literature in healthcare, emphasizing the need to address the mentioned pitfalls to ease the full utilization of generative AI potentials in health delivery.

Keywords: Generative Artificial Intelligence (AI), GenAI, Generative pre-trained models (GPT), ChatGPT, Large Language Models, Hallucination, Healthcare.

Introduction

Information resource access is a crucial element of healthcare delivery, enabling enhanced clinical outcomes, timely clinical decisions, and efficient resource utilization. However, healthcare data continuously exhibit high variability and velocity due to the complex nature of this domain, alongside data's increasing volume and the perpetual system complexity[1]. An imminent solution has been to develop and evolve healthcare information technology, which, unfortunately, presents more

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

challenges. The healthcare industry is revolutionizing with increasing personalized medication demands, patient data privacy enhancement, the panoptic goal of enhancing diagnostic accuracy and clinical outcomes, and the need to bolster chronic disease management[2, 3].Moreover, medical errors, misdiagnoses, and surgical mishaps are burgeoning in this critical field, raising a major public health concern.

Notably, the healthcare industry is experiencing the fusion of advanced technologies that offer innovative solutions in service delivery. A significant example is generative artificial intelligence (GenAI), such as large language models (LLMs)[4]. This is one of the artificial intelligence (AI) technologies known for its capability to craft new content in varied forms, including images, texts, audio, video, computer code, and speech. GenAI's immense success has been attributed to its unprecedented adoption across various fields; however, its application in healthcare exhibits excitement and controversy. GenAI models, for example, the Generative Pre-trained Transformer (GPT) developed by OpenAI, with its popular model ChatGPT, have received attention as a powerful tool to reshape healthcare due to their natural language processing (NPL) uncanny abilities[5, 6]. Other models include Bard, Bing Chat, LLaMA, Midjourney, Stable Diffusion, and DALL-E. By leveraging the potential for clinical decision support, GPT helps healthcare professionals formulate suggestions for optimized decision-making[7, 8]. Proponents attest to the application of GenAI in radiologic decision-making[9], neurologic clinical decisions[10], cancer diagnosis[11], detection[12], cardiopulmonary resuscitation[13], and bariatric surgery[14], to mention a few. The LLM's natural language understanding also improves its efficacy in facilitating patient-care provider communication to boost patient engagement and other healthcare administrative tasks[15].

Nevertheless, GenAI has invaluably attracted intense opposition due to its limitations in the medical field. There have been concerns about ethics, mainly privacy violations and patient data confidentiality. The "black box" interpretability challenge is also explicit, given the criticality of understanding the decision processes and conclusions on each health issue[16]. Lacking interpretability would render them untrustworthy. Additionally, the possibility of producing unreliable, ungrounded, or misleading content, regarded as hallucination, can harm healthcare where accuracy and reliability are crucial[17]. Other concerns include generalization and the potential for biased outputs caused by AI algorithms.

While studies have presented valuable insights, it's inevitable that harnessing this transformative technology ethically, effectively, and adequately remains a puzzle, especially for healthcare adoption needs. In light of the above, this systematic scoping review provides a comprehensive overview of how generative artificial intelligence technologies have been implemented in healthcare. By exploring the vast promises alongside possible perils that cannot be ignored, this study seeks to contribute to the ongoing debate of harnessing GenAI's significant transformative capabilities to improve overall clinical practice and care delivery experience.

Methods

Review Protocol

This systematic scoping review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, ensuring reliability, transparency, and rigorous reporting. The protocol aimed to identify themes in recent literature on the promises and pitfalls of generative AI (GenAI) applications in healthcare, guiding future research in alignment with CISAT's focus on AI-driven solutions.

Information Sources and Search Strategy

We searched five electronic databases, PubMed, Web of Science, Scopus, IEEE Xplore, and Google Scholar, to capture interdisciplinary studies on GenAI in healthcare. These databases were selected for their comprehensive coverage of clinical, informatics, and technical literature, ensuring a robust

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

evidence base. The search used a combination of keywords and MeSH terms: ("generative artificial intelligence" OR "GenAI" OR "large language models" OR "LLMs" OR "Generative Pre-trained Transformer" OR "ChatGPT") AND ("healthcare" OR "health" OR "medical" OR "clinical")). Searches covered studies published from January 2020 to April 2024, reflecting the rapid evolution of GenAI technologies.

Eligibility (Inclusion & Exclusion) Criteria

Studies were included if peer-reviewed, published in English or select non-English languages (with translated abstracts), and focused on specific GenAI applications in healthcare (e.g., clinical, administrative, research). Original research (e.g., clinical studies, case reports, RCTs, validation studies) was prioritized, with three high-quality reviews or editorials included as supplementary sources for contextual insights. Articles were excluded if they were uncategorized, books, news, nonscientific, systematic reviews, conference proceedings, or lacked full-text availability. Studies discussing non-GenAI machine learning or out-of-scope topics were also excluded.

Study Selection and Screening

The selection process involved two stages. First, two reviewers independently screened titles and abstracts for relevance based on inclusion/exclusion criteria. Second, full-text articles were reviewed to confirm eligibility. Supplementary sources (reviews/editorials) followed a separate screening track to ensure clarity in analysis. Discrepancies were resolved through consensus or consultation with a third reviewer. The process yielded 67 studies, including 62 original research articles, 2 non-English studies (translated abstracts), and 3 supplementary sources.

Data Charting and Extraction

Data were extracted into a bespoke Microsoft Word form, piloted on 5% of included studies. Extracted fields included study characteristics (author, year), GenAI application, promises/benefits, and pitfalls/challenges. Reviewers independently charted data, ensuring consistency through regular calibration.

Data Analysis

Data were synthesized narratively to map GenAI applications, promises, and pitfalls. Thematic clustering, conducted using NVivo, grouped findings into meta-themes: Clinical Efficacy (e.g., disease detection, decision support), Ethical Challenges (e.g., hallucination, bias, privacy), and Operational Efficiency (e.g., administrative tasks, research support). Co-citation analysis, performed with VOSviewer, identified influential studies shaping the GenAI discourse, highlighting key authors and works. The sample size of 67 studies was sufficient to capture the diversity of GenAI applications and challenges in this emerging field, consistent with scoping review methodology that prioritizes conceptual mapping over exhaustive inclusion. These analytical methods enhanced the exploratory depth of the review, providing a structured synthesis of the literature.

Quality Assessment

A simplified Critical Appraisal Skills Programme (CASP) checklist was applied to assess the methodological quality of included studies, focusing on clarity of objectives, study design appropriateness, data collection rigor, and reporting transparency. Each study was rated as high, moderate, or low quality, with results summarized in Table 1: Quality Assessment Summary. This assessment balanced rigor with the scoping review's broad scope.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 1: Quality Assessment Summary

Quality Rating	Number of Studies	Description
High	40	Clear objectives, robust design, transparent reporting
Moderate	22	Minor limitations in design or reporting
Low	5	Significant methodological or reporting issues

Note: The majority of studies (59.7%) were rated high quality, supporting the robustness of the evidence base. Low-quality studies were included for completeness but interpreted cautiously.

Results

Our initial search identified and retrieved 2178 unique titles. After deduplicating 567 records, 1611 were subjected to screening. In the abstract and title screening, 918 articles were found to be unrelated to the specific research interest and excluded, and further 189 studies not focusing primarily on GenAI or for discussing non-health related GenAI use. Five hundred and four studies were sought for retrieval, but only 432 studies were retrieved for full-text screening. After full-text screening based on the predefined inclusion criteria, only 67 articles were included in the final analysis. The PRISMA diagram, Fig.1, depicts our search and selection process.

Characteristics of Included Studies

Most literature inclusions in this study were original research articles, with few being opinionated articles such as conferences.

Of the 34 articles reporting various applications of GenAI in the medical field, 35% reported Gen AI disease detection applications, 18% medical decision support, 35% healthcare administrative support, and 12% reported applications in medical research fields (Table 2).

More than half of the studies also reported decision-making support by providing enhanced, accurate, clinically valuable information to caregivers and patients, which is one of the successful promises delivered by GPT models (Table 3).

Hallucination was the most frequently discussed limitation of GenAI adoption in healthcare, with possible harm to most clinical decisions, as well as compromised medical research (Table 4).

Tables 2, Table 3, and Table 4 provide a detailed summary of the themes and key findings of each of the included studies.

GenAI models and Applications in Healthcare Domains

Table 2 highlights various implementation domains of GenAI in healthcare. Based on the results, disease detection is among the key domains of Gen AI applications, spanning across conditions including cancer screening, surgical procedures, mental health, and managing comorbidities such as diabetes by predicting CKD risks[9, 19, 28]. Scholars also reported the robustness of Gen AI in supporting medical decision-making, for instance, quality cancer information, ophthalmic diagnosis, and heart disease management[19, 29, 33]. With its wider applications in general administrative functions, findings show the technology's significance in health data documentation, for example, in preparing discharge summaries, operative notes, radiologic reports, and health reports[34, 44]. Nevertheless, findings also unearth the tech's application in medical research to solve complex medical and clinical questions, improve data collection, analysis, and research communication, and increase the production of top-quality medical papers[45, 48] (Table 2).

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

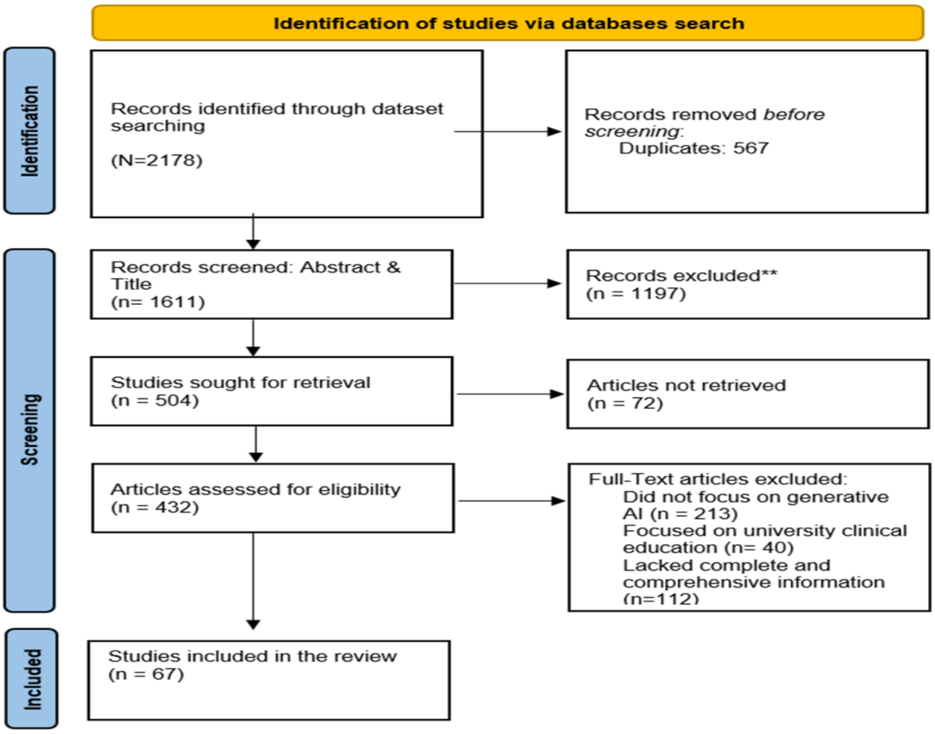

Figure 1: PRISMA Flow diagram

Table 2: GenAI Applications in Healthcare Summary

Service/Implementation domains	***Sub-Themes***	***Supporting studies***
Disease detection (diagnosis & imaging) and treatment	*Diagnosing ophthalmic conditions, obstructive sleep apnea*	*19–22*
	Breast cancer screening	*9,23*
	Providing accurate information for surgery	*24,25,49*
	Diabetes management, i.e., Predicting and identifying risk drivers for Chronic Kidney Disease (CKD) type 2 diabetes comorbidity	*26*
	Mental health/ Depression and anxiety management, i.e., Chatbot therapy for reducing depression and anxiety	*27,28*
Medical decision support	*Clinical decision support*	*29,30*
	Quality cancer information	*31*
	Radiology/ imaging; Ophthalmic diagnosis	*19*
	Cardiology/ heart disease	*32*
	Performing triage	*33*
Administrative support (Clinical documentation)	*Discharge summary preparation,*	*34*
	Health message generation, health-related texts, and reports	*35–37*
	Generating Operative notes	*34,38*
	Radiologic report labeling & image diagnosis	*39,40*
	Quality Patient interactions through medical responses and appropriate medical advice	*41–43*

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	Streamlining collection and analysis of patient data.	*44*
Medical research	*Solving complex medical and clinical questions.*	*45*
	Improving data collection and analysis, communication, and accessibility for medical researchers	*46*
	Producing and summarizing medical articles	*47,48*

Potential Promises of GenAI models in healthcare

Decision-making support

With a formidable evidence of application in various healthcare activities, one of the applauded benefit or promises of this technology is decision-making support. According to most studies, GenAI has supported decision-making by providing enhanced, accurate, clinically valuable information to caregivers and patients in the healthcare field[26, 31, 32, 49-60]. Interestingly, these studies further assert that the GenAI models and chatbots could be invaluable physician assistance tools, helpful in contextualizing risk prediction explanations to inform clinical assessment decisions. Besides, the models are also reported to support clinical decision-making through their immense potential in patient categorization, providing accurate and valuable medical information and response, precise interpretation of symptoms, and aiding clinical reasoning (Table 3).

Personalized healthcare

Although not largely reported, one study has underscored that personalized healthcare as another strategic promise of GenAI. As a revolutionary tool to reckon with in the advent of advanced technology, it is reported that large language models (LLM), such as ChatGPT, depicted the possibility of improving customized care. For example, one study noted that the models effectively enhanced therapeutic alliance for reducing depression and anxiety[27] (Table 3).

Transforming medical/ disease detection & evidence-based treatment

The large language models have also shown great promise to transform medical or disease detection alongside evidence-based treatment. Included studies demonstrate that GenAI has transformed disease detection by improving disease diagnosis accuracy, for instance, in ophthalmic conditions, enhanced medical screening and imaging, which is evident in current cancer screening and treatment, or even providing solutions in complex clinical scenarios by providing accurate information, like in the case of oromaxillofacial surgery[9, 19-25] (Table 3).

Revolutionizing medical research

Generally, GenAI has also shown great promise in revolutionizing medical research. For instance, the inclusions have reported the potential of these LLMs in generating contextual, accurate answers to complex clinical questions as part of revolutionizing and advancing medical research. These models also support data collection, analysis, and communication of research findings. Studies also acknowledge the versatility of using these models to increase the production of top-quality medical scientific research papers (Table 3)[45, 46, 61].

Improved healthcare operations (Administrative)

Finally, GenAI models such as ChatGPT have significantly impacted healthcare administrative operations. The included articles asserted the capacity of these models to generate quality and clear health messages as crucial administrative support potential. Further, the models aid patient interactions since they can provide automated medical responses to patient questions and equitable medical advice. Further administrative tasks include generating discharge summaries, patient information leaflets, operative notes for surgeons, and refined clinical documentation.

These models have also shown potential for use in medical dialogue summarization tasks or even providing intraoperative support, including streamlined patient data collection and analysis to enhance physician-patient communication for planned care[34, 35, 38, 41-44, 62-65]. Again, these tools also act as research tools to increase the production of top-quality medical decision-making[76] (Table 3).

Table 3: Promises or Benefits of GenAI Application in Healthcare Delivery

Key Finding	***Description***	***Supporting Studies***
Decision-making support by providing enhanced, accurate, clinically valuable information to caregivers and patients	*Informing clinical assessment decisions by contextualizing risk prediction explanations/ physician assistance*	*26,50*
	Recategorizing refractive surgery patients	*49,51*
	Providing accurate and valuable medical information & responses, including in medical licensing examinations and lab tests	*43,52–55*
	Accurate interpretation of symptoms of cardiac conditions for clinical decisions	*56,57*
	Debunking sleep-related myths	*32*
	Supporting clinical reasoning for informed decision-making	*58–60*
Personalized healthcare	*Enhanced therapeutic alliance for reducing depression and anxiety*	*27*
Improved healthcare operations (administrative)	*Quality and clear health messages*	*35*
	Patient interactions: responding to patient questions (medical responses) and providing appropriate and equitable medical advice	*41–43*
	Constructing discharge documents/summaries and operative notes	*34,38*
	Generating patient information leaflets	*65*
	Medical dialogue summarization	*64*
	Refined clinical documentation	*62,63*
	Providing intraoperative support, streamlining collection and analysis of patient data, and facilitating communication between physicians, patients and relatives for care planning.	*44*
Transforming medical/ disease detection & evidence-based treatment	*Enhanced disease diagnosis accuracy, i.e., diagnosing ophthalmic conditions, obstructive sleep apnea*	*19–22*
	Transforming medical screening and imaging, i.e., breast cancer screening	*9,23*
	Solving complex clinical scenarios, i.e., providing accurate information for oromaxillofacial surgery	*24,25*
Revolutionizing medical research	*Generating accurate answers to complex medical and clinical questions.*	*45*
	Supporting data collection and analysis and communication	*46*

Potential Pitfalls of GenAI models in healthcare

One of the main concerns about GenAI application in healthcare is the challenge of interpretability and transparency. Precision and trustworthiness of AI-generated content are fundamental in healthcare due to possible implications and harm if handled carelessly. Studies noted that the models were not optimally producing verifiable quality content in some instances to guide informed clinical

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

decisions. Audibility issues were also inevitable for some generated content, making it impossible to interpret and support healthcare delivery processes[28, 47, 63, 66-69] (Table 4).

Data privacy and security were also a great risk due to the nature and sensitivity of patient data and clinical information. Studies emphasized concerns over the trust and accuracy of these models' training data and the models' capability to accord privacy and security to patient information[70-74] (Table 4).

Another notable challenge is the propensity of these LLMs to generate and propagate bias or misinformation. Irrespective of the models' remarkable performance and capabilities, studies expressed reservations about training data issues that lead to bias and technical limitations[20, 28, 46, 67, 73, 75-77] (Table 4).

Various studies also raised concerns regarding the scalability and integration of most GenAI models and chatbots. Studies noted limited or lack of technical and non-technical clarity in responding to complex ethical vignettes. Studies also exhibited the inability of the models to substitute physicians, such as orthodontists' essential critical thinking and comprehensive subject knowledge, potentially due to the model's limited knowledge base and training data[19, 22, 54, 75, 78] (Table 4).

Furthermore, there is an overemphasis on GenAI-based hallucination, a crucial critique of GenAI's successful application and integration in various healthcare-related tasks and procedures[30, 47, 75, 97-83]. Quite a number of included studies pointed out the model's potential for perpetuating or fabricating information or significant responses to critical clinical inquiries. For example, studies noted the chatbot's weakness in providing confident responses that are incorrect or fabricated, creating and disseminating harmful and inaccurate health-related information, and inconsistent, occasional clinically inappropriate replies. In the context of medical research, studies faulted the models for generating fraudulent yet authentic-looking scientific medical articles and limited the ability to create reliable references for medical research proposals and publications. Finally, studies also noted the weakness of the models by providing inconsistent and precise drug assessment responses that are clinically irrelevant (Table 4).

Table 4: Pitfalls and Challenges of GenAI Adoption in Healthcare

Key Finding	***Description***	***Supporting Studies***
Interpretability and transparency/quality of content	*Auditability issues*	28,63,69
	Error-prone process of human adjudication	47,67,68
	Decreasing accuracy and poor performance on more complex tasks	66
	Black box nature of GenAI models	47
Data privacy and security	*Concerns about trust and accuracy*	70,71
	Ethical issues related to the use of patient data	72–74
Information accuracy and accountability (bias)	*Training data issues leading to bias and technical limitations*	28,46,73
	Low accuracy in answering questions	20,77,84
	Propagating training data bias	76
	Low accuracy in diagnosis and treatment suggestions	77,84
Hallucination	*Chatbots provide confident responses that are incorrect or fabricated.*	30
	Creating and spreading misinformation/ harmful and inaccurate content	79,84
	Lack of consistent justification and occasional clinically inappropriate responses	80

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	Generating fraudulent but authentic-looking scientific medical articles	*47*
	Limited ability to generate reliable references for medical research proposals	*81,82*
	Inconsistently precise drug assessment responses that are not clinically meaningful	*83*
Scalability and integration (usability)	*Technical and non-technical clarity in responding to complex ethical vignettes*	*78*
	Models cannot serve as a substitute for the orthodontist's essential critical thinking and comprehensive subject knowledge	*20,22,75*
	Unaccountability for specific patient preferences; Dependency on pre-existing data and inefficient training datasets for the models	*20,22*
	Challenges with domain integration in a multilingual setting	*54*
	Difficulty in explaining nuanced ethical dilemmas.	*19*

Discussion

The adoption and integration of generative artificial intelligence (GenAI) in healthcare presents a promising future in various clinical applications, yet it is not short of some shortcomings that could hinder its maximum utilization[6, 85]. GenAI is poised to revolutionize healthcare in the near future, and the mentioned application domains exhibited in this scoping review of 67 studies provide a glimpse of its potential within the realms of healthcare[6, 85]. While ChatGPT has undoubtedly been perceived as the dominant GenAI technology since 2022, it is noteworthy that more examples of the generative pre-trained (GPT) models and architecture are proving useful in various capacities[6] . Grounded in socio-technical systems theory[86] and technology acceptance models (TAM[87]; UTAUT[88]), this scoping delves into the multifaceted landscape of GenAI applications in the medical and healthcare domain, offering a comprehensive summary of the latest research and evidence of potential promises and benefits, as well as limitations or challenges of the technology's integration in healthcare domains.

The review summarized diverse healthcare applications of GenAI. These large language models (LLMs) technologies exhibit multiple roles within the medical domain. For instance, they are widely applied in disease detection and treatment and have proved effective[6] . These LLMs can extract medical insights from physician records, diagnostic reports, and patient histories to facilitate a precise diagnosis and treatment[6] . Studies have also highlighted the critical application of LLM models such as chatbots and ChatGPT in clinical decision support, medical-related research, and administrative tasks (clinical documentation)[6] . Based on the findings, one insightful GenAI strength is revolutionizing disease detection and evidence-based treatment. GenAI has been used in disease diagnosis and imaging to enhance the accuracy and efficiency of diagnostic results and interpretation[6] . Moreover, the technology guarantees high quality and improves the diagnosis of complex medical conditions[89] . Leibig *et al.* extrapolated that artificial intelligence algorithms outperformed human radiologists in disease detection, particularly during mammograms for breast cancer screening[89] . Regarding individual care, GenAI technologies have ensured and improved personalized healthcare delivery by tailoring treatment options to each patient's needs and conditions[90] . Studies postulate that these AI technologies analyze health profile information, including patient lifestyle factors, disease history, genomics, and current health information, to develop a customized treatment plan[91] . This approach ensures that treatment choices enhance patients' quality of life more effectively[91] .

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Additionally, this study elucidates AI's promise of supporting decision-making. By harnessing GenAI's potential, healthcare providers can make transformative, evidence-based clinical decisions for accurate, personalized treatments, diagnoses, and solutions to healthcare-related dilemmas[6] . Through their findings, Hayward *et al.* corroborate the aforementioned by emphasizing the role of GenAI in boosting patient engagement through personalized health information[92] . This study affirmed that ChatGPT yielded efficient treatment recommendations tailored to the patient's clinical profile, illness severity, and diagnoses, according to appropriate customized care plans and treatment decisions[92] . Further, this review also attests to the revolutionizing of medical-related research. According to Ghebrehiwet *et al.*, GenAI models have grown to be powerful tools in medical research, able to generate synthetic information mirroring real patient data[93] . This enhances its versatility in medical research in instances of scarce data or where data is difficult to obtain, like in the case of rare medical conditions[93] . Also, synthetic data can simulate clinical trials, disease progression models, and test hypotheses without necessarily contravening patient data and privacy [90]. Nonetheless, improved healthcare administrative functions cannot go unnoticed. Beyond notable improvements in patient care, GenAI has also optimized administrative operations from documentation resource allocation to patient communication[94] . For instance, ChatGPT has been efficient in helping hospitals optimize their staffing and resource allocations and predict patient admissions to bolster administrative efficiency[94] .

Comparative Analysis of GenAI Performance

GenAI's performance in healthcare applications can be contrasted with traditional AI (e.g., rule-based systems, supervised machine learning) and non-AI methods (e.g., manual processes) to clarify its unique contributions and limitations. In diagnostics, GenAI's adaptability, driven by LLMs, enables contextual understanding of complex medical texts, outperforming traditional AI's rigid rule-based or feature-engineered models in tasks like cancer screening[95] . For example, Leibig *et al.* demonstrate GenAI's superior accuracy in breast cancer detection compared to classical machine learning (e.g., SVM), which requires extensive feature selection[89] . However, GenAI's propensity for hallucination, generating fabricated outputs, introduces risks absent in more predictable traditional AI, necessitating robust validation[96] . Compared to non-AI methods, GenAI significantly enhances efficiency and reduces error rates. Manual radiology interpretation, for instance, is time-intensive and prone to human error, whereas GenAI automates image analysis with high sensitivity, as seen in ophthalmic diagnostics[97] . Similarly, administrative tasks like discharge summary preparation are streamlined by GenAI's natural language generation, reducing clinician workload compared to paper-based documentation. However, non-AI methods avoid GenAI's privacy risks, as manual processes do not require large-scale data training[98, 99] . In clinical decision support, GenAI's ability to process unstructured data surpasses traditional AI's reliance on structured inputs, enabling more nuanced recommendations[100, 103] . Yet, traditional AI's deterministic outputs foster greater clinician trust than GenAI's probabilistic, sometimes erroneous suggestions[101, 102, 103]. Non-AI decision-making, relying on clinician expertise alone, ensures accountability but is slower and less scalable, underscoring GenAI's efficiency gains despite its risks[104] . These comparisons highlight GenAI's transformative edge, adaptability and efficiency, while emphasizing the need to address hallucination and bias to compete with traditional AI's reliability and non-AI methods' simplicity.

Theoretical Implications

GenAI's adoption reflects socio-technical dynamics, where technical reliability (e.g., diagnostic accuracy) interacts with human factors like clinician trust[86] . The Technology Acceptance Model (TAM) posits that GenAI's enhanced diagnostics increase perceived usefulness, but hallucination undermines perceived ease of use, reducing adoption[87] . The Unified Theory of Acceptance and Use of Technology (UTAUT) suggests that social influence, such as peer endorsement and training, can

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

mitigate barriers to trust and adoption[88] . Ethical challenges, such as bias from unrepresentative training data, align with algorithmic fairness theories[105] . Transparent training and documentation ensure equitable outcomes, particularly in diverse settings[106] . Trust in LLMs requires transparent error reporting, reinforcing CISAT's ethical AI priorities[97] .

Challenges and Pitfalls

Notwithstanding, this analysis also revealed that the model's interpretability, transparency, scalability, and integration are critical limitations[71] . For instance, previous studies noted that one of the vital challenges of GenAI models is ensuring that their outputs are interpretable by physicians, hindering decision-making processes relating to diagnosis and treatment planning insights generated by the models[71] . Notably, there is also an increasing demand for transparency in the use of AI models. This highly depends on the documentation of model training, training data sources, and the decision-making process for accountability and trust for use[71] . Shaikh and colleagues have also criticized GenAI models for their "black box" nature, making them challenging for physicians to understand[98] . Regarding information accuracy and accountability or bias, it's argued that large language models can perpetuate or amplify bias in training data, leading to unequal treatment outcomes or misleading information[71] . Similarly, data privacy and security are fundamental concerns inhibiting the seamless application of GenAI models in healthcare[99] . The need for larger datasets for training the LLM models cannot be underestimated, as it is the main reason for data privacy and security challenges[99] . Thapa and Camtepe opine that health data is sensitive; thus, handling it must follow set standards and regulations[99] . In matters of hallucination, findings revealed that GenAI models had been faulted for providing confident yet incorrect or fabricated responses or medical-related content that is misleading and harmful within the healthcare domain[101] . Studies also noted a lack of consistent justification and the generation of unreliable clinically inappropriate responses for making clinical decisions, including fabricated citations and references[101] . According to Gravel *et al.*, the prevalence of fabricated information sources and the generation of fake references in GenAI-generated medical content exacerbates concerns about the integrity and validity of training data sources[102] .

Crucial to our findings is the element of hallucination, one of the major setbacks of GenAI adoption in healthcare[101] . As a consequence of continued system perpetuation and the generation of nonsensical or incorrect information, hallucination negatively impacts the reliability of AI technology[101] . In essence, when GenAI systems such as chatbots hallucinate, it equally undermines the dependability or accuracy of its generated outputs, hence, the perception of the system's unreliability[101] . Consequently, when the AI system is perceived as unreliable, the trust and willingness of users to adopt the system are corrupted[103] . For instance, there is likely to be diminished trust when users experience hallucinations, which arguably reduces confidence in its information outputs as the users begin to question the overall system's reliability[103] . Sun and Medaglia reported, with reference to a lack of trust in AI features, that individuals are unlikely to trust AI's diagnostic ability and predictive power for treatment purposes[103] . Unreliable systems are also thought to be risky, particularly in critical applications such as healthcare; hence, a lack of trust in them equates to fear of being used to avoid adverse outcomes or significant medical errors[107] . With regards to user willingness and intention to switch to AI technology, hallucination acts as a significant barrier since users are likely to become hesitant to adopt AI systems perceived to be unreliable and untrustworthy with respect to system inadequacy, limitation, and inoperability in healthcare[108] . The other crucial factor rests on long-term engagement. Due to reliability issues, hallucination compromises user experience due to high regard for dependable technology systems that give accurate and consistent outputs or results[109] . In this accord, hallucination is likely to compromise long-term user engagement since it makes the system non-dependable and inefficient[109] . Literature argues that reliable AI technologies foster user loyalty and long-term engagement, which determines their intention to switch to whichever AI system[109].

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Policy and Practice Recommendations

To address these challenges and maximize GenAI's potential in healthcare, stakeholders should adopt targeted strategies:

• Researchers: Develop interpretable LLMs by fine-tuning with healthcare-specific data and implementing human-in-the-loop validation to reduce hallucination and bias, ensuring reliable outputs[98] . Expand research to explore GenAI's applications in medical education and non-Western healthcare systems for broader impact.

• Policymakers: Enforce robust data privacy regulations, such as HIPAA-compliant frameworks, and mandate transparent documentation of model training and decision processes to foster trust and equity[100] .

• Practitioners: Integrate GenAI with clinician oversight in hybrid systems to ensure accurate outputs, supported by training and peer endorsement to enhance adoption and trust[88] .

Study Limitations

One key limitation of our study was the consideration of only three databases, PubMed, Web of Science, and Google Scholar, for literature searches. Amongst other limitations, for instance, including studies that were only original research articles, particularly those published in the English language, could potentially limit other articles that could offer more insights into the investigation of this scoping review. Finally, considering the scope of this investigation, consideration was only focused on clinical or medical perspectives of GenAI applications, implying that literature resources that investigated its application in medical education setup were eliminated.

Future Research Directions

This scoping review highlights GenAI's transformative potential and challenges in healthcare, yet several gaps warrant further exploration to maximize its ethical and effective adoption [6, 85]. Future research should address underexplored applications, contextual diversity, and technological advancements to advance the field and align with CISAT's focus on trustworthy AI solutions.

First, GenAI's application in non-Western and resource-constrained healthcare systems remains underexplored. Most included studies focused on high-resource settings (e.g., cancer screening in developed nations), limiting generalizability to diverse contexts. Future research should examine how GenAI can address global health challenges, such as infectious disease management in low-income regions or multilingual patient communication in multicultural settings. Key questions include: Can GenAI models be fine-tuned with region-specific data to improve diagnostic accuracy in under-resourced hospitals? How can privacy-preserving techniques ensure equitable adoption in data-scarce environments? Such studies would promote inclusive innovation, aligning with algorithmic fairness theories.

Second, the rapid evolution of GenAI technologies, including multimodal LLMs that integrate text, images, and audio, necessitates forward-looking research. While ChatGPT dominated this review's findings, emerging models (e.g., successors to DALL-E or LLaMA) offer potential for advanced applications, such as real-time surgical guidance or integrated diagnostic workflows. Research should explore: What are the performance trade-offs of multimodal GenAI compared to text-only models in clinical settings? How can human-in-the-loop validation mitigate hallucination in these advanced systems? These inquiries could position CISAT at the forefront of next-generation AI development.

Finally, developing novel frameworks to address GenAI's pitfalls, hallucination, bias, and interpretability, requires interdisciplinary efforts. Future studies could propose socio-technical models integrating TAM and UTAUT to predict clinician adoption in diverse contexts. Research questions include: How can explainable AI techniques enhance GenAI's transparency in high-stakes decisions? What training protocols best foster clinician trust in hybrid GenAI systems? Such frameworks would guide ethical integration, ensuring GenAI's reliability and equity in healthcare delivery.

By pursuing these directions, researchers can address the limitations of this review's scope (e.g., English-language bias, exclusion of medical education) and build on its findings to create trustworthy, globally relevant GenAI solutions. These efforts would reinforce CISAT's mission to advance AI-driven healthcare innovation while contributing to the broader discourse on responsible AI adoption.

Conclusion

This scoping review underscores GenAI's transformative potential in healthcare, from enhancing disease detection and decision support to streamlining administrative tasks, while highlighting critical challenges like hallucination, bias, and privacy. These findings offer a roadmap for ethical and effective GenAI adoption, aligning with CISAT's mission to advance AI-driven solutions. To realize GenAI's full potential, researchers must develop interpretable models to mitigate hallucination and bias, leveraging CISAT's expertise in natural language processing and data analytics. Policymakers should enforce robust data privacy regulations, such as HIPAA-compliant frameworks, to protect patient data. Practitioners are encouraged to adopt human-in-the-loop validation systems to ensure reliable GenAI outputs, fostering trust in clinical settings. Future studies should explore GenAI's applications in medical education and non-Western healthcare systems, ensuring equitable innovation globally. By addressing these priorities, CISAT and the broader research community can lead the development of trustworthy, impactful GenAI technologies for healthcare.

Declarations

Consent for publication

Not applicable

Availability of data and material

The authors confirm that the data supporting the findings of this study are available within the article

Competing interests

None

Funding

Nil

Authors' contributions

A. A. was involved in the conceptualization, investigation, methodology, data curation, and writing original draft. I. E. S. and T. B. were involved in supervision and reviewing of draft. All authors have read and approved the final manuscript.

Acknowledgements

Authors acknowledge Dr Ekom Edem for his immense assistance.

References

1. Suter-Crazzolara C. Better Patient Outcomes Through Mining of Biomedical Big Data. *Frontiers in ICT*. 2018;5. doi:10.3389/fict.2018.00030
2. Alotaibi YK, Federico F. The impact of health information technology on patient safety. *Saudi Med J*. 2017;38(12):1173-1180. doi:10.15537/smj.2017.12.20631
3. Matcha A. Innovations in Healthcare: Transforming Patient Care through Technology, Personalized Medicine, and Global Health Crises. *International Journal of Science and Research (IJSR)*. 2023;12(12):1668-1672. doi:10.21275/SR231222082955

4. Raiaan MAK, Mukta MSH, Fatema K, *et al*. A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access*. 2024;12. doi:10.1109/ACCESS.2024.3365742
5. De Angelis L, Baglivo F, Arzilli G, *et al*. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health*. 2023;11:1166120. doi:10.3389/fpubh.2023.1166120
6. Zhang P, Kamel Boulos MN. Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. *Future Internet*. 2023;15(9):286. doi:10.3390/fi15090286
7. Liu S, Wright AP, Patterson BL, *et al*. Assessing the Value of ChatGPT for Clinical Decision Support Optimization. *medRxiv*. Published online 2023. doi:10.1101/2023.02.21.23286254
8. Miao J, Thongprayoon C, Fülöp T, Cheungpasitporn W. Enhancing clinical decision-making: Optimizing ChatGPT's performance in hypertension care. *The Journal of Clinical Hypertension*. 2024;26(5):588-593. doi:10.1111/jch.14822
9. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD. Evaluating ChatGPT as an Adjunct for Radiologic Decision-Making. *medRxiv*. Published online February 7, 2023. doi:10.1101/2023.02.02.23285399
10. Pedersen M, Verspoor K, Jenkinson M, Law M, Abbott DF, Jackson GD. Artificial intelligence for clinical decision support in neurology. *Brain Commun*. 2020;2(2). doi:10.1093/braincomms/fcaa096
11. Zheng D, He X, Jing J. Overview of Artificial Intelligence in Breast Cancer Medical Imaging. *J Clin Med*. 2023;12(2). doi:10.3390/jcm12020419
12. Cirillo D, Núñez-Carpintero I, Valencia A. Artificial intelligence in cancer research: learning at different levels of data granularity. *Mol Oncol*. 2021;15(4). doi:10.1002/1878-0261.12920
13. Ahn C. Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation*. 2023;185. doi:10.1016/j.resuscitation.2023.109729
14. Samaan JS, Yeo YH, Rajeev N, *et al*. Assessing the Accuracy of Responses by the Language Model ChatGPT to Questions Regarding Bariatric Surgery. *Obes Surg*. 2023;33(6):1790-1796. doi:10.1007/s11695-023-06603-5
15. Nashwan AJ, AbuJaber AA. Harnessing the Power of Large Language Models (LLMs) for Electronic Health Records (EHRs) Optimization. *Cureus*. 2023;15(7):e42634. doi:10.7759/cureus.42634
16. Hassija V, Chamola V, Mahapatra A, *et al*. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognit Comput*. 2024;16(1):45-74. doi:10.1007/s12559-023-10179-8
17. Hatem R, Simmons B, Thornton JE. A Call to Address AI "Hallucinations" and How Healthcare Professionals Can Mitigate Their Risks. *Cureus*. Published online September 5, 2023. doi:10.7759/cureus.44720
18. Tricco AC, Lillie E, Zarin W, *et al*. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-473. doi:10.7326/M18-0850
19. Balas M, Ing EB. Conversational AI Models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel Pro Differential Diagnosis Generator. *JFO Open Ophthalmology*. 2023;1:100005. doi:10.1016/j.jfop.2023.100005
20. Chen J, Liu L, Ruan S, Li M, Yin C. Are Different Versions of ChatGPT's Ability Comparable to the Clinical Diagnosis Presented in Case Reports? A Descriptive Study. *J Multidiscip Healthc*. 2023;Volume 16:3825-3831. doi:10.2147/JMDH.S441790
21. Mira FA, Favier V, dos Santos Sobreira Nunes H, *et al*. Chat GPT for the management of obstructive sleep apnea: do we have a polar star? *European Archives of Oto-Rhino-Laryngology*. 2024;281(4):2087-2093. doi:10.1007/s00405-023-08270-9

2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

22. Pagano S, Holzapfel S, Kappenschneider T, *et al*. Arthrosis diagnosis and treatment recommendations in clinical practice: an exploratory investigation with the generative AI model GPT-4. *Journal of Orthopaedics and Traumatology*. 2023;24(1):61. doi:10.1186/s10195-023-00740-4
23. Yang F, Yan D, Wang Z. Large-Scale assessment of ChatGPT's performance in benign and malignant bone tumors imaging report diagnosis and its potential for clinical applications. *J Bone Oncol*. 2024;44:100525. doi:10.1016/j.jbo.2024.100525
24. Vaira LA, Lechien JR, Abbate V, *et al*. Accuracy of ChatGPT-Generated Information on Head and Neck and Oromaxillofacial Surgery: A Multicenter Collaborative Analysis. *Otolaryngology–Head and Neck Surgery*. 2024;170(6):1492-1503. doi:10.1002/ohn.489
25. Frosolini A, Franz L, Benedetti S, *et al*. Assessing the accuracy of ChatGPT references in head and neck and ENT disciplines. *European Archives of Oto-Rhino-Laryngology*. 2023;280(11):5129-5133. doi:10.1007/s00405-023-08205-4
26. Chari S, Acharya P, Gruen DM, *et al*. Informing clinical assessment by contextualizing post-hoc explanations of risk prediction models in type-2 diabetes. Published online February 11, 2023. doi:10.1016/j.artmed.2023.102498
27. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interv*. 2022;27:100495. doi:10.1016/j.invent.2022.100495
28. Chung NC, Dyer G, Brocki L. Challenges of Large Language Models for Mental Health Counseling. Published online November 23, 2023.
29. Nazary F, Deldjoo Y, Di Noia T. ChatGPT-HealthPrompt. Harnessing the Power of XAI in Prompt-Based Healthcare Decision Support using ChatGPT. In: *Communications in Computer and Information Science*. Vol 1947. ; 2024. doi:10.1007/978-3-031-50396-2_22
30. Ha LT, Kelley KD. Artificial Intelligence: Promise or Pitfalls? A Clinical Vignette of Real-Life ChatGPT Implementation in Perioperative Medicine. *J Gen Intern Med*. 2024;39(6):1063-1067. doi:10.1007/s11606-024-08611-2
31. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectr*. 2023;7(2). doi:10.1093/jncics/pkad015
32. Harskamp RE, De Clercq L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). *Acta Cardiol*. 2024;79(3):358-366. doi:10.1080/00015385.2024.2303528
33. Ayoub M, Ballout AA, Zayek RA, Ayoub NF. Mind + Machine: ChatGPT as a Basic Clinical Decisions Support Tool. *Cureus*. Published online August 18, 2023. doi:10.7759/cureus.43690
34. Singh S, Djalilian A, Ali MJ. ChatGPT and Ophthalmology: Exploring Its Potential with Discharge Summaries and Operative Notes. *Semin Ophthalmol*. 2023;38(5):503-507. doi:10.1080/08820538.2023.2209166
35. Lim S, Schmälzle R. Artificial intelligence for health message generation: an empirical study using a large language model (LLM) and prompt engineering. *Front Commun (Lausanne)*. 2023;8. doi:10.3389/fcomm.2023.1129082
36. Hanna JJ, Wakene AD, Lehmann CU, Medford RJ. Assessing Racial and Ethnic Bias in Text Generation for Healthcare-Related Tasks by ChatGPT1. *medRxiv*. Published online August 28, 2023. doi:10.1101/2023.08.28.23294730
37. Jeblick K, Schachtner B, Dexl J, *et al*. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *Eur Radiol*. 2023;34(5):2817-2825. doi:10.1007/s00330-023-10213-1

38. Rosenberg GS, Magnéli M, Barle N, *et al*. ChatGPT-4 generates orthopedic discharge documents faster than humans maintaining comparable quality: a pilot study of 6 cases. *Acta Orthop*. 2024;95:152-156. doi:10.2340/17453674.2024.40182
39. Mukherjee P, Hou B, Lanfredi RB, Summers RM. Feasibility of Using the Privacy-preserving Large Language Model Vicuna for Labeling Radiology Reports. *Radiology*. 2023;309(1). doi:10.1148/radiol.231147
40. Fink MA, Bischoff A, Fink CA, *et al*. Potential of ChatGPT and GPT-4 for Data Mining of Free-Text CT Reports on Lung Cancer. *Radiology*. 2023;308(3). doi:10.1148/radiol.231362
41. Nov O, Singh N, Mann D. Putting ChatGPT's Medical Advice to the (Turing) Test: Survey Study. *JMIR Med Educ*. 2023;9:e46939. doi:10.2196/46939
42. Nastasi AJ, Courtright KR, Halpern SD, Weissman GE. A vignette-based evaluation of ChatGPT's ability to provide appropriate and equitable medical advice across care contexts. *Sci Rep*. 2023;13(1):17885. doi:10.1038/s41598-023-45223-y
43. Johnson D, Goodman R, Patrinely J, *et al*. Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model. *Res Sq*. Published online 2023. doi:10.21203/rs.3.rs-2566942/v1
44. He Y, Tang H, Wang D, Gu S, Ni G, Wu H. Will ChatGPT/GPT-4 be a Lighthouse to Guide Spinal Surgeons? *Ann Biomed Eng*. 2023;51(7):1362-1365. doi:10.1007/s10439-023-03206-0
45. Sharma P, Thapa K, Thapa D, *et al*. Performance of ChatGPT on USMLE: Unlocking the Potential of Large Language Models for AI-Assisted Medical Education. Published online June 30, 2023.
46. Alsadhan A, Al-Anezi F, Almohanna A, *et al*. The opportunities and challenges of adopting ChatGPT in medical research. *Front Med (Lausanne)*. 2023;10. doi:10.3389/fmed.2023.1259640
47. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. *J Med Internet Res*. 2023;25:e46924. doi:10.2196/46924
48. Ilgaz HB, Çelik Z. The Significance of Artificial Intelligence Platforms in Anatomy Education: An Experience With ChatGPT and Google Bard. *Cureus*. Published online September 15, 2023. doi:10.7759/cureus.45301
49. Chiesa-Estomba CM, Lechien JR, Vaira LA, *et al*. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. *European Archives of Oto-Rhino-Laryngology*. 2024;281(4):2081-2086. doi:10.1007/s00405-023-08104-8
50. Lim DYZ, Tan Y Bin, Koh JTE, *et al*. ChatGPT on guidelines: Providing contextual knowledge to GPT allows it to provide advice on appropriate colonoscopy intervals. *J Gastroenterol Hepatol*. 2024;39(1):81-106. doi:10.1111/jgh.16375
51. Ćirković A, Katz T. Exploring the Potential of ChatGPT-4 in Predicting Refractive Surgery Categorizations: Comparative Study. *JMIR Form Res*. 2023;7:e51798. doi:10.2196/51798
52. Gwak G tae, Hwang U jae, Jung S hoon, Kim J hee. Search for Medical Information and Treatment Options for Musculoskeletal Disorders through an Artificial Intelligence Chatbot: Focusing on Shoulder Impingement Syndrome. *Journal of Musculoskeletal Science and Technology*. 2023;7(1). doi:10.29273/jmst.2023.7.1.8
53. Ebrahimian M, Behnam B, Ghayebi N, Sobhrakhshankhah E. ChatGPT in Iranian medical licensing examination: evaluating the diagnostic accuracy and decision-making capabilities of an AI-based model. *BMJ Health Care Inform*. 2023;30(1):e100815. doi:10.1136/bmjhci-2023-100815

54. Walker HL, Ghani S, Kuemmerli C, *et al*. Reliability of Medical Information Provided by ChatGPT: Assessment Against Clinical Guidelines and Patient Information Quality Instrument. *J Med Internet Res*. 2023;25:e47479. doi:10.2196/47479
55. Goodman RS, Patrinely JR, Stone CA, *et al*. Accuracy and Reliability of Chatbot Responses to Physician Questions. *JAMA Netw Open*. 2023;6(10):e2336483. doi:10.1001/jamanetworkopen.2023.36483
56. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: A Pilot Study. *Int J Environ Res Public Health*. 2023;20(4):3378. doi:10.3390/ijerph20043378
57. He Z, Bhasuran B, Jin Q, *et al*. Quality of Answers of Generative Large Language Models Versus Peer Users for Interpreting Laboratory Test Results for Lay Patients: Evaluation Study. *J Med Internet Res*. 2024;26:e56655. doi:10.2196/56655
58. Bragazzi NL, Garbarino S. Assessing the Accuracy of Generative Conversational Artificial Intelligence in Debunking Sleep Health Myths: Mixed Methods Comparative Study With Expert Analysis. *JMIR Form Res*. 2024;8:e55762. doi:10.2196/55762
59. Fonseca Â, Ferreira A, Ribeiro L, Moreira S, Duque C. Embracing the future, is artificial intelligence already better? A comparative study of artificial intelligence performance in diagnostic accuracy and decision-making. *Eur J Neurol*. 2024;31(4). doi:10.1111/ene.16195
60. Kaneda Y, Takita M, Hamaki T, Ozaki A, Tanimoto T. ChatGPT's Potential in Enhancing Physician Efficiency: A Japanese Case Study. *Cureus*. 2023;15(11):e48235. doi:10.7759/cureus.48235
61. Benichou L. The role of using ChatGPT AI in writing medical scientific articles. *J Stomatol Oral Maxillofac Surg*. 2023;124(5):101456. doi:10.1016/j.jormas.2023.101456
62. Blease C, Worthen A, Torous J. Psychiatrists' experiences and opinions of generative artificial intelligence in mental healthcare: An online mixed methods survey. *Psychiatry Res*. 2024;333:115724. doi:10.1016/j.psychres.2024.115724
63. Denecke K, May R, Rivera-Romero O. Transformer Models in Healthcare: A Survey and Thematic Analysis of Potentials, Shortcomings and Risks. *J Med Syst*. 2024;48(1):23. doi:10.1007/s10916-024-02043-5
64. Liu Y, Ju S, Wang J. Exploring the potential of ChatGPT in medical dialogue summarization: a study on consistency with human preferences. *BMC Med Inform Decis Mak*. 2024;24(1):75. doi:10.1186/s12911-024-02481-8
65. Lockie E, Choi J. Evaluation of a chat <scp>GPT</scp> generated patient information leaflet about laparoscopic cholecystectomy. *ANZ J Surg*. 2024;94(3):353-355. doi:10.1111/ans.18834
66. Agbavor F, Liang H. Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health*. 2022;1(12):e0000168. doi:10.1371/journal.pdig.0000168
67. Choi HS, Song JY, Shin KH, Chang JH, Jang BS. Developing prompts from large language model for extracting clinical information from pathology and ultrasound reports in breast cancer. *Radiat Oncol J*. 2023;41(3):209-216. doi:10.3857/roj.2023.00633
68. Kung TH, Cheatham M, Medenilla A, *et al*. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198
69. Wilhelm TI, Roos J, Kaczmarczyk R. Large Language Models for Therapy Recommendations Across 3 Clinical Specialties: Comparative Study. *J Med Internet Res*. 2023;25:e49324. doi:10.2196/49324
70. Raina A, Mishra P, goyal H, Kumar D. AI as a Medical Ally: Evaluating ChatGPT's Usage and Impact in Indian Healthcare. Published online January 28, 2024.
71. Chen Y, Esmaeilzadeh P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. *J Med Internet Res*. 2024;26:e53008. doi:10.2196/53008

72. Delsoz M, Madadi Y, Munir WM, *et al*. Performance of ChatGPT in Diagnosis of Corneal Eye Diseases. *medRxiv*. Published online August 28, 2023. doi:10.1101/2023.08.25.23294635
73. Bartal A, Jagodnik KM, Chan SJ, Dekel S. OpenAI's Narrative Embeddings Can Be Used for Detecting Post-Traumatic Stress Following Childbirth Via Birth Stories. *Res Sq*. Published online February 26, 2024. doi:10.21203/rs.3.rs-3428787/v2
74. Agarwal M, Goswami A, Sharma P. Evaluating ChatGPT-3.5 and Claude-2 in Answering and Explaining Conceptual Medical Physiology Multiple-Choice Questions. *Cureus*. Published online September 29, 2023. doi:10.7759/cureus.46222
75. Makrygiannakis MA, Giannakopoulos K, Kaklamanos EG. Evidence-based potential of generative artificial intelligence large language models in orthodontics: a comparative study of ChatGPT, Google Bard, and Microsoft Bing. *Eur J Orthod*. Published online April 13, 2024. doi:10.1093/ejo/cjae017
76. Spotnitz M, Idnay B, Gordon ER, *et al*. A Survey of Clinicians' Views of the Utility of Large Language Models. *Appl Clin Inform*. 2024;15(2):306-312. doi:10.1055/a-2281-7092
77. Oh N, Choi GS, Lee WY. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Ann Surg Treat Res*. 2023;104(5):269. doi:10.4174/astr.2023.104.5.269
78. Balas M, Wadden JJ, Hébert PC, *et al*. Exploring the potential utility of AI large language models for medical ethics: an expert panel evaluation of GPT-4. *J Med Ethics*. 2024;50(2):90-96. doi:10.1136/jme-2023-109549
79. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*. 2023;90:104512. doi:10.1016/j.ebiom.2023.104512
80. Zhou Y, Moon C, Szatkowski J, Moore D, Stevens J. Evaluating ChatGPT responses in the context of a 53-year-old male with a femoral neck fracture: a qualitative analysis. *European Journal of Orthopaedic Surgery & Traumatology*. Published online September 30, 2023. doi:10.1007/s00590-023-03742-4
81. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*. Published online April 11, 2023. doi:10.7759/cureus.37432
82. Sawamura S, Bito T, Ando T, Masuda K, Kameyama S, Ishida H. Evaluation of the accuracy of ChatGPT's responses to and references for clinical questions in physical therapy. *J Phys Ther Sci*. 2024;36(5):234-239. doi:10.1589/jpts.36.234
83. Montastruc F, Storck W, de Canecaude C, *et al*. Will artificial intelligence chatbots replace clinical pharmacologists? An exploratory study in clinical practice. *Eur J Clin Pharmacol*. 2023;79(10):1375-1384. doi:10.1007/s00228-023-03547-8
84. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatopoulos V, Kaklamanos EG. Evaluation of the Performance of Generative AI Large Language Models ChatGPT, Google Bard, and Microsoft Bing Chat in Supporting Evidence-Based Dentistry: Comparative Mixed Methods Study. *J Med Internet Res*. 2023;25:e51580. doi:10.2196/51580
85. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Science*. 2024;19(1):27. doi:10.1186/s13012-024-01357-9
86 Emery FE, Trist EL. Socio-technical systems. In: Churchman CW, Verhulst M, editors. Management sciences, models and techniques. Vol. 2. Oxford: Pergamon Press; 1960. p. 83-97.
87 Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 1989;13(3):319-40. doi:10.2307/249008

88 Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q. 2003;27(3):425-78. doi:10.2307/30036540

89. Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digit Health*. 2022;4(7). doi:10.1016/S2589-7500(22)00070-X

90. Yim D, Khuntia J, Parameswaran V, Meyers A. Preliminary Evidence of the Use of Generative AI in Health Care Clinical Services: Systematic Narrative Review. *JMIR Med Inform*. 2024;12:e52073. doi:10.2196/52073

91. Alowais SA, Alghamdi SS, Alsuhebany N, *et al*. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23(1):689. doi:10.1186/s12909-023-04698-z

92. Hayward R. Generative Artificial Intelligence-driven Healthcare Systems in Medical Imaging Analysis, in Clinical Decision Support, and in Patient Engagement and Monitoring. *Contemporary Readings in Law and Social Justice*. 2023;15(1). doi:10.22381/CRLSJ15120234

94. Preiksaitis C, Ashenburg N, Bunney G, *et al*. The Role of Large Language Models in Transforming Emergency Medicine: Scoping Review. *JMIR Med Inform*. 2024;12:e53787. doi:10.2196/53787

95 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. Lancet. 2019;393(10177):1347-54. doi:10.1016/S0140-6736(19)30086-4

96 Zhang Y, Li Y, Wang Z, Chen X, Liu J. Trust in large language models: a survey. Nat Mach Intell. 2023;5(10):1056-65. doi:10.1038/s42256-023-00703-x

97 Gao Y, Zhang H, Li J, Wang Q, Liu X. Clinical applications of large language models. J Med Syst. 2023;47(1):1-12. doi:10.1007/s10916-023-01916-2

98 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7

98. Shaikh TA, Mir WA, Sofi S. Decrypting the Black Boxing of Artificial Intelligence Using Explainable Artificial Intelligence in Smart Healthcare. In: *Studies in Computational Intelligence*. Vol 1021. ; 2022. doi:10.1007/978-3-030-97929-4_3

99. Thapa C, Camtepe S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Comput Biol Med*. 2021;129:104130. doi:10.1016/j.compbiomed.2020.104130

100. Ghebrehiwet I, Zaki N, Damseh R, Mohamad MS. Revolutionizing personalized medicine with generative AI: a systematic review. *Artif Intell Rev*. 2024;57(5):128. doi:10.1007/s10462-024-10768-5

101. Day T. A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT. *The Professional Geographer*. 2023;75(6):1024-1027. doi:10.1080/00330124.2023.2190373

102. Gravel J, D'Amours-Gravel M, Osmanlliu E. Learning to Fake It: Limited Responses and Fabricated References Provided by ChatGPT for Medical Questions. *Mayo Clinic Proceedings: Digital Health*. 2023;1(3):226-234. doi:10.1016/j.mcpdig.2023.05.004

103. Sun TQ, Medaglia R. Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Gov Inf Q*. 2019;36(2):368-383. doi:10.1016/j.giq.2018.09.008

105 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi:10.1126/science.aax2342

106 Chen IY, Joshi S, Ghassemi M. Ethical machine learning in healthcare. Lancet Digit Health. 2021;3(10):e627-8. doi:10.1016/S2589-7500(21)00187-0

107. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30-36. doi:10.1038/s41591-018-0307-0
108 . Prakash AV, Das S. Medical practitioner's adoption of intelligent clinical diagnostic decision support systems: A mixed-methods study. *Information & Management*. 2021;58(7):103524. doi:10.1016/j.im.2021.103524
109. Oelschlager R. Evaluating the Impact of Hallucinations on User Trust and Satisfaction in LLM-based Systems. Published online 2024.