# Preprints.org

Article

# ChatGPT: Transforming Healthcare with AI

Fnu Neha * , Deepshikha Bhati , Deepak Kumar Shukla , Md Amiruzzaman *

*Article*

# ChatGPT: Transforming Healthcare with AI

**Fnu Neha** [1,*] , **Deepshikha Bhati** [1] , **Deepak Kumar Shukla** [2] **and Md Amiruzzaman** [3]

1 Department of Computer Science, Kent State University, Kent, OH 44242; {neha@kent.edu; dbhati@kent.edu}
2 Rutgers Business School, Rutgers University, Newark, NJ 07102; ds1640@scarletmail.rutgers.edu
3 Department of Computer Science, West Chester University, West Chester, PA 19383; mamiruzzaman@wcupa.edu
* Correspondence: neha@kent.edu (F.N.)

**Abstract:** ChatGPT, developed by OpenAI, is an advanced large language model (LLM) that leverages artificial intelligence (AI) and deep learning to generate human-like responses. This paper explores its applications in digital health and healthcare, focusing on enhancing patient engagement through medical history collection, symptom assessment, and decision support to improve diagnostic accuracy. We also examine its role in promoting health literacy and clinical decision-making across various organ systems. Despite its advantages, the integration of ChatGPT in healthcare poses limitations and risks, including the potential for inaccurate information, overreliance by users, and ethical concerns related to informed consent. We discuss the challenges of keeping pace with rapidly evolving medical knowledge and highlight the need for continuous updates. Additionally, the paper addresses intellectual property and regulatory issues associated with AI deployment in medicine. Through this analysis, we provide insights into the benefits and potential risks of employing ChatGPT in healthcare.

**Keywords:** ChatGPT; large language models (LLM); artificial intelligence (AI); deep learning; digital health; healthcare; health literacy; medicine

---

## 1. Introduction

Artificial Intelligence (AI) has transformed various sectors, including finance, education, transportation, manufacturing, retail, agriculture, entertainment, telecommunications, and cybersecurity. Among these sectors, healthcare stands out as one of the most significantly impacted areas [1]. In healthcare, AI enhances diagnostic accuracy, streamlines administration, and improves patient care by analyzing large amounts of data with machine learning (ML), natural language processing (NLP), and deep learning (DL) [2]. AI applications range from early disease detection to automated patient management, significantly improving outcomes and reducing costs.

A key driver of AI's impact in healthcare is the emergence of Large Language Models (LLMs) which makes it essential in NLP tasks [3]. These models mimic human language processing using neural networks trained on extensive text datasets and excel in tasks such as machine translation, text generation, and summarization. The integration of LLMs into healthcare enables professionals to efficiently process vast amounts of medical literature, make informed decisions, and improve communication with patients, illustrating how these advancements build on AI's foundational contributions to the field.

One notable LLM is the Generative Pre-trained Transformer (GPT), particularly ChatGPT, which has shown impressive results in healthcare-specific evaluations, including medical exams and datasets like MedMCQA and PubMedQA [4]. This highlights the growing potential of conversational models in healthcare, where they can assist with patient communication and decision support.

ChatGPT stands out for its capacity to continuously learn and improve through interactions, delivering increasingly accurate and context-aware responses [5]. This adaptability makes it a valuable tool in healthcare, supporting tasks such as answering medical questions, resolving technical issues, and automating administrative functions. Additionally, its scalability enhances healthcare operations, streamlining patient triaging and information management, vital for improving overall healthcare delivery.

Despite its promise, challenges remain, particularly concerning real-world applicability and the ethical implications of AI-driven decision-making in medicine.

Based on the findings derived from existing publications on ChatGPT and its role in transforming healthcare with AI, this review contributes to the field through several key aspects:

1. **Comprehensive Background**: We provide a comprehensive overview of Natural Language Processing (NLP), Large Language Models (LLMs), the Generative Pre-trained Transformer (GPT) architecture, and ChatGPT, highlighting their evolution.
2. **Clinical Relevance**: We discuss the clinical implications of ChatGPT, emphasizing how the integration of ChatGPT in healthcare is transforming patient care, administrative tasks, and research.
3. **ChatGPT Applications Across Organ Systems**: We review literature highlighting ChatGPT's effectiveness in diagnostics, treatment recommendations, patient education, and clinician-patient communication across various organ systems.
4. **Risk Analysis**: We examine the potential risks of ChatGPT in healthcare, addressing reliability, accuracy, and ethical concerns, while evaluating existing methodologies.
5. **Future Directions**: We identify research gaps in ChatGPT applications in healthcare and propose a taxonomy to categorize the literature, enhancing understanding of its diverse applications.

The remainder of this paper is organized into several sections, each with multiple subsections. Section 2 describes the research methodology of this survey. Section 3 provides background on natural language processing (NLP), large language models (LLMs), GPT, and ChatGPT. Section 4 discusses ChatGPT's capabilities in empowering patients. Section 5 surveys existing literature on the use of ChatGPT across different organ systems. Section 6 addresses potential risks associated with the use of ChatGPT in healthcare. Section 7 presents a general discussion on ChatGPT in healthcare. Section 8 outlines the limitations of this study, while Section 9 explores future directions. Finally, Section 10 concludes the paper.

## 2. Research Methodology

To explore the applications of ChatGPT in healthcare, we employed a systematic review methodology, following established protocols to ensure a thorough and unbiased analysis. The review involved the following key steps:

1. **Literature Search**: We conducted a comprehensive search across multiple databases, including PubMed, IEEE Xplore, Scopus, Web of Science, Google Scholar, ACM Digital Library, arXiv, and ScienceDirect. The search utilized targeted keywords related to ChatGPT, natural language processing (NLP), healthcare applications, and ethical considerations, adhering to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [6].
2. **Inclusion and Exclusion Criteria**:

    • **Inclusion**: Articles discussing the application of ChatGPT in various healthcare contexts.
    • **Exclusion**: Non-English articles, publications without full-text access, and studies not specifically focused on ChatGPT or healthcare.

3. **Data Retrieval**: By August 20, 2024, we retrieved approximately 150 publications across the selected databases. Each article was assessed for eligibility based on the established criteria.
4. **Taxonomy Development**: We developed a taxonomy to categorize the identified literature based on medical applications, such as names of body parts, diagnosis, medical education, patient consultation, and specialties like, telemedicine, clinical decision support, and personalized medicine.
5. **Synthesis of Findings**: The review concluded with a synthesis of findings that highlighted the current use of ChatGPT in healthcare, along with potential risks, and future research directions.

## 2.1. Research Objective

The primary objectives of this review were:

- **RO1**: To provide an overview of ChatGPT, detailing its functionalities and rationale for integration into healthcare systems.
- **RO2**: To explore scope of ChatGPT in healthcare, focusing on its assistance with routine tasks like patient management.
- **RO3**: To see how ChatGPT is utilized in managing various organ-related diseases, analyzing its effectiveness in diagnostics, treatment recommendations, and patient education.
- **RO4**: To identify the significant applications and limitations of ChatGPT, examining its impact on patient care, administrative efficiency, and potential inaccuracies in record handling.

This study seeks to clarify the role of ChatGPT in enhancing healthcare delivery and provide a robust framework for understanding the implications of integrating AI tools into healthcare practices, particularly in the organ-specific context.

## 3. Background

In this section, we will cover key concepts related to Natural Language Processing (NLP), Transformer architecture, Large Language Models (LLMs), enhancements through bidirectional language representation, and the development of ChatGPT.

### 3.1. Natural Language Processing (NLP)

Natural Language Processing (NLP) is a subfield of AI that enables machines to understand, interpret, and generate human language [7]. Emerging from foundational linguistics work in the 1950s and 1960s, NLP connects human communication with computer understanding, facilitating the processing of large text datasets [8].

Key techniques in NLP include tokenization, part-of-speech tagging, named entity recognition, and sentiment analysis, which are essential for various applications [9]. Tokenization facilitates the division of text into manageable units, supporting tasks such as text analysis and machine translation. Part-of-speech tagging enhances syntactic understanding, aiding in grammar checking and text summarization. Named entity recognition identifies and classifies entities, crucial for information extraction and automated customer support. The introduction of transformer architecture and attention mechanisms has significantly improved NLP performance, allowing for the development of advanced models that learn from extensive datasets.

Today, NLP is used in diverse areas such as chatbots, language translation, sentiment analysis, and medical text processing. Despite this, challenges remain, including language ambiguity, context understanding, and biases in training data.

### 3.2. Transformers and Large Language Models (LLMs)

Transformers, introduced in paper titled, "Attention is All You Need" by Vaswani et al. in 2017, revolutionized NLP [10]. Unlike earlier models like recurrent neural networks (RNNs) [11] and Long Short-Term Memory networks (LSTMs) [12], transformers use a self-attention mechanism to assess the importance of words in a sequence.

RNNs and LSTMs process sequences sequentially, which can result in longer training times and difficulties in capturing long-range dependencies. In contrast, transformers enable parallelization during training, significantly speeding up the learning process and effectively capturing relationships between distant words. While LSTMs help mitigate the vanishing gradient problem found in RNNs, they still face challenges with long sequences due to their sequential nature. The self-attention mechanism in transformers thus provides a more efficient understanding of context in text. Mathematically, the self-attention mechanism can be described as follows:

1. **Input Representation**: The input sequence $X$ is transformed into three matrices: queries $Q$, keys $K$, and values $V$.

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

   where $W_Q, W_K, W_V$ are learned weight matrices.

2. **Attention Scores**: The attention scores are computed as the dot product of queries and keys, scaled by the square root of the dimensionality of the keys $d_k$:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

This mechanism allows transformers to focus on relevant parts of the input sequence, enhancing performance across various NLP tasks.

Large Language Models (LLMs), such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT), are built upon the transformer architecture and leverage massive datasets to learn language patterns, structures, and semantics [13–15]. These models have demonstrated remarkable capabilities in generating coherent text, answering questions, and performing language translations.

*3.3. Bidirectional Language Representation*

Bidirectional language representation models, such as BERT (Bidirectional Encoder Representations from Transformers), have significantly enhanced the performance of LLMs in medical tasks [16]. BERT's architecture is based on the transformer model and employs a bidirectional self-attention mechanism that allows it to process input text in both directions—left-to-right and right-to-left [17,18] This feature improves the model's understanding of polysemous words, contextual nuances, and intricate relationships between words in a sentence.

The core components of BERT's architecture include:

- **Multi-Head Self-Attention**: BERT utilizes multiple attention heads to capture various contextual meanings of words simultaneously. Each head learns to focus on different parts of the input sequence, allowing the model to understand complex dependencies.
- **Masked Language Modeling (MLM)**: During pre-training, BERT randomly masks a percentage of input tokens and trains the model to predict these masked tokens based on their context. This approach enables the model to learn a rich understanding of language structures.
- **Next Sentence Prediction (NSP)**: BERT is also trained with a next sentence prediction objective, where it learns to predict whether a given sentence follows another in a text. This helps the model understand sentence relationships, which is crucial for tasks like question answering.

Influential medical LLMs, such as PubMedBERT [19], ClinicalBERT [20], and BioBERT [21], leverage BERT's architecture, fine-tuning it on medical datasets to achieve state-of-the-art performance across various medical Natural Language Processing (NLP) tasks.

Incorporating existing medical knowledge bases, such as the Unified Medical Language System (UMLS) [22], into language models further enhances their capabilities. The integration of domain-specific terminologies and ontologies helps the model understand medical jargon and relationships more effectively.

Moreover, studies have shown that pre-training LLMs on diverse datasets, even those not directly related to healthcare, yield improved performance on medical NLP tasks compared to training solely on domain-specific datasets [23]. This approach highlights the importance of comprehensive data exposure, as it enables models to generalize better and understand a wider range of language variations.

*3.4. ChatGPT*

ChatGPT is a conversational AI model developed by OpenAI, based on the GPT architecture [13]. Its evolution traces back to the initial GPT model introduced in 2018, followed by iterations such as GPT-2 and GPT-3 [24,25]. Each version exhibited exponential growth in parameters and training data, enhancing language understanding and generation capabilities. The architecture is designed to predict the next word in a sentence, leveraging extensive training on diverse internet text.

The evolution of ChatGPT can be summarized as follows:

- **GPT**: The initial model introduced the transformer architecture, outperforming previous RNN models [26].
- **GPT-2**: Increased parameters from 117 million to 1.5 billion, showcasing the ability to generate coherent text; its release was initially withheld due to concerns about misuse [27].
- **GPT-3**: Expanded to 175 billion parameters, significantly enhancing language generation capabilities and gaining attention for its versatility across various tasks with minimal fine-tuning [28].

ChatGPT itself emerged between 2020 and 2021, specifically fine-tuned for conversational tasks [29]. OpenAI later introduced subscription models like ChatGPT Plus to provide users with access to more powerful versions while continuously improving the model's accuracy and reducing biases.

## 4. ChatGPT Applications in Healthcare

The integration of ChatGPT in healthcare is transforming patient care, administrative tasks, and research. This AI technology enhances service delivery through personalized interactions, improved clinical decision-making, and streamlined operations. This section explores key applications of ChatGPT in healthcare, including patient education and support, clinical monitoring, information access, administrative tasks, health promotion, research, and emergency response. Figure 1 gives snippet of the areas covered in this section.
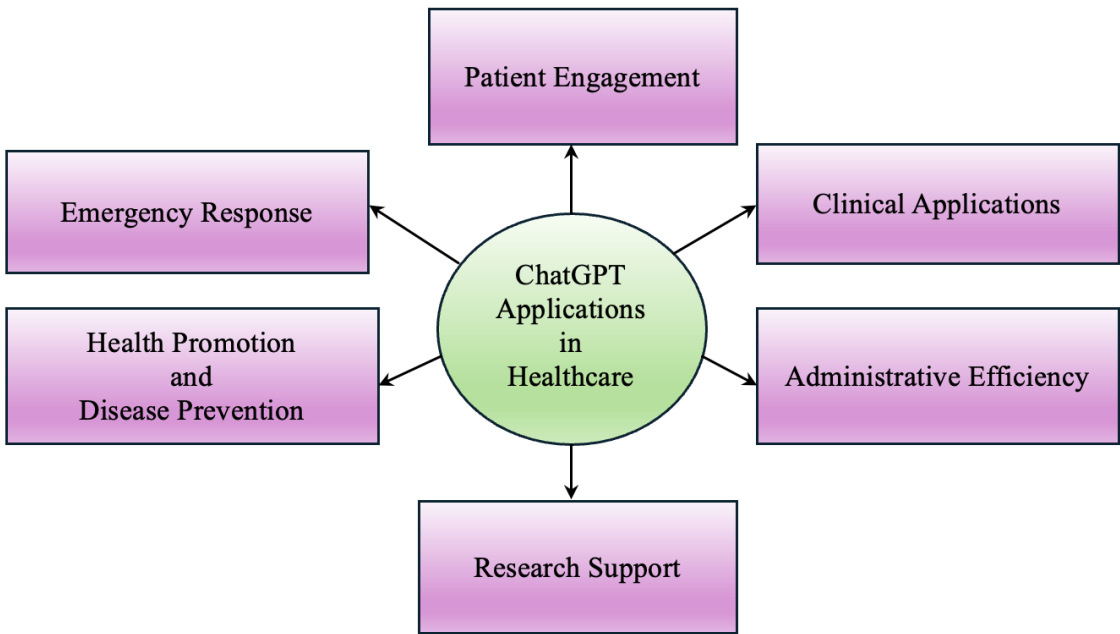


**Figure 1.** ChatGPT Applications in Healthcare.

*4.1. Patient Engagement*

ChatGPT plays a vital role in patient education and support by empowering individuals with knowledge about their health issues [30]. It addresses queries related to medical procedures and

medications. ChatGPT also provides personalized support through individualized health plans and lifestyle recommendations. Additionally, it offers 24/7 assistance and compiles insights on prospective patients. The model delivers timely medication reminders, enhancing adherence by notifying patients about potential side effects and drug interactions [5].

As a virtual health assistant, ChatGPT responds immediately to patients' questions, thereby reducing the workload on healthcare staff and improving patient satisfaction [31]. Through personalized health insights, it empowers patients to make informed decisions, leading to better adherence to treatment plans.

### 4.2. Clinical Applications

In clinical settings, ChatGPT analyzes patient data, medical histories, and symptoms to suggest diagnoses and treatment options, improving decision-making and minimizing diagnostic errors [32]. It supports clinical monitoring by tracking patient health remotely, reminding them to check vital signs, and alerting healthcare providers about potential issues [33]. It also automates routine tasks, such as generating reports and handling chatbot interactions.

In clinical studies, ChatGPT aids in data collection, informs patients about ongoing trials, and enhances the articulation of symptoms. For chronic disease management, it assists patients in monitoring their conditions, sending reminders for check-ups and medications, and managing their daily health routines effectively.

### 4.3. Administrative Efficiency

ChatGPT helps in administrative tasks, such as appointment scheduling, cancellations, and reminders, alleviating the burden on staff and enhancing the patient experience [34]. As a digital assistant, it facilitates data collection and classification from patient records, speeding up assessments and allowing healthcare professionals more time for patient care.

ChatGPT enhances communication by responding to patient inquiries and reducing staff workload through task automation [35]. It can assist in processing insurance claims, verifying patient information, and providing information on billing and payment options. Additionally, ChatGPT can generate summaries of patient interactions, track referral statuses, and monitor follow-up appointments, ensuring continuity of care.

Furthermore, it can help streamline onboarding processes for new staff by providing training materials and answering common questions. In emergency situations, ChatGPT can rapidly disseminate information regarding protocols or alerts to relevant personnel, ensuring a swift response. Overall, these capabilities contribute to a more efficient and responsive administrative framework in healthcare settings.

### 4.4. Research Support

In research and development, ChatGPT supports data analysis, hypothesis testing, and automates literature reviews for efficient findings [36]. It helps identify clinical trials based on patients' medical histories, increasing access to new treatments and enabling researchers to analyze health data to uncover patterns and evaluate intervention effectiveness.

### 4.5. Health Promotion and Disease Prevention

ChatGPT plays a vital role in public health by sharing information on vaccination campaigns, healthy lifestyle tips, and disease prevention strategies [37,38]. It enhances community health literacy and provides personalized advice on nutrition and exercise, empowering patients to adopt healthier habits.

### 4.6. Emergency Response

ChatGPT provides instant information during emergencies, guiding users on first aid procedures and facilitating communication between patients and emergency services [39]. It also offers resources

and support for mental health issues, including coping strategies and exercises, while ensuring access to mental health professionals.

## 5. ChatGPT Applications Across Diverse Organ Systems in Healthcare

Recently, ChatGPT and similar AI models have shown great promise in improving healthcare. By using NLP and ML, these systems are being applied in different medical fields to support diagnostics, treatment planning, and patient care. In the following section, we look at how ChatGPT is used across various organ systems in healthcare, as shown in Figure 2. Each study highlights the main focus, methods, key results, and impact on clinical practice.
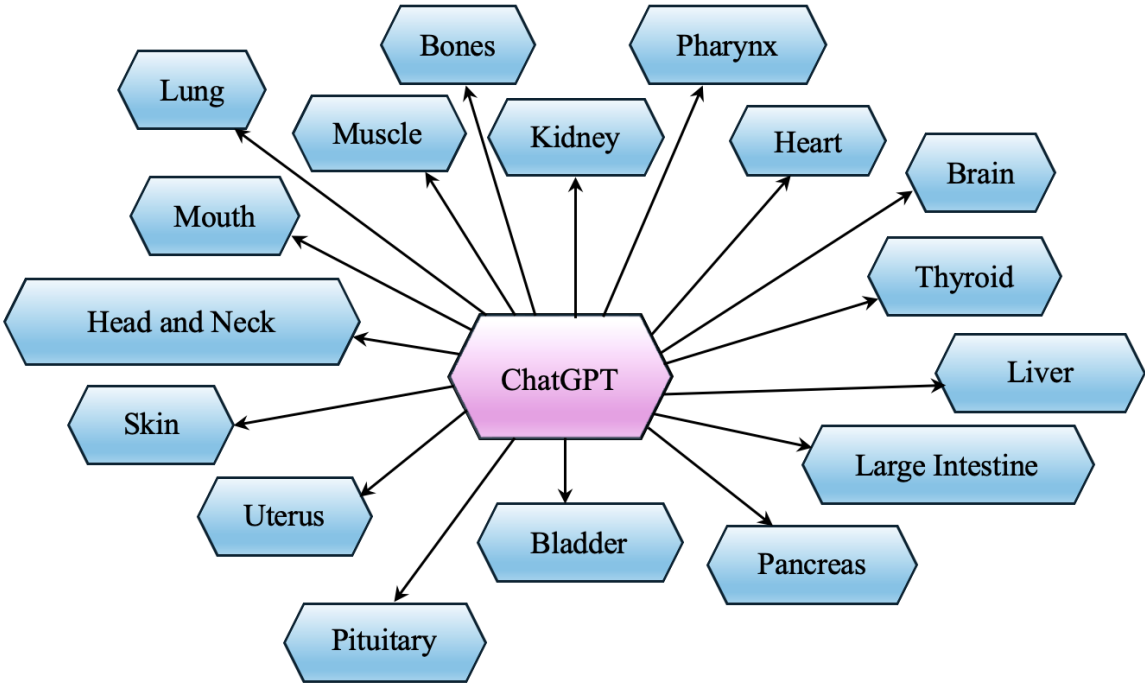


**Figure 2.** ChatGPT Applications Across Diverse Organ Systems in Healthcare.

### 5.1. Kidney

Kidney is a vital organ in the human body responsible for filtering blood to remove waste products, excess fluids, and toxins [40,41]. It plays a crucial role in maintaining overall health by regulating fluid balance, electrolytes (such as sodium and potassium), and blood pressure. Given the complexity of kidney function and its central role in maintaining health, advancements in AI, such as ChatGPT, are increasingly being explored to support kidney disease diagnosis and management. Table 1 summarizes the recent studies which highlight the integration of ChatGPT into the fields of kidney cancer and nephrology.

**Table 1.** Related Studies on ChatGPT in Kidney Cancer and Nephrology.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Choi et al. (2024) [42] | Kidney Cancer | Assessment of ChatGPT responses | Generally appropriate; 70.8% of urologists felt it could not replace specialist consultations. | Highlights the importance of urologist counseling. |
| Miao et al. (2024)[43] | Nephrology | Performance analysis of GPT-4 | Enhanced performance through chain-of-thought prompting and retrieval-augmented generation. | Emphasizes refining AI models for nephrologists. |
| Janus (2023) [44] | Renal Insufficiency | Comparative analysis of ChatGPT 3.5 | Only 5.6% agreement with expert recommendations for managing anticancer drugs. | Necessity of human expertise in complex cases. |
| Łaszkiewicz et al. (2024) [45] | Urothelial Carcinoma | Evaluation of ChatGPT responses | Comprehensible but inadequate and potentially misleading treatment information. | Caution advised for patient inquiries; useful for basic epidemiology. |
| Javid et al. (2024) [46] | Urology Stone | Clinical Performance assessment tool | Higher ratings in accuracy, empathy, completeness, and practicality than urologists, but not reliable as a direct source. | Potential value in supplementing urologist responses; develop FAQs. |
| Miao et al. (2023) [47] | Nephrology Integration | Review of AI integration | Highlights benefits in dataset management, diagnostics, treatment planning, and patient communication. | Need for thorough evaluation and validation of AI in practice. |
| Qarajeh et al. (2023) [48] | Renal Dietary Support | Accuracy evaluation of AI models | ChatGPT 4 excelled in potassium detection; Bard AI achieved perfect phosphorus accuracy. | Potential for AI in renal dietary planning; need for refinement. |

*5.2. Pharynx*

Pharynx, a muscular tube that connects the nose and mouth to the esophagus and larynx, plays a crucial role in swallowing, breathing, and vocalization [49]. Disorders of the pharynx and related structures, such as the larynx, are complex and often require precise diagnosis and treatment strategies.

Lechien et al. evaluates ChatGPT's performance in managing laryngology and head and neck cases [50]. It found that ChatGPT achieved 90% accuracy in differential diagnoses and 60.0-68.0% accuracy in treatment options. However, it was noted that ChatGPT tends to over-recommend tests and misses some important examinations. The findings suggest that while ChatGPT can serve as a promising adjunctive tool in laryngology and head and neck practice, there is a need for refinement in its recommendations for additional examinations.

*5.3. Heart*

Cardiovascular system, responsible for circulating blood and delivering oxygen and nutrients to the body, is critical for maintaining overall health [51]. As cardiovascular diseases remain a leading cause of mortality worldwide, improving patient care and education in this area is essential.

Table 2 summarizes the recent studies which highlight the integration of ChatGPT into the field of cardiovascular health advice, highlighting various studies that evaluate its capabilities, limitations, and implications for patient education and care.

**Table 2.** Related Studies on ChatGPT in Cardiovascular Health

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Lautrup et al. (2023) [52] | Cardiovascular Health Advice | Mixed-methods review | Responses varied in quality; some were dangerously incorrect. | Highlights risk of exacerbating health inequalities. |
| Anaya et al. (2024) [53] | Heart Failure Education | Readability evaluation | Answers longer but readable; low actionability score. | AI chatbots can enhance patient education, further research needed. |
| King et al. (2024) [54] | Heart Failure Question and Answers | Knowledge evaluation of GPT-3.5 and GPT-4 | GPT-4 showed 100.0% accuracy; GPT-3.5 had > 94.0% accuracy. | ChatGPT could be a valuable educational resource for patients. |
| BULBOACĂ et al. (2024) [55] | Heart Pathophysiology | Comparative analysis of responses | ChatGPT answered 4/5 questions correctly; responses lacked completeness. | Requires expert supervision for validation of responses. |
| Chlorogiannis et al. (2023) [56] | Cardiovascular Diseases | Review of applications | Discusses potential benefits and limitations of ChatGPT in diagnosis and management. | Emphasizes ethical and equitable use to maximize benefits. |

*5.4. Brain*

Brain, as the central organ of the nervous system, plays a crucial role in controlling bodily functions, processing information, and facilitating cognitive processes such as memory, learning, and emotional regulation [57]. Given the complexity of neurological disorders and the increasing prevalence of brain-related health issues, enhancing our understanding and management of brain health is essential. Keeping this in mind, Table 3 summarizes the recent studies which highlight the integration of ChatGPT into the field of brain health.

**Table 3.** Related Studies on ChatGPT in Brain Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Kozel et al. (2024) [58] | Brain Tumors | Performance evaluation of ChatGPT-3.5 and 4 | ChatGPT-4 achieved 85.0% accuracy in diagnoses and 75.0% in treatment plans. Significantly outperforms ChatGPT-3.5. | Shows potential as a diagnostic tool in neuro-oncology. |
| Adesso (2023) [59] | Brain related Discovery | Method for theory evaluation | Demonstrated ChatGPT's ability to benchmark physical theories through a gamified environment; promotes AI-human collaboration. | Highlights importance of effective AI integration in research. |
| Fei et al. (2024) [60] | Cognitive Function Assessment | Performance evaluation of GPT-3.5 and GPT-4 | Identified significant discrepancies in memory and speech evaluations, particularly with GPT-3.5; refinements improved alignment with physician assessments for GPT-4. | Suggests GPT-4's potential as a supplementary tool in cognitive evaluations; highlights need for further development in scoring methods and interaction protocols to enhance accuracy. |

*5.5. Thyroid*

Thyroid, a butterfly-shaped gland located in the neck, plays a crucial role in regulating metabolism, energy levels, and overall hormonal balance [61]. Given its importance in various physiological processes, recent studies are highlighting the integration of ChatGPT into the field of thyroid health. Table 4 provides a summary of these studies.

**Table 4.** Related Studies on ChatGPT in Thyroid Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Köroğlu et al. (2023) [62] | Thyroid Nodules Management | Effectiveness assessment by endocrinologists | Mostly correct and reliable answers; not suitable as primary resource for physicians. | Can serve as an informative tool but requires professional oversight. |
| Sievert et al. (2024) [63] | Risk Stratification of Thyroid Nodules | Assessment using Thyroid Imaging Reporting and Data System. | Moderate potential with sensitivity of 86.7% and overall accuracy of 68.0%. | Assists in personalized treatment decisions, but needs further validation. |
| Stevenson et al. (2024) [64] | Thyroid Function Test Interpretation | Comparison with practicing biochemists | ChatGPT and Google Bard only interpreted 33.3% and 20.0% correctly, respectively. | Safety concerns highlighted; AI cannot replace human consultations for test interpretation. |
| Helvaci et al. (2024) [65] | Thyroid Cancer Information | Accuracy and reliability assessment | Moderately accurate (76.7%) for general information; effective in offering emotional support. | Useful for general inquiries but insufficient for specific case management. |

Cazzato et al. conducted a review to evaluate the potential of ChatGPT in the field of pathology, analyzing five relevant publications out of an initial 103 records [66]. The findings indicated that while ChatGPT holds promise for assisting pathologists by providing substantial amounts of scientific data, it also faces significant limitations, including outdated training data and the occurrence of hallucinations. The review featured a query session addressing various pathologies, emphasizing that ChatGPT can aid the diagnostic process but should not be relied upon for clinical decision-making. Overall, the study concluded that ChatGPT's role in pathology is primarily supportive, necessitating further advancements to overcome its current challenges.

### 5.6. Liver

Liver, an organ responsible for numerous functions including detoxification, metabolism, and production of essential proteins, plays a crucial role in maintaining overall health [67]. Due to its significance in various diseases, many research studies are exploring applications of ChatGPT, to enhance the understanding and management of liver-related conditions.

In related works, Yeo et al. assessed ChatGPT's accuracy and reproducibility in answering questions about cirrhosis and hepatocellular carcinoma (HCC) management [68]. The study reported high overall accuracy, with ChatGPT scoring 79.1% for cirrhosis and 74.0% for HCC. However, comprehensive responses were limited, particularly in areas of diagnosis and regional guidelines. Despite these limitations, ChatGPT provided practical advice for patients and caregivers, suggesting its potential as an adjunct informational tool to improve patient outcomes in cirrhosis and HCC management.

In another study, Yeo et al. compared the capabilities of ChatGPT and GPT-4 in responding to cirrhosis-related questions across multiple languages, including English, Korean, Mandarin, and Spanish [69]. The results indicated that GPT-4 significantly outperformed ChatGPT in both accuracy and comprehensiveness, especially in non-English responses, with notable improvements in Mandarin and Korean. This underscores GPT-4's potential to enhance patient care by addressing language barriers and promoting equitable health literacy globally.

### 5.7. Large Intestine

Large intestine, a crucial component of the gastrointestinal system, is responsible for absorbing water and electrolytes from indigestible food matter, as well as storing and eliminating waste products [70]. Given the prevalence of gastrointestinal disorders, ChatGPT is being explored to enhance diagnosis, management, and patient education related to large intestine health. Table 5 provides an overview of its integration into gastrointestinal pathology and large intestine care.

**Table 5.** Related Studies on ChatGPT in Gastrointestinal Pathology and Large Intestine Management.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Cankurtaran et al. (2023) [71] | Inflammatory Bowel Disease | Performance evaluation for healthcare professionals and patients | Professional-directed responses scored higher in reliability and usefulness than patient-directed ones. | Highlights potential as an informative tool; need for improvement in information quality for both patients and professionals. |
| Ma (2023) [72] | Gastrointestinal Pathology | Assessment of applications in digital pathology | Benefits in summarizing charts, education, and research; limitations include biases and inaccuracies from training datasets. | Emphasizes enhancement of human expertise in healthcare quality rather than replacement. |
| Liu et al. (2024) [73] | Colonoscopy Assessment | Accuracy evaluation using the Boston Bowel Preparation Scale | Achieved accuracy lower than experienced endoscopists, indicating need for fine-tuning. | Suggests potential for scoring but underscores the necessity of professional expertise. |

*5.8. Pancreas*

Pancreas, an essential gland located behind the stomach, plays a vital role in digestion and glucose regulation by producing digestive enzymes and hormones such as insulin [74]. Recognizing its significance in metabolic and digestive disorders, ChatGPT is being applied to improve the understanding and management of pancreatic health.

In one of the related works, Du et al. assessed the performance of ChatGPT-3.5 and ChatGPT-4.0 in answering questions related to acute pancreatitis (AP) [75]. The study found that ChatGPT-4.0 achieved a higher accuracy rate than ChatGPT-3.5, answering 94.0% of subjective questions correctly compared to 80%. It also performed better on objective questions, with an accuracy of 78.1% versus 68.5%, with a statistically significant difference (P = 0.01). The concordance rates between the two versions were reported as 80.8% for ChatGPT-3.5 and 83.6% for ChatGPT-4.0. Both models excelled particularly in the etiology category, highlighting their potential utility in improving awareness and understanding of acute pancreatitis.

Qiu et al. evaluated the accuracy of ChatGPT-3.5 in answering clinical questions based on the 2019 guidelines for severe acute pancreatitis [76]. The results indicated that ChatGPT-3.5 was more accurate when responding in English (71.0%) compared to Chinese (59.0%), although the difference was not statistically significant (P = 0.203). Furthermore, the model performed better on short-answer questions (76%) compared to true/false questions (60.0%) (P = 0.405). While ChatGPT-3.5 shows potential value for clinicians managing severe acute pancreatitis, the study suggests it should not be overly relied upon for clinical decision-making.

*5.9. Bladder*

Bladder, a hollow muscular organ, plays a crucial role in storing and expelling urine, which is essential for maintaining fluid balance and overall health [77]. Table 6 summarizes recent studies which have explored the integration of ChatGPT into the field of bladder health.

**Table 6.** Related Studies on ChatGPT in Urology - Focus on Bladder Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Guo et al. (2024) [78] | Bladder Cancer Patient Education | Evaluation of ChatGPT's answers to hypothetical patient questions | ChatGPT provided coherent but incomplete responses, with accuracy scores between 3.7 and 6.0. | Further AI development is necessary to enhance its response completeness in urological consultations. |
| Braga et al. (2024) [79] | Urological Diagnoses (including Bladder Cancer) | Assessment of ChatGPT's responses to specific urological conditions | ChatGPT provided partially correct answers, but critical details were missing in certain conditions. | Caution is needed when using AI for clinical decision-making due to incomplete responses. |
| Ozgor et al. (2024) [80] | Urological Cancers (including Bladder Cancer) | Comparison of ChatGPT responses to the European Association of Urology (EAU) Guideline | ChatGPT achieved high quality scores for FAQs, but scored lower on guideline-based questions. | Promising for general inquiries, but guideline-aligned responses need improvement. |
| Cakir et al. (2024) [81] | Urogenital Tract Infections (UTIs) Information | Comparative analysis of FAQs and EAU guidelines | Achieved 96.2% accuracy for FAQs; 89.7% scored Global Quality Score (GQS) 5 for guideline responses, reproducibility > 90.0%. | ChatGPT is reliable for public and guideline-based inquiries regarding UTIs. |
| Sagir et al. (2022) [82] | Urological Disease Accuracy | Comparative analysis of 112 questions | Accuracy levels: 40.0% for Urolithiasis, 50% for Bladder cancer, 63.6% for Renal cancer, 52% for Urethroplasty; emphasizes potential as a supportive tool but needs human oversight. | Highlights the need for careful interpretation and clinical supervision before making medical decisions based on ChatGPT responses. |
| Cakir et al. (2024) [83] | Urolithiasis Information | Comparative analysis of FAQs and EAU guidelines | ChatGPT correctly answered 94.6% of FAQs with no completely incorrect responses; 83.3% top score for guideline questions. | Can be a valuable tool in urology clinics, aiding patient understanding when supervised by urologists. |
| Szczesniewski et al. (2023) [84] | Bladder Cancer and Other Urological Diseases | Analysis using DISCERN questionnaire | ChatGPT provided well-balanced information but the overall quality was moderate. | Users should apply caution due to potential biases and lack of source citation. |

## *5.10. Pituitary*

Pituitary gland, often referred to as the "master gland," plays a critical role in regulating various hormonal functions throughout the body, including growth, metabolism, and stress response [85]. Given its central role in endocrine health, ChatGPT is being studied to improve the understanding and management of pituitary disorders, including adenomas.

In one of the related works, Sambangi et al. evaluated the accuracy, readability, and grade level of ChatGPT responses regarding pituitary adenoma resection, using different prompting styles: physician-level, patient-friendly, and no prompting as a control [86]. The study found that responses without prompting were longer, while physician-level and patient-friendly prompts resulted in more concise answers. Patient-friendly prompting led to significantly easier-to-read responses. The accuracy of responses was highest with physician-level prompting, although the differences among prompting styles were not statistically significant due to the small sample size. Overall, the study suggests that ChatGPT has potential as a patient education tool, though further development and data collection are needed.

Şenoymak et al. assessed ChatGPT's ability to respond to 46 common queries regarding hyperprolactinemia and prolactinoma, evaluating accuracy and adequacy using Likert scales [87]. The median accuracy score was 6.0, indicating high accuracy, while the adequacy score was 4.5, reflecting generally adequate responses. Significant agreement was found between two independent endocrinologists assessing the responses. However, pregnancy-related queries received the lowest scores for both accuracy and adequacy, indicating limitations in ChatGPT's responses in medical contexts. The findings suggest that while ChatGPT shows promise, there is a need for improvement, particularly regarding pregnancy-related information.

Taşkaldıran et al. examined the accuracy and quality of ChatGPT-4's responses to ten hyperparathyroidism cases discussed at multidisciplinary endocrinology meetings [88]. Two endocrinologists independently scored the responses for accuracy, completeness, and overall quality. Results showed high mean accuracy scores (4.9 for diagnosis and treatment) and completeness scores (3.0 for diagnosis, 2.6 for further examination, and 2.4 for treatment). Overall, 80.0% of responses were rated as high quality, suggesting that ChatGPT can be a valuable tool in healthcare, though its limitations and risks should be considered.

### 5.11. Uterus

Uterus, a vital organ in the female reproductive system, plays a crucial role in menstruation, pregnancy, and childbirth [89]. Recognizing its importance in women's health, application of ChatGPT to improve the understanding, diagnosis, and management of uterine and gynecological conditions is being explored.

Table 7 summarizes the integration of ChatGPT into the field of uterus and gynecologic health.

**Table 7.** Related Studies on ChatGPT in Uterus Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Patel et al. (2024) [90] | Genetic Counseling for Gynecologic Cancers | Assessment of 40 questions with oncologist input | ChatGPT achieved 82.5% accuracy; 100.0% accuracy in genetic counseling category; 88.2% for hereditary breast/ovarian cancer; 66.6% for Lynch syndrome. | ChatGPT could be a valuable resource for patient information, needing further oncologist input for comprehensive education. |
| Peled et al. (2024) [91] | Obstetric Questions from pregnant individuals | Evaluation by 20 obstetric experts | 75.0% of responses rated positive; accuracy mean of 4.2; completeness and safety lower, at means of 3.8 and 3.9. | ChatGPT can provide accurate obstetric responses but requires caution regarding maternal and fetal safety. |
| Psilopatis et al. (2024) [92] | Intrauterine Growth Restriction | Assessment of comprehension of S2k (a specific set of clinical practice guidelines developed by the German Society for Gynecology and Obstetrics (DGGG)) | Most responses about definitions and timing were adequate; over half of delivery mode suggestions needed correction. | ChatGPT could assist in clinical practice but responses require expert supervision for accuracy. |
| Winograd et al. (2024) [93] | Female Puberty | Evaluation of responses to ten puberty questions | 60.0% of responses deemed acceptable; 40.0% unacceptable; no verifiable references provided. | While generally accurate, further study and development are needed before endorsing ChatGPT for adolescent health information. |

### 5.12. Skin

Skin, the body's largest organ, serves as a critical barrier protecting against external threats while playing essential roles in thermoregulation, sensation, and immune response [94].

Given its importance, Lantz examined the use of ChatGPT in a case report involving a critically ill african american woman diagnosed with toxic epidermal necrolysis (TEN), which affected over 30.0% of her body surface area [95]. The condition, triggered by medications, poses a high mortality

risk, and the report highlighted the challenges of identifying the offending drug due to the patient's complex medical history. It also discussed potential genetic or epigenetic predispositions in African Americans to conditions such as Stevens-Johnson syndrome (SJS) and TEN, underscoring the necessity for increased representation of skin of color in medical literature. While the report acknowledged the advantages of utilizing ChatGPT in medical documentation, it also pointed out its limitations and the need for careful consideration of its use in clinical settings.

Table 8 summarizes the integration of ChatGPT into the field of dermatology and skin health in the recent works.

**Table 8.** Related Studies on ChatGPT in Skin Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Sanchez-Zapata et al. (2024) [96] | Inflammatory Dermatoses | Evaluated ChatGPT's quality in answering questions on conditions like acne and psoriasis, rated by dermatology residents | Responses were generally rated between "acceptable" to "very good," with median scores around 4, indicating potential as a patient information tool. | Suggests that ChatGPT can provide valuable primary information on skin conditions when used cautiously by clinicians. |
| Passby et al. (2024) [97] | Dermatology Examination Performance | Assessed ChatGPT-3.5 and ChatGPT-4 on 84 questions from the Specialty Certificate Examination in Dermatology | ChatGPT-3.5 scored 63.0%, while ChatGPT-4 achieved 90.0%, exceeding typical pass marks and highlighting its potential in medical education. | Indicates that advanced AI can effectively answer clinical questions, though its limitations in complex cases must be acknowledged for patient safety. |
| Stoneham et al. (2024) [98] | Diagnostic Accuracy in Dermatology | Compared ChatGPT's diagnostic capabilities with specialists using dermatologist-provided and nonspecialist data | ChatGPT diagnosed correctly 56.0% of the time with specialist data and 39.0% with nonspecialist data, showing some ability but less than dermatologists (83.0%). | Highlights ChatGPT's potential for providing differential diagnoses but indicates it does not yet enhance overall diagnostic accuracy significantly. |
| Mondal et al. (2023) [99] | Dermatological diseases Education | Evaluated ChatGPT's capability in generating educational content on dermatological diseases | Generated texts averaged 377 words with satisfactory accuracy; however, a high text similarity index (27.1%) raised plagiarism concerns. | Suggests that while ChatGPT can produce useful educational content, generated texts should be reviewed by doctors to mitigate plagiarism risks. |

### 5.13. Head and Neck

Head and neck region houses various critical structures, including the oral cavity, pharynx, larynx, and salivary glands, and is integral to functions such as breathing, swallowing, and speech [100].

Keeping this in mind, Vaira et al. evaluated the accuracy of ChatGPT-4 in answering clinical questions and scenarios related to head and neck surgery [101]. The study involved 18 surgeons across 14 Italian units and assessed a total of 144 clinical questions and 15 scenarios. ChatGPT achieved a median accuracy score of 6 (IQR: 5-6) and a completeness score of 3 (IQR: 2-3). Notably, 87.2% of the answers were deemed nearly correct, while 73.0% of the responses were considered comprehensive. The AI model successfully answered 84.7% of closed-ended questions and provided correct diagnoses in 81.7% of the scenarios. Only 56.7% of the proposed procedures were complete, and the quality of bibliographic references was lacking, with 46.4% of responses missing sources. Despite, ChatGPT shows promise in addressing complex scenarios, it is not yet a reliable tool for specialist decision-making in head and neck surgery.

### 5.14. Mouth

Mouth is essential for functions like eating, speaking, and breathing, and its health significantly impacts overall well-being [102]. Early detection of oral cancer is crucial for improving treatment outcomes, and AI applications like ChatGPT are being applied for their potential to enhance awareness and education about oral health.

In recent work, Hassona et al. evaluated the quality, reliability, readability, and usefulness of ChatGPT in promoting early detection of oral cancer [103]. The study analyzed a total of 108 patient-

oriented questions, with ChatGPT providing "very useful" responses for 75.0% of the inquiries. The mean Global Quality Score was 4.24 out of 5, and the reliability score was high, achieving 23.17 out of 25. However, the mean actionability score was notably lower at 47.3%, and concerns were raised regarding readability, reflected in a mean Flesch-Kincaid Score (FKS) reading ease score of 38.4%. Despite these readability challenges, no misleading information was found, suggesting that ChatGPT could serve as a valuable resource for patient education regarding oral cancer detection. Table 9 summarizes ChatGPT's integration into mouth health.

**Table 9.** Related Studies on ChatGPT in Mouth Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Babayiğit et al. (2023) [104] | Periodontal diseases and dental implants | Evaluated accuracy and completeness of ChatGPT responses to 70 FAQs in periodontology | Median accuracy score of 6 and completeness score of 2, indicating responses were "nearly completely correct" and "adequate". Highest accuracy in peri-implant diseases. | ChatGPT can be a useful informational resource, but expert supervision is essential due to potential inaccuracies. |
| Mago and Sharma (2023) [105] | Oral and Maxillofacial Radiology | Assessed ChatGPT-3's ability to identify radiographic anatomical landmarks and understand pathologies using an 80-question questionnaire | Achieved 100.0% accuracy in describing radiographic landmarks, with mean scores of 3.94, 3.85, and 3.96 across categories. | Effective as an adjunct tool for information, but lacks detail, limiting its use as a primary reference; can enhance knowledge and reduce patient anxiety. |
| Puladi et al. (2024) [106] | Oral and Maxillofacial Surgery | Reviewed the impact of LLMs like ChatGPT in OMS, identifying 57 records with 37 relevant studies focusing on GPT-3.5 and GPT-4 | Current research is limited, primarily addressing scientific writing and patient communication, with classic OMS diseases underrepresented. | While LLMs may enhance certain healthcare aspects, ethical and regulatory concerns need to be resolved before widespread adoption. |

*5.15. Lung*

Lungs are essential organs in the respiratory system, responsible for gas exchange and oxygenating the blood. Lung health is critical, as conditions such as lung cancer can significantly impact overall well-being and quality of life [107]. Effective diagnosis and management of lung-related diseases require accurate data extraction and analysis from medical records, an area where ChatGPT is being explored for their potential.

Fink et al. compared the performance of ChatGPT and GPT-4 in extracting oncologic phenotypes from free-text CT reports for lung cancer [108]. The study analyzed a total of 424 reports and found that GPT-4 significantly outperformed ChatGPT in several key areas, including extracting lesion parameters (98.6% for GPT-4 vs. 84.0% for ChatGPT), identifying metastatic disease (98.1% vs. 90.3%), and labeling oncologic progression, where GPT-4 achieved an F1 score of 0.96 compared to 0.91 for ChatGPT. Additionally, GPT-4 scored higher on measures of factual correctness (4.3 vs. 3.9) and accuracy (4.4 vs. 3.3) on a Likert scale, with a notably lower confabulation rate (1.7% vs. 13.7%). Overall, the findings indicate that GPT-4 demonstrated superior capability in data mining from medical records related to lung cancer. Table 10 summarizes the integration of ChatGPT into the field of lung health.

**Table 10.** Related Studies on ChatGPT in Lung Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Rahsepar et al. (2023) [109] | Lung Cancer | Comparison of ChatGPT-3.5, Google Bard, Bing, and Google search engines on 40 questions | ChatGPT-3.5 achieved 70.8% accuracy, outperforming Google Bard at 51.7%, Bing at 61.7%, and Google search at 55.0%. Notably, ChatGPT-3.5 and Google search demonstrated greater consistency in their responses. | Highlights risks in ChatGPT accuracy, indicating a critical need for reliable health information tools that can enhance patient understanding and decision-making. |
| Nakamura et al. (2023) [110] | Lung Cancer Staging Automation | Compared GPT-3.5 Turbo and GPT-4 on 135 reports using TNM classification rule | GPT-4 achieved the highest accuracy: 52.2% for T, 78.9% for N, and 86.7% for M categories. The main errors were attributed to challenges in numerical reasoning and anatomical knowledge. | Suggests potential for automating lung cancer staging, emphasizing the need for further enhancements in numerical reasoning and anatomical knowledge to improve clinical utility. |
| Lee et al. (2024) [111] | Lung Cancer Staging Performance Comparison | Evaluated three ChatGPT large language models (LLMs) vs. human readers on 700 patients' reports | GPT-4o achieved the highest overall accuracy at 74.1%, followed by GPT-4 at 70.1% and GPT-3.5 at 57.4%. All LLMs performed below fellowship-trained radiologists (82.3% and 85.4%). | Indicates that while LLMs can assist in lung cancer staging, they should not replace expert radiologists for complex assessments, reinforcing the importance of domain expertise. |
| Schulte et al. (2023) [112] | Lung Cancer Treatment Identification | Evaluated ability to identify therapies for 51 advanced solid cancer diagnoses | ChatGPT successfully identified 91 distinct medications, achieving a valid therapy quotient (VTQ) of 0.77, which demonstrates good concordance with National Comprehensive Cancer Network (NCCN) guidelines, providing at least one NCCN-recommended therapy for each malignancy. | Shows promise in assisting oncologists with treatment decision-making, but underscores the need for accuracy improvements to maximize its clinical utility in oncology. |

## 5.16. Bone

Bones are crucial components of the human skeletal system, providing structure, support, and protection to vital organs [113]. Bone health is essential for overall well-being, as conditions such as osteoporosis can lead to increased fracture risk and diminished quality of life. Ensuring accurate information on bone health and associated disorders is vital for patient education and management.

In a related study, Ghanem et al. evaluated the accuracy of ChatGPT-3.5 in providing evidence-based answers to 20 frequently asked questions about osteoporosis [114]. The responses were reviewed by three orthopedic surgeons and one advanced practice provider, resulting in an overall mean accuracy score of 91.0%. The responses were categorized as either "accurate requiring minimal clarification" or "excellent," with no answers found to be inaccurate or harmful. Additionally, there were no significant differences in accuracy across categories such as diagnosis, risk factors, and treatment. While ChatGPT demonstrated high-quality educational content, the authors recommend it as a supplement to, rather than a replacement for, human expertise and clinical judgment in patient education. Table 11 summarizes recent work in the field of integrating ChatGPT with bone health.

**Table 11.** Related Studies on ChatGPT in Bone Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Son et al. (2023) [115] | Bone Metastases Diagnosis | Developed a deep learning model using ChatGPT-3.5 and ResNet50 on bone scans from 4,626 cancer patients | The model achieved an AUC of 81.6%, sensitivity of 56.0%, and specificity of 88.7%. Class activation maps revealed a focus on spinal metastases but confusion with benign lesions. | Suggests that clinicians with basic programming skills can effectively leverage AI for medical image analysis, potentially improving clinical decision-making and diagnostics. |
| Cinar (2023) [116] | ChatGPT's Knowledge of Osteoporosis | Assessed responses to 72 FAQs based on National Osteoporosis Guideline Group guidelines | ChatGPT achieved an overall accuracy of 80.6%, highest in prevention (91.7%) and general knowledge (85.8%). However, only 61.3% aligned with guidelines, indicating limitations. | While showing adequate performance, the study highlights the need for improvements in adherence to clinical guidelines for reliable patient education. |
| Yang et al. (2024) [117] | Diagnostic Accuracy of Bone Tumors | Evaluated ChatGPT's performance on 1,366 imaging reports diagnosed by experienced physicians | Initial diagnostic accuracy was 73.0%, improving to 87.0% with few-shot learning, achieving sensitivity of 99.0% and specificity of 73.0%. Misdiagnoses included benign cases misidentified as malignant. | Highlights ChatGPT's potential to enhance diagnostic processes for bone tumors, while emphasizing the need for collaboration with experienced physicians in clinical settings to mitigate misdiagnosis risks. |

### 5.17. Muscles

Muscle health is vital for overall physical function, mobility, and quality of life [118]. It includes the maintenance and improvement of muscle strength, endurance, and flexibility, which are essential for daily activities and overall well-being. Understanding muscle-related conditions and their management is crucial for effective rehabilitation and enhancing patient outcomes.

In related works, Sawamura et al. evaluated ChatGPT 4.0's performance on Japan's national exam for physical therapists, specifically assessing its ability to handle complex questions that involve images and tables [119]. The study revealed that ChatGPT achieved an overall accuracy of 73.4%, successfully passing the exam. Notably, it excelled in text-based questions with an accuracy of 80.5%, but faced challenges with practical questions, achieving only 46.6%, and those requiring visual interpretation, where it scored 35.4%. The findings suggest that while ChatGPT shows promise for use in rehabilitation and Japanese medical education, there is a significant need for improvements in its handling of practical and visually complex questions.

In a study, Agarwal et al. evaluated the capabilities of ChatGPT, Bard, and Bing in generating reasoning-based multiple-choice questions (MCQs) in medical physiology for MBBS students [120]. ChatGPT and Bard produced a total of 110 MCQs, while Bing generated 100, encountering issues with two competencies. Among the models, ChatGPT achieved the highest validity score of 3, while Bing received the lowest, indicating notable differences in performance. Despite these variations, all models received comparable ratings for difficulty and reasoning ability, with no significant differences observed. The findings underscore the need for further development of AI tools to enhance their effectiveness in creating reasoning-based MCQs for medical education. Table 12 summarizes the integration of ChatGPT into the field of muscle health.

**Table 12.** Related Studies on ChatGPT in Muscle Health.

| Study | Focus Area | Methodology | Key Findings | Implications |
|---|---|---|---|---|
| Saluja and Tigga (2024) [121] | Anatomy Education | Evaluated ChatGPT-4's effectiveness in explaining anatomical structures, generating quizzes, and summarizing lectures | ChatGPT proved useful for clinical relevance explanations and summarizing material. However, it struggled with accurately depicting anatomical images, especially complex structures. | Highlights the potential of ChatGPT as an educational tool for medical students, emphasizing that while it can enhance teaching, it cannot replace the role of teachers in anatomy education. |
| Kaarre et al. (2023) [122] | Information on ACL Surgery | Evaluated ChatGPT's responses to ACL surgery-related questions aimed at patients and non-orthopaedic medical doctors, with assessments from four orthopaedic surgeons | ChatGPT achieved approximately 65.0% accuracy, demonstrating adaptability in providing relevant information, but it should be viewed as a supplementary tool rather than a replacement for orthopaedic expertise due to its limitations in understanding complex medical concepts. | Reinforces the notion that ChatGPT can aid in patient education but cannot substitute the nuanced understanding of experienced medical professionals. |
| Meng et al. (2024) [123] | Genu Valgum Prediction | Developed a deep learning architecture for predicting genu valgum using non-contact pose analysis data combined with ChatGPT-generated features from subject images | The combined approach outperformed a baseline model, achieving an accuracy of 77.2%, showcasing ChatGPT's effectiveness in semantic information extraction for medical imaging applications. | Promises a method for assessing genu valgum, emphasizing the potential of ChatGPT in enhancing the capabilities of traditional medical assessments through integrated technologies. |
| Mantzou et al. (2024) [124] | Quality of Responses on Musculoskeletal Anatomy | Assessed the efficacy of ChatGPT in answering questions related to musculoskeletal anatomy at different time points, rated by three experts using a 5-point Likert scale | Results showed variability in response quality; 50.0% of responses were rated as good quality, and 66.6% consistent across time points. However, low-quality responses frequently contained significant mistakes or conflicting information. | Indicates that while ChatGPT can provide useful insights, its reliability as an independent learning resource for musculoskeletal anatomy is limited, necessitating validation against established anatomical literature. |
| Li et al. (2023) [125] | Mobile Rehabilitation for Osteoarthritis | Evaluated the clinical efficacy and cost-effectiveness of a mobile rehabilitation system integrating ChatGPT-4 and wearable devices for patients with osteoarthritis and sarcopenia in a prospective randomized trial | 278 patients will be assigned to an intervention group receiving personalized exercise therapy through mobile platforms and wearables, while a control group receives traditional face-to-face therapy. Outcome measures will include pain assessment and functional scores at multiple time points over six months. | Suggests that integrating ChatGPT with wearable technology could enhance rehabilitation service efficiency and availability, potentially improving therapeutic outcomes for patients with muscle-related conditions. |

## 6. Potential Risks of ChatGPT in Healthcare

ChatGPT offers promising applications in healthcare. However, its integration also presents several potential risks. These risks can impact patient safety, the efficacy of healthcare delivery, and the ethical landscape surrounding AI in medicine. This section outlines key concerns related to the use of ChatGPT in healthcare settings, emphasizing the importance of cautious implementation and oversight. Figure 3 points out these risks.
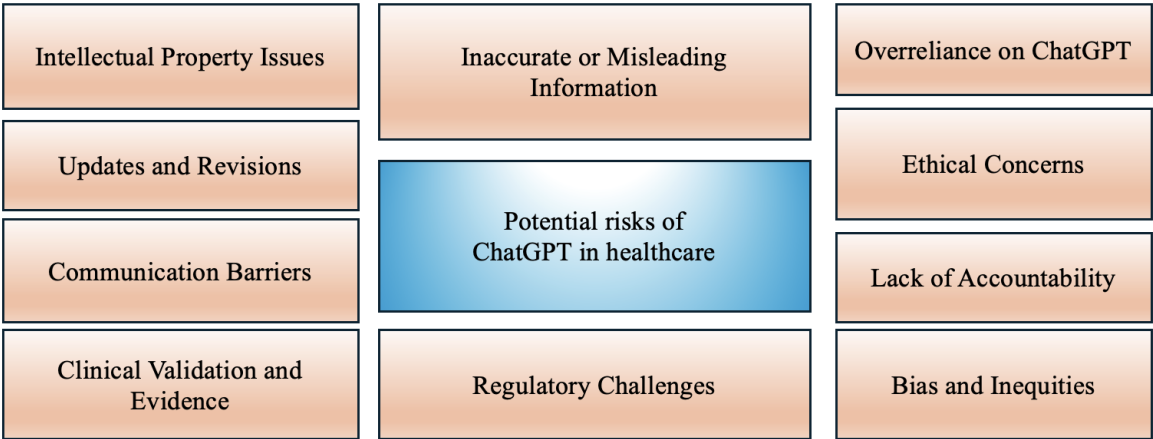
**Figure 3.** Potential Risks of ChatGPT in Healthcare

### 6.1. Inaccurate or Misleading Information

ChatGPT can generate inaccurate or misleading information due to its reliance on vast datasets that contain errors [126]. This risk is significant when patients depend on ChatGPT-generated advice for self-diagnosis or treatment, potentially leading to misdiagnosis or delayed care. The model lacks the nuanced understanding of individual patient situations, reinforcing the need to use ChatGPT as a supplementary tool rather than a primary source of medical guidance.

### 6.2. Overreliance on ChatGPT

The integration of ChatGPT in healthcare can lead to overreliance among patients and professionals [127]. Patients may depend on it for health queries instead of consulting healthcare providers, risking delayed diagnoses. Healthcare professionals might also defer too much to ChatGPT, undermining their critical thinking and diagnostic skills, which are essential for accurate patient care.

### 6.3. Ethical Concerns

Ethical implications include issues of informed consent, as patients misunderstand that ChatGPT information is not a substitute for professional medical advice [128]. Privacy and confidentiality concerns arise when sharing sensitive health information with ChatGPT, highlighting the need for robust data protection measures.

### 6.4. Lack of Accountability

ChatGPT lacks legal accountability for its information, complicating liability when harm results from incorrect advice [129,130]. This raises ethical questions about responsibility, especially if patients misinterpret ChatGPT responses.

### 6.5. Bias and Inequities

ChatGPT training datasets may contain societal biases, leading to inequitable healthcare advice that disadvantages specific demographics [131]. This risk exacerbates existing disparities in healthcare access and outcomes, necessitating careful scrutiny of its applications in diverse populations.

### 6.6. Regulatory Challenges

The rapid advancement of AI technologies often outpaces existing regulatory frameworks, which are not adequately address the unique challenges posed by models like ChatGPT [132]. Comprehensive regulations and quality control measures are needed to ensure safety and efficacy in clinical settings.

*6.7. Clinical Validation and Evidence*

Current research on ChatGPT efficacy in clinical settings is limited and often lacks rigorous clinical validation [133,134]. Ensuring its applications are evidence-based is crucial for patient safety and effective healthcare practices.

*6.8. Communication Barriers*

Communication barriers can arise from misinterpretations by ChatGPT, leading to incorrect or irrelevant responses [135]. Additionally, the absence of emotional support and empathy in ChatGPT interactions can negatively affect patient care and increase anxiety.

*6.9. Updates and Revisions*

The fast-evolving nature of medical knowledge poses challenges for ChatGPT [136]. If not regularly updated, it may disseminate outdated advice, which can severely impact patient health. Continuous updates and revisions are essential for maintaining accuracy in clinical practice.

*6.10. Intellectual Property Issues*

Using ChatGPT for generating medical content raises intellectual property concerns [137]. Healthcare professionals might inadvertently engage in plagiarism or fail to provide proper attribution, compromising the integrity of their work and raising questions about ownership.

## 7. Discussion

ChatGPT, developed by OpenAI, is an AI language model with substantial potential in healthcare. Its human-like responses can assist clinicians, enhance patient care, and improve communication. A primary advantage is 24/7 patient support with minimal human involvement. ChatGPT aids patients in scheduling appointments, answering inquiries, and sending medication reminders, enhancing engagement and satisfaction. Additionally, it serves as a clinical decision support tool. By analyzing symptoms and medical histories, it provides evidence-based recommendations, streamlining decision-making in cases like health related concerns.

ChatGPT also enhances medical education by answering students' questions and generating quizzes. Its interactive approach fosters deeper understanding of complex medical concepts. Moreover, it automates administrative tasks such as generating clinical reports and discharge summaries, allowing healthcare providers to focus more on patient care, thereby improving efficiency. As ChatGPT is adopted in healthcare, it is expected to improve communication between patients and providers, enhance clinical workflows, and support high-quality care. It clarifies treatment plans and boosts health literacy by delivering tailored educational materials.

## 8. Limitations

Despite its potential, the integration of ChatGPT in healthcare presents significant limitations. A primary concern is the risk of misinformation; outdated or incorrect information could lead to ill-informed health decisions. While ChatGPT automates customer support, this could result in job losses for administrative staff. Ethical implications, including patient privacy and informed consent, are critical; mishandling sensitive data could undermine trust in the healthcare system.

Additionally, it cannot provide personalized medical advice, lacking the capacity to tailor responses based on individual patient histories. Its responses may lack nuance, particularly in mental health contexts, where empathy is essential. While capable of assisting with information, it struggles with fact based information, requiring excessive detail for effective results. Furthermore, it cannot fully replace the critical thinking of human professionals, especially in rapidly evolving medical fields, as it may miss crucial developments.

## 9. Future Directions

ChatGPT is set to significantly impact healthcare, particularly in patient engagement, diagnostics, and medical education. It can enhance patient interactions by facilitating real-time communication, providing information on treatment options, and answering medication queries, thereby improving satisfaction and adherence to treatment plans.

As real-world data is integrated into healthcare systems, it can analyze this information to support clinical decision-making, helping physicians identify symptom patterns and suggest appropriate diagnostic tests for personalized treatment plans.

In medical education, it can act as a virtual tutor, offering immediate feedback on clinical scenarios and generating practice questions for exam preparation. This tailored approach enhances learning outcomes and better prepares future healthcare professionals.

The future of ChatGPT in healthcare is promising, with anticipated improvements in accuracy and speed as OpenAI refines its models. Integrating ChatGPT with electronic health records (EHR) could enable it to analyze patient data and facilitate informed clinical decisions.

Additionally, it can streamline recruitment in healthcare organizations by automating candidate screening and interview scheduling, thereby reducing administrative burdens and accelerating on-boarding. In medical tourism, it could assist patients in booking travel arrangements, providing real-time updates on flight statuses, and offering information on local healthcare services.

In eLearning, it is set to revolutionize medical education by generating tailored learning materials and assessments, thereby enhancing training efficiency. As the technology evolves, collaborations between healthcare institutions and AI solution providers will be crucial to maximizing the potential of generative AI, ultimately improving patient outcomes and operational efficiency across the sector.

## 10. Conclusion

ChatGPT has the potential to significantly transform the healthcare industry by enhancing communication between patients and providers through its ability to generate human-like text responses. It can streamline administrative tasks by producing organized medical reports, thereby improving clinical efficiency. Additionally, it can analyze medical research and adverse event reports to identify critical trends that contribute to patient safety and quality of care.

However, it is important to recognize that ChatGPT cannot replace the expertise of healthcare professionals. Limitations such as ethical considerations, accountability, and the need for accurate data interpretation must be carefully addressed. Healthcare providers should remain vigilant in verifying the information generated by AI tools to ensure alignment with established medical evidence. As healthcare continues to evolve, the thoughtful integration of technologies like ChatGPT can enhance patient engagement and operational efficiency while prioritizing patient safety and ethical standards.

## References

1.  Bajwa, J.; Munir, U.; Nori, A.; Williams, B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal* **2021**, *8*, e188–e194. doi:10.7861/fhj.2021-0095.
2.  Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal* **2019**, *6*, 94–98. doi:10.7861/futurehosp.6-2-94.
3.  Tan, T.; Thirunavukarasu, A.; Campbell, J.; Keane, P.; Pasquale, L.; Abramoff, M.; Kalpathy-Cramer, J.; Lum, F.; Kim, J.; Baxter, S.; Ting, D. Generative Artificial Intelligence Through ChatGPT and Other Large Language Models in Ophthalmology: Clinical Applications and Challenges. *Ophthalmology Science* **2023**, *3*, 100394. doi:10.1016/j.xops.2023.100394.
4.  Liévin, V.; Hother, C.E.; Motzfeldt, A.G.; Winther, O. Can large language models reason about medical questions?, 2023, [arXiv:cs.CL/2207.08143].
5.  Haleem, A.; Javaid, M.; Singh, R.P. An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* **2022**, *2*, 100089. doi:https://doi.org/10.1016/j.tbench.2023.100089.

6.  Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; Chou, R.; Glanville, J.; Grimshaw, J.M.; Hróbjartsson, A.; Lalu, M.M.; Li, T.; Loder, E.W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L.A.; Stewart, L.A.; Thomas, J.; Tricco, A.C.; Welch, V.A.; Whiting, P.; McKenzie, J.E. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* **2021**, *372*, n160. doi:10.1136/bmj.n160.

7.  Khurana, D.; Koli, A.; Khatter, K.; others. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications* **2023**, *82*, 3713–3744. doi:10.1007/s11042-022-13428-4.

8.  Khan, W.; Daud, A.; Khan, K.; Muhammad, S.; Haq, R. Exploring the Frontiers of Deep Learning and Natural Language Processing: A Comprehensive Overview of Key Challenges and Emerging Trends. *Natural Language Processing Journal* **2023**, *4*, 100026. doi:10.1016/j.nlp.2023.100026.

9.  Tyagi, N.; Bhushan, B. Demystifying the Role of Natural Language Processing (NLP) in Smart City Applications: Background, Motivation, Recent Advances, and Future Research Directions. *Wireless Personal Communications* **2023**, *130*, 857–908. Epub 2023 Mar 16, doi:10.1007/s11277-023-10312-8.

10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need, 2023, [arXiv:cs.CL/1706.03762].

11. Mikolov, T.; Karafiát, M.; Burget, L.; Cernockỳ, J.; Khudanpur, S. Recurrent neural network based language model. Interspeech. Makuhari, 2010, Vol. 2, pp. 1045–1048.

12. Graves, A.; Graves, A. Long short-term memory. *Supervised sequence labelling with recurrent neural networks* **2012**, pp. 37–45.

13. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y.; others. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* **2024**, *15*, 1–45.

14. Alaparthi, S.; Mishra, M. Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. *arXiv preprint arXiv:2007.01127* **2020**.

15. Liu, X.; Zheng, Y.; Du, Z.; Ding, M.; Qian, Y.; Yang, Z.; Tang, J. GPT understands, too. *AI Open* **2023**.

16. Gorenstein, L.; Konen, E.; Green, M.; Klang, E. Bidirectional Encoder Representations from Transformers in Radiology: A Systematic Review of Natural Language Processing Applications. *Journal of the American College of Radiology* **2024**, *21*, 914–941. doi:https://doi.org/10.1016/j.jacr.2024.01.012.

17. Devlin, J. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* **2018**.

18. Shah Jahan, M.; Khan, H.U.; Akbar, S.; Umar Farooq, M.; Gul, S.; Amjad, A. Bidirectional language modeling: a systematic literature review. *Scientific Programming* **2021**, *2021*, 6641832.

19. Gu, Y.; Tinn, R.; Cheng, H.; Lucas, M.; Usuyama, N.; Liu, X.; Naumann, T.; Gao, J.; Poon, H. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)* **2021**, *3*, 1–23.

20. Huang, K.; Altosaar, J.; Ranganath, R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission, 2020, [arXiv:cs.CL/1904.05342].

21. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2020**, *36*, 1234–1240. doi:10.1093/bioinformatics/btz682.

22. National Library of Medicine (US). UMLS Knowledge Sources. Available from: http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html, 2024. Release 2024AA. [dataset on the Internet] [cited 2024 Jul 15].

23. Nazi, Z.A.; Peng, W. Large language models in healthcare and medical domain: A review. *Informatics* **2024**, *11*, 57.

24. Fan, L.; Li, L.; Ma, Z.; Lee, S.; Yu, H.; Hemphill, L. A bibliometric review of large language models research from 2017 to 2023. *arXiv preprint arXiv:2304.02020* **2023**.

25. Sanderson, K. GPT-4 is here: what scientists think. *Nature* **2023**, *615*, 773.

26. Radford, A. Improving language understanding by generative pre-training, 2018.

27. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language Models are Unsupervised Multitask Learners. *OpenAI blog* **2019**, *1*, 9.

28. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shi, K.; Sastry, G.; Askell, A.; Sutskever, I. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems* **2020**, *33*, 1877–1901.

29. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D.M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; Amodei, D. Language Models are Few-Shot Learners, 2020, [arXiv:cs.CL/2005.14165].

30. Xu, T.; Weng, H.; Liu, F.; Yang, L.; Luo, Y.; Ding, Z.; Wang, Q. Current Status of ChatGPT Use in Medical Education: Potentials, Challenges, and Strategies. *Journal of Medical Internet Research* **2024**, *26*, e57896. doi:10.2196/57896.

31. Samala, A.D.; Rawas, S. Generative AI as Virtual Healthcare Assistant for Enhancing Patient Care Quality. *International Journal of Online & Biomedical Engineering* **2024**, *20*.

32. Ferdush, J.; Begum, M.; Hossain, S.T. ChatGPT and clinical decision support: scope, application, and limitations. *Annals of Biomedical Engineering* **2024**, *52*, 1119–1124.

33. Iftikhar, L.; Iftikhar, M.F.; Hanif, M.I. Docgpt: Impact of chatgpt-3 on health services as a virtual doctor. *EC Paediatrics* **2023**, *12*, 45–55.

34. Zheng, Y.; Wang, L.; Feng, B.; Zhao, A.; Wu, Y. Innovating healthcare: the role of ChatGPT in streamlining hospital workflow in the future. *Annals of Biomedical Engineering* **2024**, *52*, 750–753.

35. Awal, S.; Awal, S. ChatGPT and the healthcare industry: a comprehensive analysis of its impact on medical writing. *J Public Health (Berl.)* **2023**. doi:10.1007/s10389-023-02170-2.

36. Temsah, O.; Khan, S.A.; Chaiah, Y.; Senjab, A.; Alhasan, K.; Jamal, A.; Aljamaan, F.; Malki, K.H.; Halwani, R.; Al-Tawfiq, J.A.; others. Overview of early ChatGPT's presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. *Cureus* **2023**, *15*.

37. Parray, A.A.; Inam, Z.M.; Ramonfaur, D.; Haider, S.S.; Mistry, S.K.; Pandya, A.K. ChatGPT and global public health: Applications, challenges, ethical considerations and mitigation strategies. *Global Transitions* **2023**, *5*, 50–54. doi:https://doi.org/10.1016/j.glt.2023.05.001.

38. Abd Karim, R.; Cakir, G.K. Investigating ChatGPT Usability in Promoting Smart Health Awareness. In *Industry 5.0 for Smart Healthcare Technologies*; CRC Press, 2024; pp. 227–237.

39. Baldwin, A.J. An artificial intelligence language model improves readability of burns first aid information. *Burns* **2024**, *50*, 1122–1127.

40. Neha, F. Kidney Localization and Stone Segmentation from a CT Scan Image. 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, 2023, pp. 1–6.

41. Neha, F.; Bansal, A.K. Multi-Layer Feature Fusion with Cross-Channel Attention-Based U-Net for Kidney Tumor Segmentation.

42. Choi, J.; Kim, J.W.; Lee, Y.S.; Tae, J.H.; Choi, S.Y.; Chang, I.H.; Kim, J.H. Availability of ChatGPT to provide medical information for patients with kidney cancer. *Scientific Reports* **2024**, *14*, 1542.

43. Miao, J.; Thongprayoon, C.; Craici, I.M.; Cheungpasitporn, W. How to improve ChatGPT performance for nephrologists: a technique guide. *Journal of Nephrology* **2024**, pp. 1–7.

44. Janus, N. A Comparative Analysis of Chatgpt Vs Expert in Managing Anticancer Drug in Patients Renal Insufficiency. *Blood* **2023**, *142*, 7186.

45. Łaszkiewicz, J.; Krajewski, W.; Tomczak, W.; Chorbińska, J.; Nowak, Ł.; Chełmoński, A.; Krajewski, P.; Sójka, A.; Małkiewicz, B.; Szydełko, T. Performance of ChatGPT in providing patient information about upper tract urothelial carcinoma. *Contemporary Oncology/Współczesna Onkologia*, *28*.

46. Javid, M.; Bhandari, M.; Parameshwari, P.; Reddiboina, M.; Prasad, S. Evaluation of ChatGPT for patient counseling in kidney stone clinic: A prospective study. *Journal of endourology* **2024**, *38*, 377–383.

47. Miao, J.; Thongprayoon, C.; Suppadungsuk, S.; Garcia Valencia, O.A.; Qureshi, F.; Cheungpasitporn, W. Innovating personalized nephrology care: exploring the potential utilization of ChatGPT. *Journal of Personalized Medicine* **2023**, *13*, 1681.

48. Qarajeh, A.; Tangpanithandee, S.; Thongprayoon, C.; Suppadungsuk, S.; Krisanapan, P.; Aiumtrakul, N.; Garcia Valencia, O.A.; Miao, J.; Qureshi, F.; Cheungpasitporn, W. AI-Powered Renal Diet Support: Performance of ChatGPT, Bard AI, and Bing Chat. *Clinics and Practice* **2023**, *13*, 1160–1172.

49. German, R.Z.; Palmer, J.B. Anatomy and development of oral cavity and pharynx. *GI Motility online* **2006**.

50. Lechien, J.R.; Georgescu, B.M.; Hans, S.; Chiesa-Estomba, C.M. ChatGPT performance in laryngology and head and neck surgery: a clinical case-series. *European Archives of Oto-Rhino-Laryngology* **2024**, *281*, 319–333.

51. Aaronson, P.I.; Ward, J.P.; Connolly, M.J. *The cardiovascular system at a glance*; John Wiley & Sons, 2020.

52. Lautrup, A.D.; Hyrup, T.; Schneider-Kamp, A.; Dahl, M.; Lindholt, J.S.; Schneider-Kamp, P. Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice. *Open heart* **2023**, *10*, e002455.

53. Anaya, F.; Prasad, R.; Bashour, M.; Yaghmour, R.; Alameh, A.; Blakumaran, K. Evaluating ChatGPT platform in delivering heart failure educational material: A comparison with the leading national cardiology institutes. *Current Problems in Cardiology* **2024**, p. 102797.

54. King, R.C.; Samaan, J.S.; Yeo, Y.H.; Mody, B.; Lombardo, D.M.; Ghashghaei, R. Appropriateness of ChatGPT in answering heart failure related questions. *Heart, Lung and Circulation* **2024**.

55. BULBOACĂ, A.I.; BORLEA, B.; BULBOACĂ, A.E.; STĂNESCU, I.C.; BOLBOACĂ, S.D. Exploring ChatGPT's Efficacy in Pathophysiological Analysis: A Comparative Study of Ischemic Heart Disease and Anaphylactic Shock Cases. *Applied Medical Informatics* **2024**, *46*, 16–28.

56. Chlorogiannis, D.D.; Apostolos, A.; Chlorogiannis, A.; Palaiodimos, L.; Giannakoulas, G.; Pargaonkar, S.; Xesfingi, S.; Kokkinidis, D.G. The role of ChatGPT in the advancement of diagnosis, management, and prognosis of cardiovascular and cerebrovascular disease. Healthcare. MDPI, 2023, Vol. 11, p. 2906.

57. Ayub, M.; Mallamaci, A. An Introduction: Overview of Nervous System and Brain Disorders. *The Role of Natural Antioxidants in Brain Disorders* **2023**, pp. 1–24.

58. Kozel, G.; Gurses, M.E.; Gecici, N.N.; Gökalp, E.; Bahadir, S.; Merenzon, M.A.; Shah, A.H.; Komotar, R.J.; Ivan, M.E. Chat-GPT on brain tumors: an examination of Artificial Intelligence/Machine Learning's ability to provide diagnoses and treatment plans for example neuro-oncology cases. *Clinical Neurology and Neurosurgery* **2024**, *239*, 108238.

59. Adesso, G. Towards the ultimate brain: Exploring scientific discovery with ChatGPT AI. *AI Magazine* **2023**, *44*, 328–342.

60. Fei, X.; Tang, Y.; Zhang, J.; Zhou, Z.; Yamamoto, I.; Zhang, Y. Evaluating cognitive performance: Traditional methods vs. ChatGPT. *Digital Health* **2024**, *10*, 20552076241264639.

61. Al-Suhaimi, E.A.; Khan, F.A. Thyroid glands: Physiology and structure. In *Emerging concepts in endocrine structure and functions*; Springer, 2022; pp. 133–160.

62. Köroğlu, E.Y.; Fakı, S.; Beştepe, N.; Tam, A.A.; Seyrek, N.Ç.; Topaloglu, O.; Ersoy, R.; Cakir, B. A novel approach: Evaluating ChatGPT's utility for the management of thyroid nodules. *Cureus* **2023**, *15*.

63. Sievert, M.; Conrad, O.; Mueller, S.K.; Rupp, R.; Balk, M.; Richter, D.; Mantsopoulos, K.; Iro, H.; Koch, M. Risk stratification of thyroid nodules: Assessing the suitability of ChatGPT for text-based analysis. *American Journal of Otolaryngology* **2024**, *45*, 104144.

64. Stevenson, E.; Walsh, C.; Hibberd, L. Can artificial intelligence replace biochemists? A study comparing interpretation of thyroid function test results by ChatGPT and Google Bard to practising biochemists. *Annals of Clinical Biochemistry* **2024**, *61*, 143–149.

65. Helvaci, B.C.; Hepsen, S.; Candemir, B.; Boz, O.; Durantas, H.; Houssein, M.; Cakal, E. Assessing the accuracy and reliability of ChatGPT's medical responses about thyroid cancer. *International Journal of Medical Informatics* **2024**, *191*, 105593.

66. Cazzato, G.; Capuzzolo, M.; Parente, P.; Arezzo, F.; Loizzi, V.; Macorano, E.; Marzullo, A.; Cormio, G.; Ingravallo, G. Chat GPT in diagnostic human pathology: will it be useful to pathologists? A preliminary review with 'query session'and future perspectives. *AI* **2023**, *4*, 1010–1022.

67. Alamri, Z.Z. The role of liver in metabolism: an updated review with physiological emphasis. *Int J Basic Clin Pharmacol* **2018**, *7*, 2271–2276.

68. Yeo, Y.H.; Samaan, J.S.; Ng, W.H.; Ting, P.S.; Trivedi, H.; Vipani, A.; Ayoub, W.; Yang, J.D.; Liran, O.; Spiegel, B.; others. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clinical and molecular hepatology* **2023**, *29*, 721.

69. Yeo, Y.H.; Samaan, J.S.; Ng, W.H.; Ma, X.; Ting, P.S.; Kwak, M.S.; Panduro, A.; Lizaola-Mayo, B.; Trivedi, H.; Vipani, A.; others. GPT-4 outperforms ChatGPT in answering non-English questions related to cirrhosis. *medRxiv* **2023**, pp. 2023–05.

70. Shahsavari, D.; Parkman, H.P. Normal gastrointestinal tract physiology. In *Nutrition, Weight, and Digestive Health: The Clinician's Desk Reference*; Springer, 2022; pp. 3–28.

71. Cankurtaran, R.E.; Polat, Y.H.; Aydemir, N.G.; Umay, E.; Yurekli, O.T. Reliability and usefulness of ChatGPT for inflammatory bowel diseases: an analysis for patients and healthcare professionals. *Cureus* **2023**, *15*.

72. Ma, Y. The potential application of ChatGPT in gastrointestinal pathology. *Gastroenterology & Endoscopy* **2023**, *1*, 130–131.

73. Liu, X.; Wang, Y.; Huang, Z.; Xu, B.; Zeng, Y.; Chen, X.; Wang, Z.; Yang, E.; Lei, X.; Huang, Y.; others. The Application of ChatGPT in Responding to Questions Related to the Boston Bowel Preparation Scale. *arXiv preprint arXiv:2402.08492* **2024**.

74. Chang, E.B.; Leung, P.S. Pancreatic physiology. In *The gastrointestinal system: gastrointestinal, nutritional and hepatobiliary physiology*; Springer, 2014; pp. 87–105.

75. Du, R.C.; Liu, X.; Lai, Y.K.; Hu, Y.X.; Deng, H.; Zhou, H.Q.; Lu, N.H.; Zhu, Y.; Hu, Y. Exploring the performance of ChatGPT on acute pancreatitis-related questions. *Journal of Translational Medicine* **2024**, *22*, 527.

76. Qiu, J.; Luo, L.; Zhou, Y. Accuracy of ChatGPT3. 5 in answering clinical questions on guidelines for severe acute pancreatitis. *BMC gastroenterology* **2024**, *24*, 260.

77. Lorenzo, A.J.; Bagli, D. Basic science of the urinary bladder. *Clinical Pediatric Urology. London: Informa Healthcare* **2007**.

78. Guo, A.A.; Razi, B.; Kim, P.; Canagasingham, A.; Vass, J.; Chalasani, V.; Rasiah, K.; Chung, A. The Role of Artificial Intelligence in Patient Education: A Bladder Cancer Consultation with ChatGPT. *Société Internationale d'Urologie Journal* **2024**, *5*, 214–224.

79. Braga, A.V.N.M.; Nunes, N.C.; Santos, E.N.; Veiga, M.L.; Braga, A.A.N.M.; Abreu, G.E.d.; Bessa, J.d.; Braga, L.H.; Kirsch, A.J.; Barroso, U. Use of ChatGPT in Urology and its Relevance in Clinical Practice: Is it useful? *International braz j urol* **2024**, *50*, 192–198.

80. Ozgor, F.; Caglar, U.; Halis, A.; Cakir, H.; Aksu, U.C.; Ayranci, A.; Sarilar, O. Urological cancers and ChatGPT: assessing the quality of information and possible risks for patients. *Clinical Genitourinary Cancer* **2024**, *22*, 454–457.

81. Cakir, H.; Caglar, U.; Yildiz, O.; Meric, A.; Ayranci, A.; Ozgor, F. Evaluating the performance of ChatGPT in answering questions related to urolithiasis. *International Urology and Nephrology* **2024**, *56*, 17–21.

82. Sagir, S. Evaluating the accuracy of ChatGPT addressing urological questions: A pilot study. *Journal of Clinical Trials and Experimental Investigations* **2022**, *1*, 119–123.

83. Cakir, H.; Caglar, U.; Sekkeli, S.; Zerdali, E.; Sarilar, O.; Yildiz, O.; Ozgor, F. Evaluating ChatGPT ability to answer urinary tract Infection-Related questions. *Infectious Diseases Now* **2024**, *54*, 104884.

84. Szczesniewski, J.J.; Tellez Fouz, C.; Ramos Alba, A.; Diaz Goizueta, F.J.; García Tello, A.; Llanes González, L. ChatGPT and most frequent urological diseases: analysing the quality of information and potential risks for patients. *World Journal of Urology* **2023**, *41*, 3149–3153.

85. Mashinini, M. Pituitary gland and growth hormone. *Southern African Journal of Anaesthesia and Analgesia* **2020**, *26*, S109–112.

86. Sambangi, A.; Carreras, A.; Campbell, D.; Bray, D.; Evans, J.J. Evaluating Chatgpt for Patient Education Regarding Pituitary Adenoma Resection. *Journal of Neurological Surgery Part B: Skull Base* **2024**, *85*, P216.

87. Şenoymak, M.C.; Erbatur, N.H.; Şenoymak, İ.; Fırat, S.N. The Role of Artificial Intelligence in Endocrine Management: Assessing ChatGPT's Responses to Prolactinoma Queries. *Journal of Personalized Medicine* **2024**, *14*, 330.

88. Taşkaldıran, I.; Emir Önder, Ç.; Gökbulut, P.; Koç, G.; Kuşkonmaz, Ş.M. Evaluation of the accuracy and quality of ChatGPT-4 responses for hyperparathyroidism patients discussed at multidisciplinary endocrinology meetings. *Digital health* **2024**, *10*, 20552076241278692.

89. Hafez, B.; Hafez, E. Anatomy of female reproduction. *Reproduction in farm animals* **2000**, pp. 13–29.

90. Patel, J.M.; Hermann, C.E.; Growdon, W.B.; Aviki, E.; Stasenko, M. ChatGPT accurately performs genetic counseling for gynecologic cancers. *Gynecologic Oncology* **2024**, *183*, 115–119.

91. Peled, T.; Sela, H.Y.; Weiss, A.; Grisaru-Granovsky, S.; Agrawal, S.; Rottenstreich, M. Evaluating the validity of ChatGPT responses on common obstetric issues: Potential clinical applications and implications. *International Journal of Gynecology & Obstetrics* **2024**.

92. Psilopatis, I.; Bader, S.; Krueckel, A.; Kehl, S.; Beckmann, M.W.; Emons, J. Can Chat-GPT read and understand guidelines? An example using the S2k guideline intrauterine growth restriction of the German Society for Gynecology and Obstetrics. *Archives of Gynecology and Obstetrics* **2024**, pp. 1–13.

93. Winograd, D.; Alterman, C.; Appelbaum, H.; Baum, J. 51. Evaluation of ChatGPT Responses to Common Puberty Questions. *Journal of Pediatric and Adolescent Gynecology* **2024**, *37*, 261.

94. Edwards, S. The skin. *Essential Pathophysiology For Nursing And Healthcare Students* **2014**, p. 431.

95. Lantz, R. Toxic epidermal necrolysis in a critically ill African American woman: a Case Report Written with ChatGPT Assistance. *Cureus* **2023**, *15*.

96. Sanchez-Zapata, M.J.; Rios-Duarte, J.A.; Orduz-Robledo, M.; Motta, A. 53670 Evaluating ChatGPT answers to frequently asked questions from patients with inflammatory skin diseases in a physician-patient context. *Journal of the American Academy of Dermatology* **2024**, *91*, AB205.

97. Passby, L.; Jenko, N.; Wernham, A. Performance of ChatGPT on Specialty Certificate Examination in Dermatology multiple-choice questions. *Clinical and experimental dermatology* **2024**, *49*, 722–727.

98. Stoneham, S.; Livesey, A.; Cooper, H.; Mitchell, C. ChatGPT versus clinician: challenging the diagnostic capabilities of artificial intelligence in dermatology. *Clinical and Experimental Dermatology* **2024**, *49*, 707–710.

99. Mondal, H.; Mondal, S.; Podder, I. Using ChatGPT for writing articles for patients' education for dermatological diseases: a pilot study. *Indian Dermatology Online Journal* **2023**, *14*, 482–486.

100. Mat Lazim, N. Introduction to Head and Neck Surgery. In *Head and Neck Surgery: Surgical Landmark and Dissection Guide*; Springer, 2022; pp. 1–23.

101. Vaira, L.A.; Lechien, J.R.; Abbate, V.; Allevi, F.; Audino, G.; Beltramini, G.A.; Bergonzani, M.; Bolzoni, A.; Committeri, U.; Crimi, S.; others. Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: a multicenter collaborative analysis. *Otolaryngology–Head and Neck Surgery* **2024**, *170*, 1492–1503.

102. Nischwitz, D. *It's All in Your Mouth: Biological Dentistry and the Surprising Impact of Oral Health on Whole Body Wellness*; Chelsea Green Publishing, 2020.

103. Hassona, Y.; Alqaisi, D.; Alaa, A.H.; Georgakopoulou, E.A.; Malamos, D.; Alrashdan, M.S.; Sawair, F. How good is ChatGPT at answering patients' questions related to early detection of oral (mouth) cancer? *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* **2024**.

104. Babayiğit, O.; Eroglu, Z.T.; Sen, D.O.; Yarkac, F.U. Potential use of ChatGPT for Patient Information in Periodontology: a descriptive pilot study. *Cureus* **2023**, *15*.

105. Mago, J.; Sharma, M. The potential usefulness of ChatGPT in oral and maxillofacial radiology. *Cureus* **2023**, *15*.

106. Puladi, B.; Gsaxner, C.; Kleesiek, J.; Hölzle, F.; Röhrig, R.; Egger, J. The impact and opportunities of large language models like ChatGPT in oral and maxillofacial surgery: a narrative review. *International journal of oral and maxillofacial surgery* **2024**, *53*, 78–88.

107. Fahey, J. Optimising lung health. *Journal of the Australian Traditional-Medicine Society* **2020**, *26*, 142–147.

108. Fink, M.A.; Bischoff, A.; Fink, C.A.; Moll, M.; Kroschke, J.; Dulz, L.; Heußel, C.P.; Kauczor, H.U.; Weber, T.F. Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. *Radiology* **2023**, *308*, e231362.

109. Rahsepar, A.A.; Tavakoli, N.; Kim, G.H.J.; Hassani, C.; Abtin, F.; Bedayat, A. How AI responds to common lung cancer questions: ChatGPT versus Google Bard. *Radiology* **2023**, *307*, e230922.

110. Nakamura, Y.; Kikuchi, T.; Yamagishi, Y.; Hanaoka, S.; Nakao, T.; Miki, S.; Yoshikawa, T.; Abe, O. ChatGPT for automating lung cancer staging: feasibility study on open radiology report dataset. *medRxiv* **2023**, pp. 2023–12.

111. Lee, J.E.; Park, K.S.; Kim, Y.H.; Song, H.C.; Park, B.; Jeong, Y.J. Lung Cancer Staging Using Chest CT and FDG PET/CT Free-Text Reports: Comparison Among Three ChatGPT Large-Language Models and Six Human Readers of Varying Experience. *American Journal of Roentgenology* **2024**.

112. Schulte, B. Capacity of ChatGPT to identify guideline-based treatments for advanced solid tumors. *Cureus* **2023**, *15*.

113. White, T.D.; Folkens, P.A. *The human bone manual*; Elsevier, 2005.

114. Ghanem, D.; Shu, H.; Bergstein, V.; Marrache, M.; Love, A.; Hughes, A.; Sotsky, R.; Shafiq, B. Educating patients on osteoporosis and bone health: Can "ChatGPT" provide high-quality content? *European Journal of Orthopaedic Surgery & Traumatology* **2024**, pp. 1–9.

115. Son, H.J.; Kim, S.J.; Pak, S.; Lee, S.H. ChatGPT-assisted deep learning for diagnosing bone metastasis in bone scans: Bridging the AI Gap for Clinicians. *Heliyon* **2023**, *9*.

116. Cinar, C. Analyzing the performance of ChatGPT about osteoporosis. *Cureus* **2023**, *15*.

117. 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. *Journal of Bone Oncology* **2024**, *44*, 100525.

118. Kell, R.T.; Bell, G.; Quinney, A. Musculoskeletal fitness, health outcomes and quality of life. *Sports Medicine* **2001**, *31*, 863–873.

119. Sawamura, S.; Kohiyama, K.; Takenaka, T.; Sera, T.; Inoue, T.; Nagai, T. Performance of ChatGPT 4.0 on Japan's National Physical Therapist Examination: A Comprehensive Analysis of Text and Visual Question Handling. *Cureus* **2024**, *16*.

120. Agarwal, M.; Sharma, P.; Goswami, A. Analysing the applicability of ChatGPT, Bard, and Bing to generate reasoning-based multiple-choice questions in medical physiology. *Cureus* **2023**, *15*.

121. Saluja, S.; Tigga, S.R. Capabilities and Limitations of ChatGPT in Anatomy Education: An Interaction With ChatGPT. *Cureus* **2024**, *16*.

122. Kaarre, J.; Feldt, R.; Keeling, L.E.; Dadoo, S.; Zsidai, B.; Hughes, J.D.; Samuelsson, K.; Musahl, V. Exploring the potential of ChatGPT as a supplementary tool for providing orthopaedic information. *Knee Surgery, Sports Traumatology, Arthroscopy* **2023**, *31*, 5190–5198.

123. Meng, D.; He, S.; Wei, M.; Lv, Z.; Guo, H.; Yang, G.; Wang, Z. Enhanced predicting genu valgum through integrated feature extraction: Utilizing ChatGPT with body landmarks. *Biomedical Signal Processing and Control* **2024**, *97*, 106676.

124. Mantzou, N.; Ediaroglou, V.; Drakonaki, E.; Syggelos, S.A.; Karageorgos, F.F.; Totlis, T. ChatGPT efficacy for answering musculoskeletal anatomy questions: a study evaluating quality and consistency between raters and timepoints. *Surgical and Radiologic Anatomy* **2024**, pp. 1–6.

125. Li, J.; You, M.; Chen, X.; Li, P.; Deng, Q.; Wang, K.; Wang, L.; Xu, Y.; Liu, D.; Ye, L.; others. ChatGPT-4 and Wearable Device Assisted Intelligent Exercise Therapy for Co-existing Sarcopenia and Osteoarthritis (GAISO): A feasibility study and design for a randomized controlled PROBE non-inferiority trial **2023**.

126. Walker, H.L.; Ghani, S.; Kuemmerli, C.; Nebiker, C.A.; Müller, B.P.; Raptis, D.A.; Staubli, S.M. 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.

127. Liaw, W.; Chavez, S.; Pham, C.; Tehami, S.; Govender, R. The Hazards of Using ChatGPT: A Call to Action for Medical Education Researchers. *PRiMER* **2023**, *7*, 27. doi:10.22454/PRiMER.2023.295710.

128. Wang, C.; Liu, S.; Yang, H.; Guo, J.; Wu, Y.; Liu, J. Ethical Considerations of Using ChatGPT in Health Care. *J Med Internet Res* **2023**, *25*, e48009. doi:10.2196/48009.

129. Si, Y.; Yang, Y.; Wang, X.; Zu, J.; Chen, X.; Fan, X.; An, R.; Gong, S. Quality and Accountability of ChatGPT in Health Care in Low- and Middle-Income Countries: Simulated Patient Study. *J Med Internet Res* **2024**, *26*, e56121. doi:10.2196/56121.

130. Baumgartner, C.; Baumgartner, D. A regulatory challenge for natural language processing (NLP)-based tools such as ChatGPT to be legally used for healthcare decisions. Where are we now? *Clin Transl Med* **2023**, *13*, e1362. doi:10.1002/ctm2.1362.

131. Goh, E.; Bunning, B.; Khoong, E.; Gallo, R.; Milstein, A.; Centola, D.; Chen, J.H. ChatGPT Influence on Medical Decision-Making, Bias, and Equity: A Randomized Study of Clinicians Evaluating Clinical Vignettes. *medRxiv [Preprint]* **2023**, p. 2023.11.24.23298844. doi:10.1101/2023.11.24.23298844.

132. Palaniappan, K.; Lin, E.Y.T.; Vogel, S. Global Regulatory Frameworks for the Use of Artificial Intelligence (AI) in the Healthcare Services Sector. *Healthcare (Basel)* **2024**, *12*, 562. doi:10.3390/healthcare12050562.

133. Shieh, A.; Tran, B.; He, G.; Kumar, M.; Freed, J.A.; Majety, P. Assessing ChatGPT 4.0's Test Performance and Clinical Diagnostic Accuracy on USMLE STEP 2 CK and Clinical Case Reports. *Scientific Reports* **2024**, *14*, 1–8. Accessed October 3, 2024, doi:10.1038/s41598-024-58760-x.

134. Liu, J.; Wang, C.; Liu, S. Utility of ChatGPT in Clinical Practice. *J Med Internet Res* **2023**, *25*, e48568. doi:10.2196/48568.

135. Teixeira da Silva, J.A. Can ChatGPT rescue or assist with language barriers in healthcare communication? *Patient Education and Counseling* **2023**, *115*, 107940. doi:10.1016/j.pec.2023.107940.

136. Liu, Z.; Zhang, L.; Wu, Z.; Yu, X.; Cao, C.; Dai, H.; Liu, N.; Liu, J.; Liu, W.; Li, Q.; Shen, D.; Li, X.; Zhu, D.; Liu, T. Surviving ChatGPT in healthcare. *Front Radiol* **2024**, *3*, 1224682. doi:10.3389/fradi.2023.1224682.

137. Sedaghat, S. Plagiarism and Wrong Content as Potential Challenges of Using Chatbots Like ChatGPT in Medical Research. *J Acad Ethics* **2024**. doi:10.1007/s10805-024-09533-8.