AI

MDPI

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, USA; dbhati@kent.edu
[2] Rutgers Business School, Rutgers University, Newark, NJ 07102, USA; ds1640@scarletmail.rutgers.edu
[3] Department of Computer Science, West Chester University, West Chester, PA 19383, USA; mamiruzzaman@wcupa.edu
* Correspondence: neha@kent.edu

Abstract: ChatGPT, developed by OpenAI, is a large language model (LLM) that leverages artificial intelligence (AI) and deep learning (DL) to generate human-like responses. This paper provides a broad, systematic review of ChatGPT's applications in healthcare, particularly in enhancing patient engagement through medical history collection, symptom assessment, and decision support for improved diagnostic accuracy. It assesses ChatGPT's potential across multiple organ systems and specialties, highlighting its value in clinical, educational, and administrative contexts. This analysis reveals both the benefits and limitations of ChatGPT, including health literacy promotion and support for clinical decision-making, alongside challenges such as the risk of inaccuracies, ethical considerations around informed consent, and regulatory hurdles. A quantified summary of key findings shows ChatGPT's promise in various applications while underscoring the risks associated with its integration in medical practice. Through this comprehensive approach, this review aims to provide healthcare professionals, researchers, and policymakers with a balanced view of ChatGPT's potential and limitations, emphasizing the need for ongoing updates to keep pace with evolving medical knowledge.

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

check for updates

Citation: Neha, F.; Bhati, D.; Shukla, D.K.; Amiruzzaman, M. ChatGPT: Transforming Healthcare with AI. *AI* **2024**, *5*, 2618–2650. https://doi.org/10.3390/ai5040126

Academic Editor: Arslan Munir

Received: 9 October 2024
Revised: 9 November 2024
Accepted: 21 November 2024
Published: 2 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial Intelligence (AI) has transformed various sectors, including finance, education, transportation, manufacturing, retail, agriculture, entertainment, telecommunications, and cybersecurity. Among these, healthcare is 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 using machine learning (ML), natural language processing (NLP), and deep learning (DL) [2]. Applications range from early disease detection to automated patient management, significantly improving outcomes and reducing costs. However, despite these advancements, the healthcare sector continues to face persistent challenges such as limited accessibility to health information, shortages of healthcare personnel, and the need for enhanced patient engagement.

A key driver of AI's impact in healthcare is the emergence of Large Language Models (LLMs), which are essential for 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. Integrating LLMs into healthcare enables professionals to efficiently process vast amounts of medical literature, make informed decisions, and improve communication with patients, building upon AI's foundational contributions to the field. Moreover, LLMs can assist in mitigating healthcare personnel shortages by automating routine tasks and providing decision support, allowing healthcare professionals to focus on more critical aspects of patient care.

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. For instance, ChatGPT can provide patients with accessible health information, answer common medical queries, and offer preliminary guidance, enhancing patient engagement and empowerment. In clinical settings, it can aid clinicians by summarizing patient records, suggesting possible diagnoses, and offering evidence-based recommendations, thus improving clinical decision-making.

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. In educational contexts, ChatGPT can serve as a resource for medical students and professionals seeking to stay updated with the latest research, supporting ongoing education and knowledge dissemination. In administrative applications, it can automate scheduling, handle billing inquiries, and manage patient records, alleviating the burden on administrative staff and addressing personnel shortages.

Despite its promise, challenges remain, particularly concerning real-world applicability and the ethical implications of AI-driven decision-making in medicine. Ethical issues such as patient privacy, data security, algorithmic bias, and the potential for misdiagnosis must be carefully considered. There is also a need to establish clear regulatory frameworks and guidelines to ensure the safe and effective integration of ChatGPT into healthcare practices.

Based on an extensive review of existing research on ChatGPT's transformative role in healthcare, our study offers several key contributions, each providing quantified insights where applicable to strengthen the understanding of ChatGPT's impact:

1. **Comprehensive Background:** We deliver an in-depth overview of NLP, LLMs, GPT architecture, and ChatGPT, detailing their evolution and foundational technologies.
2. **Clinical Relevance:** Our analysis demonstrates how ChatGPT is reshaping patient care, administrative workflows, and medical research. Studies have shown that ChatGPT can reduce administrative workloads, freeing healthcare professionals to focus more on direct patient care.
3. **ChatGPT Applications Across Organ Systems:** We systematically review ChatGPT's effectiveness across various medical specialties, including its roles in diagnostics, treatment recommendations, patient education, and clinician-patient communication. In fields such as dermatology and nephrology, ChatGPT has shown promising preliminary accuracy in providing educational information and supporting patient self-care.
4. **Risk Analysis:** Our risk analysis addresses the reliability, accuracy, and ethical concerns surrounding ChatGPT's use in healthcare. We evaluate these challenges through current methodologies, emphasizing the importance of mitigating misinformation and ensuring patient safety.
5. **Future Directions:** Identifying research gaps, we propose a taxonomy categorizing the literature on ChatGPT applications, facilitating a structured understanding of its diverse healthcare roles and highlighting areas for future investigation.

The paper is structured as follows: Section 2 outlines the research methodology. Section 3 provides background on NLP, LLMs, GPT, and ChatGPT. Section 4 discusses ChatGPT's role in empowering patients. Section 5 reviews its applications across various organ systems. Section 6 discusses potential risks of using ChatGPT in healthcare. Section 7 offers a general discussion on ChatGPT in healthcare. Section 8 highlights the study's limitations, Section 9 suggests future research directions, and Section 10 concludes the paper.

2. Research Methodology

To explore the applications of ChatGPT in healthcare, we employed a systematic review methodology, following established preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocols to ensure a thorough and unbiased analysis as shown in Figure 1. The review involved the following key steps:

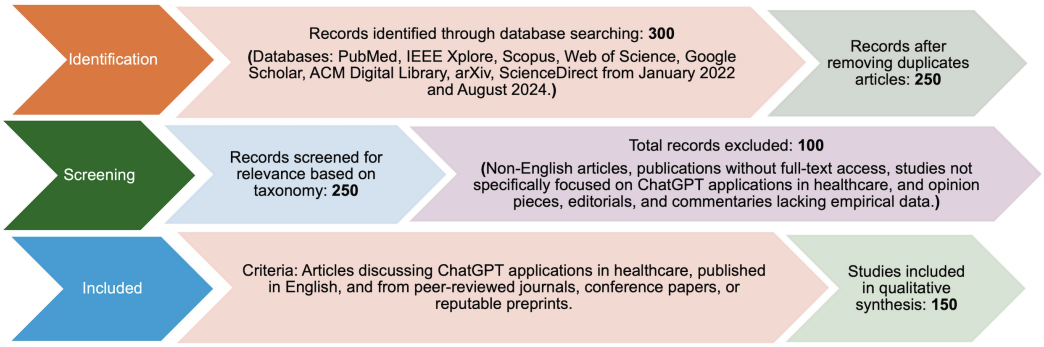


Figure 1. Research Methodology based on PRISMA protocols.

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 was performed for studies published from January 2022 to August 2024 to capture the most recent advancements in ChatGPT and its applications in healthcare. The search used targeted keywords and their combinations, such as "ChatGPT and healthcare", "ChatGPT and medical applications", "natural language processing in healthcare", "AI in patient management", and "ethical considerations of AI in medicine" [6].
2. **Inclusion and Exclusion Criteria:**
    - **Inclusion:**
        - Articles discussing the application of ChatGPT in various healthcare contexts.
        - Studies published in English between January 2022 and August 2024.
        - Peer-reviewed journal articles, conference papers, and reputable preprints that provide comprehensive reviews on the topic.
    - **Exclusion:**
        - Non-English articles.
        - Publications without full-text access.
        - Studies not specifically focused on ChatGPT and its applications in healthcare.
        - Opinion pieces, editorials, and commentaries lacking empirical data.
3. **Data Retrieval:** By 20 August 2024, we initially retrieved approximately 300 publications from the selected databases. After removing duplicates, 250 unique articles remained. Each article was assessed for eligibility based on the established inclusion and exclusion criteria, resulting in 150 publications included in the final review.
4. **Taxonomy Development:** We created a taxonomy to categorize the identified literature based on medical applications, including body parts, diagnosis, medical education, patient consultation, and specialties such as telemedicine, clinical decision support, and personalized medicine, as shown in Table 1.
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, limitations and future research directions.

Table 1. Taxonomy of ChatGPT Applications in Healthcare.

Category	Subcategories and Examples
Medical Applications	**Names of Body Parts**: Assisting in identifying and explaining conditions related to specific organs. **Diagnosis**: Preliminary symptom assessment and potential diagnosis suggestions.
Specialties	**Telemedicine**: Enhancing virtual consultations and remote patient monitoring. **Clinical Decision Support**: Providing evidence-based recommendations to clinicians. **Personalized Medicine**: Creating treatment plans based on individual patient data.

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 NLP, Transformer architecture, LLMs, enhancements through bidirectional language representation, and the development of ChatGPT.

3.1. Natural Language Processing (NLP)

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, transformed 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:

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

   In the formula, T stands for the transpose of the matrix K. In the context of the attention mechanism in transformers, taking the transpose of the keys matrix K is crucial for the matrix multiplication with the queries matrix Q. This operation aligns the dimensions appropriately, allowing each query to be compared against all keys, which is necessary for calculating the attention scores across the input sequence.
   The range of the attention scores in a transformer model is from 0 to 1. These scores are derived through the softmax function, which normalizes the computed scores so that they add up to 1 across each position in the input sequence for a given output position. A high attention score between a particular query (representing an output position) and a key (an input position) indicates that the input at that position is very relevant for producing the corresponding output. In other words, the transformer "focuses" more on that part of the input sequence.
   The model does assign scores to all parts of the input sequence, but not all parts are weighed equally. The final output at each position is a weighted sum of all values V, where the weights are the attention scores. This allows the transformer to dynamically focus on different parts of the input sequence depending on the context required by each output element. 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 [28]. 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.To fully lever-

age these benefits, it is crucial to explore how ChatGPT can be deployed across various healthcare settings and to identify potential infrastructure and resource challenges that may arise. 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 2 gives a snippet of the areas covered in this section.

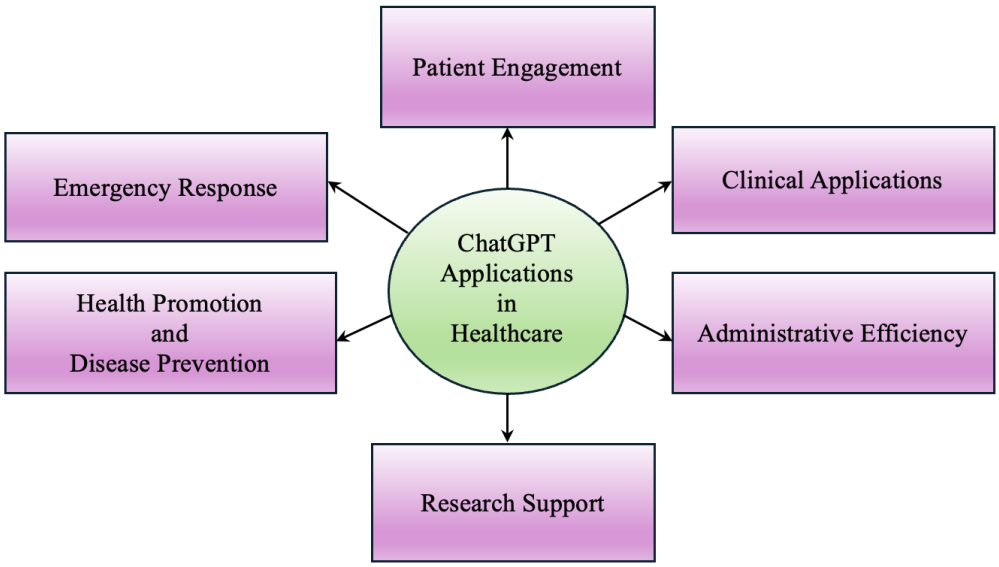


Figure 2. ChatGPT Applications in Healthcare.

4.1. Patient Engagement

ChatGPT plays a vital role in patient education and support by empowering individuals with knowledge about their health issues [29]. 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 [30]. Through personalized health insights, it empowers patients to make informed decisions, leading to better adherence to treatment plans. However, deploying ChatGPT for patient engagement requires addressing potential challenges such as ensuring patient access to the necessary technology. In rural or underserved areas, limited internet connectivity and a lack of smart devices can hinder effective use. Additionally, varying levels of digital literacy among patients may impact their ability to interact with AI tools. Healthcare providers must consider investing in infrastructure improvements, offering educational programs, and providing alternative access methods (e.g., kiosks in clinics) to ensure equitable service delivery.

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 [31]. It supports clinical monitoring by tracking patient health remotely, reminding them to check vital signs, and alerting healthcare providers about potential issues [32]. 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.

Nonetheless, integrating ChatGPT into clinical workflows presents infrastructure and resource challenges. Compatibility with existing Electronic Health Record (EHR) systems is essential but do require significant IT investment and technical expertise. Ensuring data privacy and compliance with regulations like Health Insurance Portability and Accountability Act (HIPAA) adds complexity, necessitating secure servers and encryption protocols. Smaller healthcare facilities struggle with these requirements due to limited budgets and technical staff. Overcoming these hurdles involves strategic planning, potential partnerships with technology providers, and seeking funding or grants dedicated to healthcare innovation.

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 [33]. 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 [34]. 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.

Implementing ChatGPT for administrative functions faces challenges such as the initial cost of technology adoption, including software licenses and hardware upgrades. There is also a need for staff training to effectively interact with and manage AI systems, which requires time and resources. Data security is another concern, as administrative systems handle sensitive patient and financial information. Healthcare organizations must allocate resources for cybersecurity measures and develop protocols to prevent unauthorized access or data breaches. Addressing these challenges is critical to maximizing the benefits of ChatGPT in administrative operations.

4.4. Research Support

In research and development, ChatGPT supports data analysis, hypothesis testing, and automates literature reviews for efficient findings [35]. 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.

Yet, deploying ChatGPT in research contexts encounters obstacles such as limited access to high-quality datasets due to privacy regulations or proprietary restrictions. Researchers need to invest in data anonymization techniques and secure data storage solutions to comply with ethical standards. Additionally, the computational resources required to process large datasets and run complex models can be significant, potentially necessitating investment in high-performance computing infrastructure or cloud services. Collaboration among institutions and seeking funding opportunities can help mitigate these resource challenges.

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 [36,37]. It enhances community health literacy and provides personalized advice on nutrition and exercise, empowering patients to adopt healthier habits. However, effective deployment in this

area requires overcoming infrastructure challenges such as ensuring widespread internet access and availability of devices capable of running ChatGPT applications. Language barriers and cultural differences must be considered to create content that is accessible and relevant to diverse populations. This involves translating materials into multiple languages and collaborating with community leaders to adapt messaging appropriately. Additionally, public health organizations need to allocate resources for outreach programs and campaigns to promote the use of ChatGPT tools among target populations.

4.6. Emergency Response

ChatGPT provides instant information during emergencies, guiding users on first aid procedures and facilitating communication between patients and emergency services [38]. It also offers resources and support for mental health issues, including coping strategies and exercises, while ensuring access to mental health professionals. Deploying ChatGPT in emergency scenarios requires robust and resilient infrastructure capable of operating under stress conditions. This includes reliable power supplies, backup systems, and secure communication networks that can function during disasters. Integrating ChatGPT with existing emergency management systems and protocols is essential but may require significant technical coordination and standardization efforts. Additionally, real-time data processing and dissemination demand high computational capacity and low-latency networks, which may not be available in all regions. Investment in infrastructure upgrades and collaboration with government agencies and emergency services is necessary to overcome these challenges and ensure effective utilization of ChatGPT in critical situations.

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 3. Each study highlights the main focus, methods, key results, and impact on clinical practice.

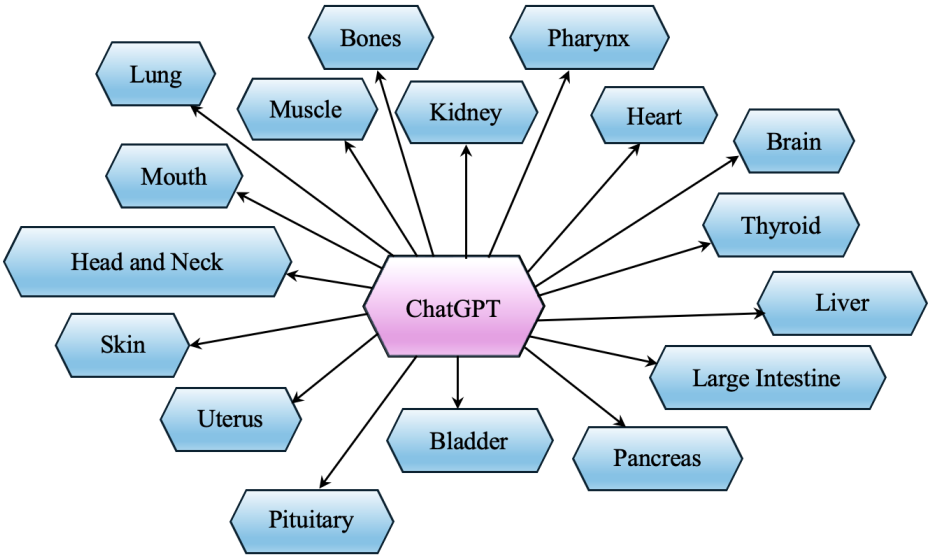


Figure 3. ChatGPT Applications Across Diverse Organ Systems in Healthcare.

5.1. Kidney

Kidney is a vital organ in the human body responsible for filtering blood to remove waste products, excess fluids, and toxins [39–41]. 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 2 summarizes the recent studies which highlight the integration of ChatGPT into the fields of kidney cancer and nephrology.

Table 2. 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 3 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 3. 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 Patho-physiology	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 4 summarizes the recent studies which highlight the integration of ChatGPT into the field of brain health.

Table 4. 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 5 provides a summary of these studies.

Table 5. 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 6 provides an overview of its integration into gastrointestinal pathology and large intestine care.

Table 6. 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 7 summarizes recent studies which have explored the integration of ChatGPT into the field of bladder health.

Table 7. 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.

Table 7. *Cont.*

Study	Focus Area	Methodology	Key Findings	Implications
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 8 summarizes the integration of ChatGPT into the field of uterus and gynecologic health.

Table 8. 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 9 summarizes the integration of ChatGPT into the field of dermatology and skin health in the recent works.

Table 9. 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 10 summarizes ChatGPT's integration into mouth health.

Table 10. 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 11 summarizes the integration of ChatGPT into the field of lung health.

Table 11. 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.

Table 11. *Cont.*

Study	Focus Area	Methodology	Key Findings	Implications
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 12 summarizes recent work in the field of integrating ChatGPT with bone health.

Table 12. 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 13 summarizes the integration of ChatGPT into the field of muscle health.

Table 13. 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. It is essential to thoroughly examine these ethical considerations, provide practical recommendations for risk mitigation, and understand the impact on underserved or vulnerable populations. This section outlines key concerns related to the use of ChatGPT in healthcare settings, emphasizing the importance of cautious implementation and oversight. Figure 4 points out these risks.

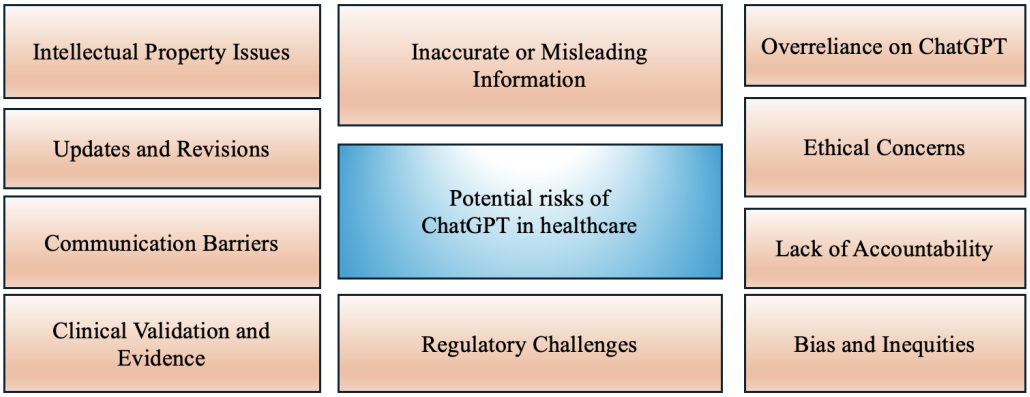


Figure 4. Potential Risks of ChatGPT in Healthcare.

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.

For example, a patient using ChatGPT for symptom checking might receive an incorrect assessment, causing unnecessary anxiety or a false sense of security. To mitigate this risk, it is crucial to implement safeguards such as integrating ChatGPT outputs with professional oversight. Healthcare organizations can establish protocols where AI-generated advice is reviewed by qualified professionals before being conveyed to patients. Additionally, providing clear disclaimers about the limitations of ChatGPT and encouraging users to consult healthcare providers for medical decisions can help reduce overreliance on AI-generated information.

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.

For instance, a clinician might overtrust ChatGPT's diagnostic suggestions without sufficient verification, potentially leading to incorrect treatment plans. To address this, training programs for healthcare professionals should emphasize the importance of maintaining critical evaluation skills when using AI tools. Establishing guidelines that define the appropriate use of ChatGPT in clinical practice can help balance the benefits of AI assistance with the necessity of human judgment.

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.

Moreover, there is a risk of violating patient autonomy if AI recommendations are presented without adequate context or options. Implementing strong privacy policies, obtaining explicit consent for data usage, and ensuring transparency about how patient information is handled are critical steps. Additionally, designing ChatGPT interactions to support patient autonomy by providing options and explaining reasoning can enhance ethical compliance.

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.

For example, if a patient follows ChatGPT's incorrect medical advice leading to adverse outcomes, determining who is legally responsible becomes challenging. To mitigate this issue, clear legal frameworks and regulations need to be established that define the accountability of AI developers, deployers, and users. Healthcare institutions should also implement policies that clarify the role of ChatGPT in care delivery and ensure that ultimate responsibility rests with qualified healthcare professionals.

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. For instance, language or cultural biases in the data might result in less accurate or inappropriate responses for non-English speakers or underrepresented groups. Underserved or vulnerable populations may face unique risks due to AI biases, exacerbating health disparities. These biases can manifest in various ways, such as misinterpretation of colloquial language, failure to recognize culturally specific symptoms, or providing recommendations that are not feasible for certain socioeconomic groups. To mitigate these issues, it is important to diversify training datasets to include a wide range of languages, cultures, and demographic information. Implementing bias detection and correction mechanisms during the development of ChatGPT can help identify and address disparities. Additionally, involving diverse stakeholders in the design and testing of AI tools can enhance cultural competency and equity.

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.

Without proper regulation, there is a risk of inconsistent standards, which can lead to variable quality of care and patient safety issues. Policymakers and regulatory bodies should collaborate with technologists and healthcare professionals to develop guidelines that address the specific challenges of AI in healthcare. Regular audits, certification processes, and compliance requirements can help ensure that ChatGPT applications meet established safety and ethical standards.

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.

For example, without proper validation, ChatGPT might suggest treatment plans that are not aligned with current medical guidelines. To mitigate this risk, extensive clinical trials and validation studies should be conducted before deploying ChatGPT in critical healthcare functions. Ongoing monitoring and post-market surveillance can help identify unforeseen issues and facilitate continuous improvement of the AI system.

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.

For instance, a patient seeking reassurance might receive a clinically accurate but emotionally insensitive response, impacting their psychological well-being. Incorporating natural language processing techniques that recognize and respond appropriately to emotional cues can enhance patient interactions. Complementing ChatGPT with human support, especially in mental health contexts, can ensure that patients receive both accurate information and the necessary emotional support.

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. For example, if ChatGPT continues to recommend a treatment that has been recently contraindicated due to new evidence, it could harm patients.

Establishing a systematic process for regular updates of the AI model with the latest medical guidelines and research findings is crucial. Collaboration with medical experts and institutions can facilitate timely incorporation of new information.

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.

For instance, if a clinician uses ChatGPT-generated text in a publication without proper citation, it may infringe on copyright laws. To address this, clear guidelines on the use of AI-generated content should be established, including attribution requirements and limitations on usage. Educating healthcare professionals about intellectual property laws related to AI-generated content can prevent inadvertent violations.

6.11. Addressing the Risks

To mitigate these risks, it is vital to implement practical strategies:

- Promote Awareness and Education: Healthcare providers and patients should be educated about the limitations of ChatGPT, emphasizing that it should complement, not replace, professional medical advice.
- Ensure Inclusive and Diverse Data: Efforts should be made to ensure that the training data for ChatGPT is diverse and representative, reducing the likelihood of bias and improving the quality of care for all demographic groups.
- Develop Robust Oversight Mechanisms: Regulatory bodies should establish oversight mechanisms to monitor the integration of AI in healthcare, ensuring compliance with ethical standards and accountability.

By proactively addressing these risks, we can better harness the potential of ChatGPT in healthcare while safeguarding patient welfare and promoting equitable access to care.

7. Discussion

ChatGPT, developed by OpenAI, demonstrates substantial potential in healthcare by offering a range of benefits alongside critical challenges that require careful consideration. With its ability to generate human-like responses, ChatGPT can assist clinicians, improve patient care, and facilitate communication in diverse healthcare settings. A primary advantage of ChatGPT is its capacity for 24/7 patient support, allowing immediate assistance in non-critical situations with minimal human intervention.

For instance, ChatGPT can help patients schedule appointments, answer inquiries, and send medication reminders—enhancing patient engagement and overall satisfaction. This functionality is particularly valuable in rural or underserved areas where access to healthcare professionals may be limited. In addition, ChatGPT can act as a clinical decision-support tool by analyzing symptoms and medical histories to provide evidence-based recommendations. For example, in a case of suspected respiratory infection, ChatGPT can

guide clinicians through relevant diagnostic criteria and management options, thereby streamlining care delivery and supporting timely decision-making.

In medical education, ChatGPT enhances learning by answering students' questions and generating quizzes. This interactive approach deepens students' understanding of complex medical concepts, which is crucial for preparing future healthcare professionals. Studies have shown that students using AI-assisted tools in medical schools perform better in assessments compared to those who use traditional learning methods.

Moreover, ChatGPT can automate administrative tasks, such as generating clinical reports and discharge summaries, which reduces the workload on healthcare providers and improves efficiency in clinical settings. For example, a hospital that implemented ChatGPT for documentation reported a 30% reduction in administrative workload, allowing physicians to allocate more time to direct patient care.

Despite these advantages, it is crucial to address the challenges and risks associated with ChatGPT's deployment in healthcare. Regular updates are essential to maintain the model's accuracy in light of rapid advancements in medical knowledge. Additionally, ethical oversight mechanisms are needed to mitigate risks related to misinformation, patient privacy, and potential biases in AI-generated content.

In examining ChatGPT's effectiveness across different organ systems and specialties, we observe that its utility varies based on clinical complexity and specialty-specific requirements. For example, in fields such as dermatology and nephrology, ChatGPT has demonstrated higher accuracy in providing preliminary educational information and assisting in patient self-care. However, in specialties requiring nuanced interpretation, such as neurology or oncology, ChatGPT may fall short without additional expert oversight. This comparison underscores the importance of customizing ChatGPT's use according to the specific demands of each medical field. One key area of concern involves intellectual property and regulatory issues surrounding AI in healthcare. Using ChatGPT to generate medical content raises questions about ownership and authorship, especially regarding the originality of AI-produced materials. Ensuring that ChatGPT's outputs comply with intellectual property laws is essential to maintaining the integrity of healthcare information. Moreover, compliance with data privacy regulations, such as HIPAA and GDPR, is critical when ChatGPT processes sensitive patient data. Implementing clear guidelines and robust data protection measures can safeguard patient information while ensuring the ethical use of AI in medical practice.

Furthermore, fostering effective AI-human collaboration is essential for maximizing ChatGPT's benefits in healthcare. Training healthcare professionals in the appropriate use of AI tools can enhance decision-making and improve patient outcomes. For instance, adopting collaborative frameworks where ChatGPT supports clinicians—while maintaining human oversight as central to patient care—could lead to more accurate diagnoses and tailored treatment plans.

By addressing these intellectual property, regulatory, and collaboration aspects, we aim to provide a comprehensive understanding of ChatGPT's potential, limitations, and responsible integration in healthcare.

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 and potentially harm patients. Ensuring the accuracy and reliability of AI-generated information is crucial, especially in critical healthcare contexts.

While ChatGPT automates customer support, this could result in job losses for administrative staff. This raises ethical considerations regarding workforce displacement and the need for retraining programs to help affected employees transition to new roles within the healthcare system. Balancing automation with human employment is essential to maintain a skilled and motivated workforce.

Ethical implications are paramount, including patient privacy and informed consent. Mishandling sensitive data could undermine trust in the healthcare system and violate regulations such as HIPAA or the General Data Protection Regulation (GDPR). Robust data protection measures and transparent data handling policies must be implemented to safeguard patient information. Additionally, clear communication about data usage and obtaining explicit patient consent are necessary to uphold ethical standards.

Additionally, it cannot provide personalized medical advice, lacking the capacity to tailor responses based on individual patient histories. This limitation can lead to generic recommendations that may not be suitable for all patients, potentially resulting in ineffective or harmful outcomes. Its responses may lack nuance, particularly in mental health contexts, where empathy is essential. The absence of emotional intelligence in AI models means they cannot replace the empathetic support provided by human professionals, which is critical in mental health care.

While capable of assisting with information, it struggles with fact based information, requiring excessive detail for effective results. Moreover, ChatGPT may sometimes produce plausible-sounding but incorrect answers, known as "AI hallucinations", which can mislead users. Implementing verification mechanisms and cross-referencing AI outputs with trusted medical sources can help mitigate this issue. Furthermore, it cannot fully replace the critical thinking of human professionals, especially in rapidly evolving medical fields, as it may miss crucial developments. AI models rely on their training data up to a certain cutoff date and may not incorporate the latest research findings or clinical guidelines. Regular updates and continuous training of the AI model are necessary to keep it current. Additionally, encouraging a collaborative approach where AI supports but does not replace human judgment can enhance decision-making while minimizing risks.

AI model biases also pose significant challenges. ChatGPT's training data may contain inherent biases that can lead to unequal treatment recommendations or misdiagnosis for certain populations. For example, minority groups or those with rare conditions might receive less accurate information due to underrepresentation in the data. Addressing these biases requires diversifying training datasets and implementing algorithms that detect and correct biased outputs.

Real-world implementation challenges include integrating ChatGPT into existing healthcare systems, which may involve technical hurdles such as compatibility with electronic health records (EHRs) and ensuring data security. Regulatory compliance is another critical factor. Healthcare AI applications must adhere to regulations like the FDA's guidelines on software as a medical device, necessitating rigorous validation and approval processes.

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, future research should focus on developing advanced algorithms that enable ChatGPT to analyze this information effectively. This includes refining natural language processing capabilities to interpret complex medical data and patient histories, supporting physicians in identifying symptom patterns, and suggesting appropriate diagnostic tests for personalized treatment plans. Collaborative studies involving data scientists, clinicians, and AI specialists are essential to create models that are both accurate and clinically relevant.

In medical education, ChatGPT can act as a virtual tutor, offering immediate feedback on clinical scenarios and generating customized practice questions for exam preparation. Research could explore the integration of ChatGPT into medical curricula, assessing its impact on student learning outcomes, knowledge retention, and critical thinking skills. Pilot programs in educational institutions can provide valuable insights into best practices for AI-assisted learning.

The future of ChatGPT in healthcare is promising, with anticipated improvements in accuracy and speed as OpenAI refines its models. Addressing biases and enhancing accuracy requires specific research into bias detection and mitigation strategies within AI systems. Developing transparent algorithms and incorporating diverse datasets can help reduce disparities in AI-generated recommendations. Additionally, establishing clear ethical and regulatory guidelines is crucial to ensure the responsible use of ChatGPT in clinical settings. This includes compliance with patient privacy laws, data security measures, and protocols for AI oversight.

Integrating ChatGPT with electronic health records (EHR) could enable it to analyze patient data and facilitate informed clinical decisions. Practical aspects of this integration involve ensuring interoperability between AI systems and existing healthcare IT infrastructure. Research should focus on developing standardized APIs and data exchange formats that allow seamless communication between ChatGPT and EHRs. Additionally, evaluating the impact of such integration on clinical workflows and patient outcomes through real-world trials can guide effective implementation strategies.

Additionally, ChatGPT can streamline recruitment in healthcare organizations by automating candidate screening and interview scheduling, thereby reducing administrative burdens and accelerating onboarding processes. Investigating the use of AI in human resources within healthcare, including compliance with employment laws and ethical considerations, can help in developing robust AI solutions for administrative functions.

In medical tourism, ChatGPT could assist patients in booking travel arrangements, providing real-time updates on flight statuses, and offering information on local healthcare services. Research into cross-border data sharing regulations, multilingual support, and cultural sensitivity is important to tailor ChatGPT's services for international patients.

In eLearning, ChatGPT is set to transform medical education by generating tailored learning materials and assessments, thereby enhancing training efficiency. Future studies could examine the long-term effects of AI-assisted education on professional competence and patient care quality. Collaborations between educational institutions and AI developers can facilitate the creation of specialized educational tools that meet the needs of learners.

As technology evolves, it is imperative to explore deeper into the practical aspects of integration and collaboration between humans and AI. This includes developing user-friendly interfaces that facilitate seamless interaction with ChatGPT, training programs for healthcare professionals to effectively use AI tools, and establishing protocols that define the roles of AI and humans in clinical decision-making. Research should also focus on understanding the psychological and organizational factors that influence the acceptance of AI in healthcare settings, aiming to foster a collaborative environment where AI augments human capabilities rather than replacing them.

10. Conclusions

ChatGPT demonstrates substantial potential to transform healthcare by enhancing communication between patients and providers, supporting clinical decision-making, and streamlining administrative processes. Its ability to generate structured, human-like responses equips it to assist with tasks such as drafting medical reports, summarizing patient interactions, and performing preliminary triage. Moreover, ChatGPT's capacity to process extensive medical literature and adverse event data allows it to support healthcare professionals in recognizing critical trends that contribute to improved patient safety and quality of care.

However, while the integration of ChatGPT into healthcare offers numerous benefits, it is crucial to recognize the model's limitations. ChatGPT cannot replace the expertise and nuanced judgment of healthcare professionals, especially in complex medical scenarios. Issues such as potential ethical concerns, data privacy challenges, lack of accountability, and difficulties in interpreting specialized medical data underscore the need for rigorous oversight. Reliance on AI-generated responses without expert validation could risk patient care, particularly when specialized clinical judgment is required.

Future work should focus on evaluating ChatGPT's performance alongside other large language models (LLMs) to provide a broader understanding of how different AI tools perform in healthcare settings. Comparative studies with healthcare-specific LLMs, such as BioBERT or PubMedBERT, could reveal insights into model strengths, limitations, and application-specific suitability. Additionally, developing a regulatory framework for the responsible integration of LLMs into healthcare is essential, with an emphasis on patient safety, accuracy, and adherence to ethical standards. Addressing these considerations will enable ChatGPT and similar models to serve as valuable adjuncts to healthcare practice, complementing human expertise and allowing healthcare professionals to deliver more efficient, high-quality care.

Author Contributions: Conceptualization, F.N.; Methodology, F.N.; Writing—original draft, F.N., D.B., D.K.S. and M.A.; Funding acquisition, F.N., D.B., D.K.S. and M.A.; Writing—review & editing, F.N., D.B., D.K.S. and M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by Kent State University's Open Access APC Support Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data were presented in the main text.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bajwa, J.; Munir, U.; Nori, A.; Williams, B. Artificial intelligence in healthcare: Transforming the practice of medicine. *Future Healthc. J.* **2021**, *8*, e188–e194. [CrossRef] [PubMed]
2. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthc. J.* **2019**, *6*, 94–98. [CrossRef] [PubMed]
3. Tan, T.; Thirunavukarasu, A.; Campbell, J.; Keane, P.; Pasquale, L.; Abramoff, M.; Kalpathy-Cramer, J.; Lum, F.; Kim, J.; Baxter, S.; et al. Generative Artificial Intelligence Through ChatGPT and Other Large Language Models in Ophthalmology: Clinical Applications and Challenges. *Ophthalmol. Sci.* **2023**, *3*, 100394. [CrossRef] [PubMed]
4. Liévin, V.; Hother, C.E.; Motzfeldt, A.G.; Winther, O. Can large language models reason about medical questions? *arXiv* **2023**, arXiv:2207.08143. [CrossRef]
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 Trans. Benchmarks Stand. Eval.* **2022**, *2*, 100089. [CrossRef]
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.; et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* **2021**, *372*, n160. [CrossRef]
7. Khurana, D.; Koli, A.; Khatter, K.; Singh, S. Natural language processing: State of the art, current trends and challenges. *Multimed. Tools Appl.* **2023**, *82*, 3713–3744. [CrossRef] [PubMed]
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. *Nat. Lang. Process. J.* **2023**, *4*, 100026. [CrossRef]
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. *Wirel. Pers. Commun.* **2023**, *130*, 857–908. [CrossRef]
10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762. [CrossRef]
11. Mikolov, T.; Karafiát, M.; Burget, L.; Cernockỳ, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the Interspeech, Makuhari, Japan, 26–30 September 2010; Volume 2, pp. 1045–1048.
12. Graves, A.; Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 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.; et al. A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–45. [CrossRef]
14. Alaparthi, S.; Mishra, M. Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. *arXiv* **2020**, arXiv:2007.01127.
15. Liu, X.; Zheng, Y.; Du, Z.; Ding, M.; Qian, Y.; Yang, Z.; Tang, J. GPT understands, too. *AI Open* **2023**, *5*, 208–215. [CrossRef]

16. Gorenstein, L.; Konen, E.; Green, M.; Klang, E. Bidirectional Encoder Representations from Transformers in Radiology: A Systematic Review of Natural Language Processing Applications. *J. Am. Coll. Radiol.* **2024**, *21*, 914–941. [CrossRef]
17. Devlin, J. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
18. Shah Jahan, M.; Khan, H.U.; Akbar, S.; Umar Farooq, M.; Gul, S.; Amjad, A. Bidirectional language modeling: A systematic literature review. *Sci. Program.* **2021**, *2021*, 6641832. [CrossRef]
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 Trans. Comput. Healthc. (HEALTH)* **2021**, *3*, 1–23. [CrossRef]
20. Huang, K.; Altosaar, J.; Ranganath, R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. *arXiv* **2020**, arXiv:1904.05342. [CrossRef]
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. [CrossRef]
22. National Library of Medicine (US). UMLS Knowledge Sources. 2024. Available online: http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html (accessed on 5 July 2024).
23. Nazi, Z.A.; Peng, W. Large language models in healthcare and medical domain: A review. *Informatics* **2024**, *11*, 57. [CrossRef]
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* **2023**, arXiv:2304.02020. [CrossRef]
25. Sanderson, K. GPT-4 is here: What scientists think. *Nature* **2023**, *615*, 773. [CrossRef] [PubMed]
26. Radford, A.; Narasimhan, K. Improving Language Understanding by Generative Pre-Training. 2018. Available online: https://api.semanticscholar.org/CorpusID:49313245 (accessed on 5 July 2024).
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.; et al. Language Models are Few-Shot Learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
29. 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. *J. Med. Internet Res.* **2024**, *26*, e57896. [CrossRef]
30. Samala, A.D.; Rawas, S. Generative AI as Virtual Healthcare Assistant for Enhancing Patient Care Quality. *Int. J. Online Biomed. Eng.* **2024**, *20*, 174. [CrossRef]
31. Ferdush, J.; Begum, M.; Hossain, S.T. ChatGPT and clinical decision support: Scope, application, and limitations. *Ann. Biomed. Eng.* **2024**, *52*, 1119–1124. [CrossRef]
32. Iftikhar, L.; Iftikhar, M.F.; Hanif, M.I. Docgpt: Impact of chatgpt-3 on health services as a virtual doctor. *EC Paediatr.* **2023**, *12*, 45–55.
33. Zheng, Y.; Wang, L.; Feng, B.; Zhao, A.; Wu, Y. Innovating healthcare: The role of ChatGPT in streamlining hospital workflow in the future. *Ann. Biomed. Eng.* **2024**, *52*, 750–753. [CrossRef]
34. Awal, S.; Awal, S. ChatGPT and the healthcare industry: A comprehensive analysis of its impact on medical writing. *J. Public Health* **2023**. [CrossRef]
35. Temsah, O.; Khan, S.A.; Chaiah, Y.; Senjab, A.; Alhasan, K.; Jamal, A.; Aljamaan, F.; Malki, K.H.; Halwani, R.; Al-Tawfiq, J.A.; et al. Overview of early ChatGPT's presence in medical literature: Insights from a hybrid literature review by ChatGPT and human experts. *Cureus* **2023**, *15*, e37281. [CrossRef] [PubMed]
36. 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. *Glob. Transit.* **2023**, *5*, 50–54. [CrossRef]
37. Abd Karim, R.; Cakir, G.K. Investigating ChatGPT Usability in Promoting Smart Health Awareness. In *Industry 5.0 for Smart Healthcare Technologies*; CRC Press: Boca Raton, FL, USA, 2024; pp. 227–237.
38. Baldwin, A.J. An artificial intelligence language model improves readability of burns first aid information. *Burns* **2024**, *50*, 1122–1127. [CrossRef]
39. Neha, F. Kidney Localization and Stone Segmentation from a CT Scan Image. In Proceedings of the 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 18–19 August 2023; pp. 1–6.
40. Neha, F.; Bansal, A.K. Multi-Layer Feature Fusion with Cross-Channel Attention-Based U-Net for Kidney Tumor Segmentation. *arXiv* **2024**, arXiv:2410.15472.
41. Neha, F.; Bansal, A.K. Convnext-PCA: A Parameter-Efficient Model for Accurate Kidney Abnormality Classification. In Proceedings of the 2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP), London, UK, 22–25 September 2024; pp. 1–6. [CrossRef]
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. *Sci. Rep.* **2024**, *14*, 1542. [CrossRef]
43. Miao, J.; Thongprayoon, C.; Craici, I.M.; Cheungpasitporn, W. How to improve ChatGPT performance for nephrologists: A technique guide. *J. Nephrol.* **2024**, *37*, 1397–1403. [CrossRef] [PubMed]
44. Janus, N. A Comparative Analysis of Chatgpt Vs Expert in Managing Anticancer Drug in Patients Renal Insufficiency. *Blood* **2023**, *142*, 7186. [CrossRef]

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. *Contemp. Oncol. Onkol.* **2024**, *28*, 172–181. [CrossRef]
46. Javid, M.; Bhandari, M.; Parameshwari, P.; Reddiboina, M.; Prasad, S. Evaluation of ChatGPT for patient counseling in kidney stone clinic: A prospective study. *J. Endourol.* **2024**, *38*, 377–383. [CrossRef]
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. *J. Pers. Med.* **2023**, *13*, 1681. [CrossRef] [PubMed]
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. *Clin. Pract.* **2023**, *13*, 1160–1172. [CrossRef] [PubMed]
49. German, R.Z.; Palmer, J.B. Anatomy and development of oral cavity and pharynx. *GI Motil. Online* **2006**. [CrossRef]
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. *Eur. Arch.-Oto-Rhino-Laryngol.* **2024**, *281*, 319–333. [CrossRef]
51. Aaronson, P.I.; Ward, J.P.; Connolly, M.J. *The Cardiovascular System at a Glance*; John Wiley & Sons: Hoboken, NJ, USA, 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. [CrossRef]
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. *Curr. Probl. Cardiol.* **2024**, *49*, 102797. [CrossRef] [PubMed]
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 Circ.* **2024**, *33*, 1314–1318. [CrossRef] [PubMed]
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. *Appl. Med. Inform.* **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* **2023**, *11*, 2906. [CrossRef]
57. Ayub, M.; Mallamaci, A. An Introduction: Overview of Nervous System and Brain Disorders. In *The Role of Natural Antioxidants in Brain Disorders*; Springer: Cham, Switzerland, 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. *Clin. Neurol. Neurosurg.* **2024**, *239*, 108238. [CrossRef]
59. Adesso, G. Towards the ultimate brain: Exploring scientific discovery with ChatGPT AI. *AI Mag.* **2023**, *44*, 328–342. [CrossRef]
60. Fei, X.; Tang, Y.; Zhang, J.; Zhou, Z.; Yamamoto, I.; Zhang, Y. Evaluating cognitive performance: Traditional methods vs. ChatGPT. *Digit. Health* **2024**, *10*, 20552076241264639. [CrossRef]
61. Al-Suhaimi, E.A.; Khan, F.A. Thyroid glands: Physiology and structure. In *Emerging Concepts in Endocrine Structure and Functions*; Springer: Berlin/Heidelberg, Germany, 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*, e47576. [CrossRef] [PubMed]
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. *Am. J. Otolaryngol.* **2024**, *45*, 104144. [CrossRef] [PubMed]
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. *Ann. Clin. Biochem.* **2024**, *61*, 143–149. [CrossRef] [PubMed]
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. *Int. J. Med. Inform.* **2024**, *191*, 105593. [CrossRef]
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. [CrossRef]
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. [CrossRef]
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.; et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin. Mol. Hepatol.* **2023**, *29*, 721. [CrossRef] [PubMed]
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.; et al. GPT-4 outperforms ChatGPT in answering non-English questions related to cirrhosis. *medRxiv* **2023**. [CrossRef]
70. Shahsavari, D.; Parkman, H.P. Normal gastrointestinal tract physiology. In *Nutrition, Weight, and Digestive Health: The Clinician's Desk Reference*; Springer: Berlin/Heidelberg, Germany, 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*, e46736. [CrossRef]
72. Ma, Y. The potential application of ChatGPT in gastrointestinal pathology. *Gastroenterol. Endosc.* **2023**, *1*, 130–131. [CrossRef]
73. Liu, X.; Wang, Y.; Huang, Z.; Xu, B.; Zeng, Y.; Chen, X.; Wang, Z.; Yang, E.; Lei, X.; Huang, Y.; et al. The Application of ChatGPT in Responding to Questions Related to the Boston Bowel Preparation Scale. *arXiv* **2024**, arXiv:2402.08492.
74. Chang, E.B.; Leung, P.S. Pancreatic physiology. In *The Gastrointestinal System: Gastrointestinal, Nutritional and Hepatobiliary Physiology*; Springer: Berlin/Heidelberg, Germany, 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. *J. Transl. Med.* **2024**, *22*, 527. [CrossRef] [PubMed]
76. Qiu, J.; Luo, L.; Zhou, Y. Accuracy of ChatGPT3. 5 in answering clinical questions on guidelines for severe acute pancreatitis. *BMC Gastroenterol.* **2024**, *24*, 260. [CrossRef] [PubMed]
77. Lorenzo, A.J.; Bagli, D. Basic science of the urinary bladder. In *Clinical Pediatric Urology*; Informa Healthcare: London, UK, 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. *Soc. Int. D'Urologie J.* **2024**, *5*, 214–224. [CrossRef]
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? *Int. Braz. J. Urol.* **2024**, *50*, 192–198. [CrossRef]
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. *Clin. Genitourin. Cancer* **2024**, *22*, 454–457. [CrossRef]
81. Cakir, H.; Caglar, U.; Yildiz, O.; Meric, A.; Ayranci, A.; Ozgor, F. Evaluating the performance of ChatGPT in answering questions related to urolithiasis. *Int. Urol. Nephrol.* **2024**, *56*, 17–21. [CrossRef]
82. Sagir, S. Evaluating the accuracy of ChatGPT addressing urological questions: A pilot study. *J. Clin. Trials Exp. Investig.* **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. *Infect. Dis. Now* **2024**, *54*, 104884. [CrossRef]
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 J. Urol.* **2023**, *41*, 3149–3153. [CrossRef] [PubMed]
85. Mashinini, M. Pituitary gland and growth hormone. *S. Afr. J. Anaesth. Analg.* **2020**, *26*, S109–S112. [CrossRef]
86. Sambangi, A.; Carreras, A.; Campbell, D.; Bray, D.; Evans, J.J. Evaluating Chatgpt for Patient Education Regarding Pituitary Adenoma Resection. *J. Neurol. Surg. 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. *J. Pers. Med.* **2024**, *14*, 330. [CrossRef] [PubMed]
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. *Digit. Health* **2024**, *10*, 20552076241278692. [CrossRef]
89. Hafez, B.; Hafez, E. Anatomy of female reproduction. In *Reproduction in Farm Animals*; Wiley: Hoboken, NJ, USA, 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. *Gynecol. Oncol.* **2024**, *183*, 115–119. [CrossRef]
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. *Int. J. Gynecol. Obstet.* **2024**, *166*, 1127–1133. [CrossRef] [PubMed]
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. *Arch. Gynecol. Obstet.* **2024**, *310*, 2425–2437. [CrossRef] [PubMed]
93. Winograd, D.; Alterman, C.; Appelbaum, H.; Baum, J. 51. Evaluation of ChatGPT Responses to Common Puberty Questions. *J. Pediatr. Adolesc. Gynecol.* **2024**, *37*, 261. [CrossRef]
94. Edwards, S. The skin. In *Essential Pathophysiology For Nursing And Healthcare Students*; McGraw-Hill Education: Maidenhead, UK, 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*, e35742. [CrossRef]
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. *J. Am. Acad. Dermatol.* **2024**, *91*, AB205. [CrossRef]
97. Passby, L.; Jenko, N.; Wernham, A. Performance of ChatGPT on Specialty Certificate Examination in Dermatology multiple-choice questions. *Clin. Exp. Dermatol.* **2024**, *49*, 722–727. [CrossRef] [PubMed]
98. Stoneham, S.; Livesey, A.; Cooper, H.; Mitchell, C. ChatGPT versus clinician: Challenging the diagnostic capabilities of artificial intelligence in dermatology. *Clin. Exp. Dermatol.* **2024**, *49*, 707–710. [CrossRef] [PubMed]

99. Mondal, H.; Mondal, S.; Podder, I. Using ChatGPT for writing articles for patients' education for dermatological diseases: A pilot study. *Indian Dermatol. Online J.* **2023**, *14*, 482–486. [CrossRef]
100. Mat Lazim, N. Introduction to Head and Neck Surgery. In *Head and Neck Surgery: Surgical Landmark and Dissection Guide*; Springer: Berlin/Heidelberg, Germany, 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.; et al. Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: A multicenter collaborative analysis. *Otolaryngol.–Head Neck Surg.* **2024**, *170*, 1492–1503. [CrossRef]
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: London, UK, 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 Surg. Oral Med. Oral Pathol. Oral Radiol.* **2024**, *138*, 269–278. [CrossRef]
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*, e48518. [CrossRef] [PubMed]
105. Mago, J.; Sharma, M. The potential usefulness of ChatGPT in oral and maxillofacial radiology. *Cureus* **2023**, *15*, e42133. [CrossRef]
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. *Int. J. Oral Maxillofac. Surg.* **2024**, *53*, 78–88. [CrossRef]
107. Fahey, J. Optimising lung health. *J. Aust.-Tradit.-Med. Soc.* **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. [CrossRef] [PubMed]
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. [CrossRef]
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**, . [CrossRef]
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. *Am. J. Roentgenol.* **2024**. [CrossRef]
112. Schulte, B. Capacity of ChatGPT to identify guideline-based treatments for advanced solid tumors. *Cureus* **2023**, *15*, e37938. [CrossRef] [PubMed]
113. White, T.D.; Folkens, P.A. *The Human Bone Manual*; Elsevier: Amsterdam, The Netherlands, 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? *Eur. J. Orthop. Surg. Traumatol.* **2024**, *34*, 2757–2765. [CrossRef]
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*, e22409. [CrossRef] [PubMed]
116. Cinar, C. Analyzing the performance of ChatGPT about osteoporosis. *Cureus* **2023**, *15*, e45890. [CrossRef]
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. *J. Bone Oncol.* **2024**, *44*, 100525. [CrossRef]
118. Kell, R.T.; Bell, G.; Quinney, A. Musculoskeletal fitness, health outcomes and quality of life. *Sport. Med.* **2001**, *31*, 863–873. [CrossRef]
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*, e67347. [CrossRef] [PubMed]
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*, e40977. [CrossRef] [PubMed]
121. Saluja, S.; Tigga, S.R. Capabilities and Limitations of ChatGPT in Anatomy Education: An Interaction with ChatGPT. *Cureus* **2024**, *16*, e69000. [CrossRef]
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 Surg. Sport. Traumatol. Arthrosc.* **2023**, *31*, 5190–5198. [CrossRef]
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. *Biomed. Signal Process. Control* **2024**, *97*, 106676. [CrossRef]
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. *Surg. Radiol. Anat.* **2024**, *46*, 1885–1890. [CrossRef]
125. Li, J.; You, M.; Chen, X.; Li, P.; Deng, Q.; Wang, K.; Wang, L.; Xu, Y.; Liu, D.; Ye, L.; et al. 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. *J. Orthop. Surg. Res.* **2024**, *19*, 635.

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. [CrossRef]
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. [CrossRef]
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. [CrossRef] [PubMed]
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. [CrossRef] [PubMed]
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. [CrossRef] [PubMed]
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* **2023**. [CrossRef]
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* **2024**, *12*, 562. [CrossRef] [PubMed]
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. *Sci. Rep.* **2024**, *14*, 9330. [CrossRef]
134. Liu, J.; Wang, C.; Liu, S. Utility of ChatGPT in Clinical Practice. *J. Med. Internet Res.* **2023**, *25*, e48568. [CrossRef]
135. Teixeira da Silva, J.A. Can ChatGPT rescue or assist with language barriers in healthcare communication? *Patient Educ. Couns.* **2023**, *115*, 107940. [CrossRef]
136. Liu, Z.; Zhang, L.; Wu, Z.; Yu, X.; Cao, C.; Dai, H.; Liu, N.; Liu, J.; Liu, W.; Li, Q.; et al. Surviving ChatGPT in healthcare. *Front. Radiol.* **2024**, *3*, 1224682. [CrossRef]
137. Sedaghat, S. Plagiarism and Wrong Content as Potential Challenges of Using Chatbots Like ChatGPT in Medical Research. *J. Acad. Ethics* **2024**. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.