

Review

Not peer-reviewed version

---

# Innovations in Medicine: Exploring ChatGPT's Impact on Rare Disorder Management

---

[Stefania Zampatti](#) , Cristina Peconi , [Domenica Megalizzi](#) , [Giulia Calvino](#) , [Giulia Trastulli](#) , [Raffaella Cascella](#) , [Claudia Strafella](#) , [Carlo Caltagirone](#) , [Emiliano Giardina](#) \*

Posted Date: 15 March 2024

doi: [10.20944/preprints202403.0876.v1](https://doi.org/10.20944/preprints202403.0876.v1)

Keywords: Artificial Intelligence; ChatGPT; Rare Disorders



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

Review

# Innovations in Medicine: Exploring ChatGPT's Impact on Rare Disorder Management

Stefania Zampatti <sup>1</sup>, Cristina Peconi <sup>1</sup>, Domenica Megalizzi <sup>1</sup>, Giulia Calvino <sup>1</sup>, Giulia Trastulli <sup>1</sup>, Raffaella Cascella <sup>1,2</sup>, Claudia Strafella <sup>1</sup>, Carlo Caltagirone <sup>3</sup> and Emiliano Giardina <sup>1,4,\*</sup>

<sup>1</sup> Genomic Medicine Laboratory UILDM, IRCCS Santa Lucia Foundation, 00179 Rome, Italy

<sup>2</sup> Department of Biomedical Sciences, Catholic University Our Lady of Good Counsel, Tirana, Albania

<sup>3</sup> Department of Clinical and Behavioral Neurology, IRCCS Fondazione Santa Lucia, Rome, Italy

<sup>4</sup> Department of Biomedicine and Prevention, Tor Vergata University, 00133 Rome, Italy

\* Correspondence: emiliano.giardina@uniroma2.it; Tel.: +39-06 5150 1550

**Abstract:** Artificial Intelligence (AI) is rapidly transforming the field of medicine, announcing a new era of innovation and efficiency. Among AI programs designed for general use ChatGPT holds a prominent position using an innovative language model developed by OpenAI. Thanks to the use of deep learning techniques, ChatGPT stands out as an exceptionally viable tool, renowned for generating human-like responses to queries. Various medical specialties, including rheumatology, oncology, psychiatry, internal medicine, and ophthalmology, have been explored for ChatGPT integration, with pilot studies and trials revealing each field's potential benefits and challenges. However, the field of genetics and genetic counseling, as well as that of rare disorders, represents an area suitable for exploration, with its complex data sets and the need for personalized patient care. In this review, we synthesize the wide range of potential applications for ChatGPT in the medical field, highlighting its benefits and limitations. We pay special attention to rare and genetic disorders, aiming to shed light on the future roles of AI-driven chatbots in healthcare. Our goal is to pave the way for a healthcare system that is more knowledgeable, efficient, and centered around patient needs.

**Keywords:** Artificial Intelligence; ChatGPT; rare disorders

---

## 1. Introduction

Artificial Intelligence (AI) is rapidly transforming the field of medicine, announcing a new era of innovation and efficiency. By integrating advanced algorithms and machine learning techniques, AI can enhance diagnostic accuracy, personalize treatment plans, and simplify healthcare operations. This technological evolution is expected to revolutionize patient care, making it more precise, predictive, and personalized than ever before. Among AI programs designed for general use ChatGPT holds a prominent position using an innovative language model developed by OpenAI, based in California, USA. Thanks to the use of deep learning techniques, ChatGPT stands out as an exceptionally viable tool, renowned for generating human-like responses to queries. This ability stems from its foundation as one of the Generative Pre-training Transformer (GPT) models, specifically designed to understand, interpret, and generate human language with remarkable proficiency [1–5]. ChatGPT, trained on extensive text datasets, generates outputs that closely mimic the style and content of its training material and maintain relevance and coherence with the input context [6]. Its deep learning basis enables the model to identify patterns in large data volumes, essential for producing accurate predictions and executing specific tasks [7]. At its core, ChatGPT facilitates interactions that resemble human conversation, which is a significant advancement in enabling machines to understand and produce human language in previously unexpected ways. This marks a crucial step in developing more intelligent, responsive, and adaptable AI systems.

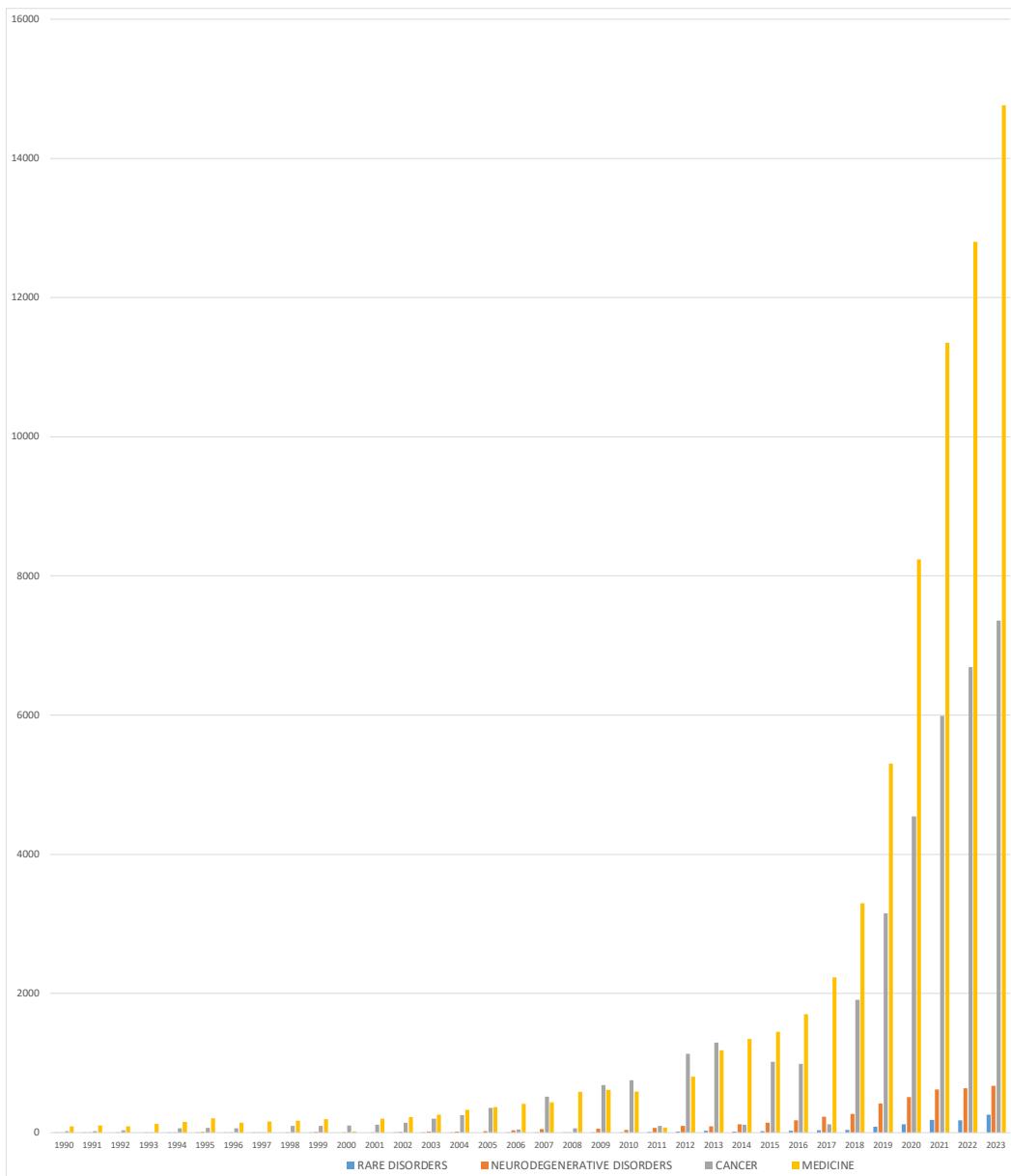
ChatGPT became the first large language model (LLM) achieving broad acceptance and curiosity across the general public [8]. As a chatbot technology, it responds to a diverse array of inquiries, engaging users in dialogues that bear a striking resemblance to natural human conversation. The model gets smarter with each conversation because it learns from its interactions [9,10].

It employs a transformer model a type of neural network architecture able to analyze the input data, understand the nuances of the language, and map out relationships between words and phrases. Through this process, ChatGPT determines the most appropriate words and phrases to construct a coherent and relevant response in a specific context [11]. The intelligence of ChatGPT lies in its ability to simulate an internet-like breadth of knowledge through its training, enabling it to provide informative and conversationally fluent outputs that are remarkably human-like in their presentation.

Upon its release, it captured the public's attention, collecting one million registered users within only five days. By the end of two months, this number had risen to over 100 million, underscoring the chatbot's widespread appeal [12]. While it does not pull information from the internet in real time, the illusion of comprehensive knowledge is crafted through its prior training on an extensive dataset. Its ability to engage in natural, human-like dialogue and provide informative answers with remarkable contextual awareness has fueled its adoption by curious users worldwide.

Furthermore, from September 2023, ChatGPT can browse the internet [13,14]. It generates responses based on a mixture of licensed data, data created by human trainers, and publicly available data, which is then used to pre-train the model. To date, its responses are generated based on a fixed dataset available up to the last training cut-off in April 2023. One of the significant limitations of articles on ChatGPT published before September 2023 is the restricted access of the chatbot only to information before 2021. The update of the ChatGPT in September 2023 allows it to access a more completed and updated dataset (dated April 2023). It is important to note that while ChatGPT can provide varied responses to repeated queries over time, this variation is not due to live internet access or real-time data updates, rather it is a result of its design to produce non-deterministic outputs. Nonetheless, the repetitive questioning in different times might results in different answers with some critical differences in completeness and correctness, as reported below.

Integrating Artificial Intelligence (AI) into the healthcare sector signs a new era, characterized by enhanced precision, accuracy, and efficiency. While AI has been widely adopted in domains such as customer service and data management, its use into healthcare and medical research needs a careful consideration. The use of AI in healthcare is not only beneficial but crucial, given its capacity to transform medical practices through time efficiency. However, it is equally essential to methodically assess its limitations to prevent unacceptable mistakes and errors in medicine [15]. Despite its considerable potential for applications across various medical fields, AI remains rarely applied to rare diseases. Specifically, an examination of the volume of scientific publications since the 1990s revealed a clear growth in several medical fields, with a slight increasing in rare disease (Figure 1).



**Figure 1.** Histogram showing the number of PubMed results per year ("Artificial Intelligence" and respectively "rare disorders", "neurodegenerative disorders", "cancer", and "medicine").

One of the primary benefits of ChatGPT in the medical domain is its contribution to research and education. With its advanced writing abilities and contextually adaptive language, ChatGPT could have the potential to be a powerful tool for synthesizing research, drafting papers, and composing coherent and context-aware literature reviews though concerns exist regarding its accuracy, potential for misuse, and the ethical implications of its application. Also, ChatGPT can play an essential role in education, though it has some limits. It can help training medical professionals by offering interactive learning and simulating clinical situations for teaching. One of the most promising applications of ChatGPT is its ability to enhance clinical practice. It can provide initial diagnostic recommendations, assist in developing differential diagnoses, and suggesting treatment options.

Various medical specialties, including rheumatology, oncology, psychiatry, internal medicine, and ophthalmology, have been explored for ChatGPT integration, with pilot studies and trials revealing each field's potential benefits and challenges. However, the field of genetics and genetic counseling, as well as that of rare disorders, represents an area suitable for exploration, with its complex data sets and the need for personalized patient care.

In this review, we synthesize the wide range of potential applications for ChatGPT in the medical field, highlighting its benefits and limitations. We pay special attention to rare and genetic disorders, aiming to shed light on the future applications of AI-driven chatbots in healthcare. Our goal is to pave the way for a healthcare setting that is more knowledgeable, efficient, and centered around patient needs.

## 2. Medical Research and Literature Production

The potential applications of AI are extensive, reflecting its adaptability and the depth of its training. In academic contexts, the ChatGPT chatbot was utilized to assist in composing theses, structuring research projects, and drafting scientific articles [16]. This highlights ChatGPT's potential to facilitate the scholarly writing process, marking it as a significant tool for students, researchers, and academics. ChatGPT has passed the United States Medical Licensing Exam (USMLE), demonstrating its ability to learn and utilize complex medical knowledge to meet high professional standards [17,18].

Despite an expected shortfall in areas requiring high creativity, ChatGPT can be helpful in structuring original research outlines on specific topics. It can deliver a comprehensive research outline that meticulously resounds the structure and detail expected in standard research projects [19]. This demonstrates its understanding of academic norms and its ability to adjust language and output to make them suitable for the context provided. Despite ChatGPT's ability to draft articles based on selected scientific literature, its feasibility for topics on rare disorders is yet to be reached. This is primarily because language models like ChatGPT exhibit frequency bias—they perform better with concepts that are extensively covered in their training data and less so with lesser-known topics [20]. Consequently, the reliability of ChatGPT's responses is higher for diseases that are more prevalent in the dataset used for pre-training the model (last updated in April 2023) compared to those with less available information. For instance, it has been observed that ChatGPT's information on common conditions like osteoarthritis and ankylosing spondylitis is more accurate than that on relatively rare diseases, such as psoriatic arthritis [21].

While ChatGPT emerged as a significant supporting tool in drafting medical research articles, its ideal fit for this task warrants a critical examination. The AI's contribution to manuscript preparation can be helpful, yet the decision by some researchers to list ChatGPT as a co-author on publications suggests caution. This practice raises ethical and practical questions about the nature of authorship and intellectual contribution. Co-authorship traditionally conveys a degree of intellectual investment and responsibility for the content, aspects that AI, by its current design, cannot fulfill. Furthermore, the implications for accountability, especially in fields as sensitive as medical research, are profound. To date (last access on February 12th, 2024), PubMed acknowledges four articles that formally list ChatGPT in the authorship [22–25], while Scopus records three [26–28]. These figures are undoubtedly an underestimate. Many papers have acknowledged ChatGPT in the author list during their initial presentation, a fact that is inconsistent with reports from PubMed [28].

The support from chatbots was tested as a replacement of human authors. In June 2023, some scientists even produced a paper entirely using ChatGPT [29]. Prestigious scientific journals such as *Nature* and *JAMA Network Science* stated that they will not accept manuscripts generated by ChatGPT. In contrast, other journals ask authors to disclose the extent of ChatGPT's use and to affirm their responsibility for the paper's content [30–32]. Authorship guidelines distinguish between contributions made by humans and those made by ChatGPT. Specifically, authors are expected to provide substantial intellectual input. They must possess the ability to consent to co-authorship, thereby assuming responsibility for the paper or for the part to which they have contributed [33].

The ability of ChatGPT to compose scientific papers and abstracts in a manner indistinguishable from human output is remarkable. A study by Northwestern University (Chicago, Illinois, USA) highlighted this ability through a blind evaluation involving 50 original abstracts alongside their counterparts generated by ChatGPT from the original articles. These abstracts were randomly presented to a panel of medical researchers asking them to identify which were original and which were produced by ChatGPT. The outcome of this experiment was very intriguing since blinded

human reviewers correctly identified only 68% of the ChatGPT-generated abstracts as such. Conversely, they mistakenly classified 14% of the actual human-written abstracts as being created by the chatbot. The challenge in distinguishing between authentic and AI-generated abstracts highlights ChatGPT's skill in creating persuasive scientific narratives [34,35]. However, this raises critical concerns about the integrity of scientific communication. The fact that these AI-generated abstracts seem genuine and score high on originality according to plagiarism detection software introduces a paradox [15,36,37]. It suggests that while ChatGPT can produce work that is apparently innovative and unique, this may not necessarily reflect true originality or contribute to genuine scientific progress. This situation calls into question the reliability of using such software as the sole metric for originality and underscores the need for more nuanced approaches for evaluating the authenticity and value of scientific work. The reliance on AI for generating scientific content without critical oversight could risk diluting the scientific literature with works that, despite being original in form, lack the depth and rigor of human-generated research [38,39].

Among the multitude of information generated by ChatGPT, a critical limitation arises from the inability to verify its data sources, as its responses are synthesized from an extensive dataset from the web without direct access to or citation of these sources [13,14]. AI will never replace the human revision. ChatGPT's inability to distinguish between credible and less credible sources leads to transparency issues, treating all information equally. This differs from traditional research methods where source credibility can be verified. Users lack the means to judge information accuracy without external checks, highlighting concerns about the potential for misinformation and the necessity for external validation to ensure content reliability [19,40].

Furthermore, sometimes ChatGPT fabricate its references. Inaccurate references were reported as one of the three key features that guide the correct identification of human and AI-generated articles. In a single-blinded observer study, human and ChatGPT dermatology case reports were evaluated by 20 medical reviewers. One of the two selected case reports described a rare disease (posterior reversible encephalopathy syndrome) associated with pharmacological therapy. Human reviewers accurately identified AI-generated case reports in 35% of times. Three key features were reported by reviewers as essential for discrimination: poorly description of cutaneous findings, imprecise report of pathophysiology, and inaccurate references [41]. One of the notable challenges with ChatGPT's outputs is the phenomenon known as "artificial hallucination", particularly evident in its provision of creative references. These hallucinations refer to information or data generated by the chatbot that do not accurately reflect reality, although their realistic appearance. This issue is especially prevalent in references, where ChatGPT might cite sources, studies, or data that seem legitimate but do not exist or are inaccurately represented [42,43]. When questioned about liver involvement in late-onset Pompe disease, the chatbot provided details about the co-occurrence of the two conditions and suggested references to support its thesis. However, it is well known that the Pompe disease (Glycogen storage disease II, OMIM#232300) involves liver disease in infantile-onset forms, but hepatic features are unique or almost rare in the late-onset form [44]. Furthermore, the chatbot provided fabricated references to support its thesis [42,43]. These inaccuracies are not due to intentional misinformation but originate from the model's design to generate responses based on patterns learned from its training dataset. In this context, it is essential to underline that ChatGPT does not have access to the PubMed database, so it cannot realistically search for references. Indeed, when a user requests ChatGPT to provide references supporting their responses, it fabricates credible but nonexistent references [45]. In a recent study, Gravel and coworker evaluated 59 references provided by ChatGPT and retrieved that almost two thirds of them were fabricated.

Interestingly, fabricated references seemed real to a first glance. About 95% (56/59) of reported references contained authors who published papers in the requested field, and 100% reported titles that seemed appropriate. Despite this truthfulness, 69% of references were fabricated [45]. Interestingly, the change of the topic did not improve the truthfulness of references [46,47]. Another described ChatGPT hallucination is in the production of medical note, sometimes fabricating patient features (i.e., it generates BMI score, without height and weight data) [18,48].

ChatGPT proved to be useful in reviewing scientific manuscripts, being able to pinpoint their strengths and areas for improvement. Its skill set goes beyond content creation, including critical analysis and error detection, making it valuable for assessing medical data, even those it produces [16–19,49,50]. Its utility in these areas suggests that ChatGPT can assist researchers and medical professionals, offering a preliminary review that can optimize the revision process. However, it's imperative to integrate ChatGPT's output with expert human revisions to achieve the highest scientific communication and medical accuracy standards. This integration becomes particularly crucial in rare diseases. ChatGPT may inadvertently introduce errors in medical documents related to these diseases due to the nuances and complexities involved. Therefore, any material prepared by ChatGPT, including drafts and preliminary analyses, should undergo a comprehensive review by specialists in the relevant field. This ensures that the final documents shared with patients or submitted to scientific journals are accurate, reliable, and reflect the current medical knowledge, avoiding potential misdiagnoses or misinformation.

### 3. Education

ChatGPT could improve the dissemination of knowledge generating manuscripts in multiple languages. In some contexts, English could be an impediment and ChatGPT can bridge the gap generating copies of a manuscript in different languages. Similarly, in the conduction of cross-cultural research studies it may support the communication processes. Nevertheless, great attention should be paid to contents, in fact chatbot can generate misleading or inaccurate content with the risk to cause misrepresentation instead of knowledge dissemination [51].

ChatGPT demonstrated a good impact in limiting misinformation derived from the internet in cancer myths. In fact, despite many harmful information available online about cancer [52], the chatbot demonstrated a good accuracy in response to cancer myths [53]. Similarly, in other contexts, ChatGPT was helpful in providing comprehensive information to patients, helping them to understand medical information and treatment options [54]. Undoubtedly, it is advisable to subject ChatGPT to specific training to maintain its ability and prevent the sharing of incorrect information. It is particularly true for cases in which it has been required to answer questions about rare diseases, for which the available information on the web may be limited.

The platform's feasibility is one reason for its widespread diffusion. Other main strengths of ChatGPT were in form and accessibility of the platform. The user-chatbot interaction is straightforward and it mimics a dialogue. Not all information provided is accurate, and mistakes are difficult to detect because of the chatbot's linguistic ability. Correct and incorrect sentences are reported entirely appropriate for the context. For these reasons, ChatGPT could be applied to provide and explain basic medical information and treatment options, even in rare disorders, but it should be used with caution in other cases.

In this context, it is essential to note that young people are prone to use online resources instead of seeking help through face-to-face methods [55]. Thus, ChatGPT is expected to become one of the most interrogated tools for every need. It has already become one of the most trusted online chatbots. Notably, the trust is greater for administrative tasks (i.e. scheduling appointments) and lower for management of complex medical situations (i.e. treatment advice) [56].

ChatGPT showed a certain ability to detect diagnosis and provide medical advice when evaluating medical scenarios. In detail, a study on 96 unique vignettes representing clinical cases with different features (scenarios, clinical histories, ages, races, genders, and insurance statuses) reported that ChatGPT offered safe medical advice, often without specificity [57]. The substantial safeness of the chatbot may support the care continuum but confirm its inability to replace the medical judgment. As an example, a French study evaluated ChatGPT responses to a virtual patient affected by systemic lupus erythematosus (SLE) and asked for his treatment. The chatbot emphasized the need for medical evaluation but provided inconsistent information on hydroxychloroquine use during pregnancy and breastfeeding, as well as incorrect dosage suggestions. This highlights the risk of using ChatGPT's responses without medical supervision [16]. Similar issues were noted with cardiovascular

conditions, where it performed better on straightforward questions and case vignettes than on complex decision-making scenarios [58].

Likewise, ChatGPT performances on question resolution were confirmed high across different specialties. A recent study enrolled 33 physicians from 17 specialties to produce 180 medical questions. Each question was classified according to difficulty levels (easy, medium, and hard) and was fed to ChatGPT. Accuracy and completeness of ChatGPT answers were evaluated: the median accuracy score was 5/6 (mean 4.4, SD 1.7), and the median completeness score was 3/3 (mean 2.4, SD 0.7) [59]. Interestingly, ChatGPT performances on rare and familial disorders (as prolactinoma and age-related macular degeneration) were in line with other diseases, with a slightly improved result for age-related macular degeneration, for which there is many data available on the web. There were no significant differences according to the difficulty level of questions, unless for completeness scores that reached a median of 2.5/3 (mean 2.4, SD 0.7) for complex answers. [59]

It's widely recognized that language models, including ChatGPT, exhibit a frequency bias, performing better on concepts extensively covered in their training data and poorly on less represented topics [20]. Consequently, ChatGPT's reliability varies with the availability of information online; it is more accurate for diseases well-documented on the internet and less so for those with limited information. For instance, ChatGPT's insights on osteoarthritis and ankylosing spondylitis are notably more accurate than its information on psoriatic arthritis [21].

## 4. Medical Practice

### 4.1. Support in Communications

Mental health care is one of the most promising topics in which AI-driven chatbots have been developed. Several chatbots have been designed for psychoeducation, emotional support, and cognitive behavioral therapy [60–62]. These tools have been developed mainly because of known barriers in assessing treatment for mental disorders. Beyond the long waiting times and geographical limitations similar to other medical specialties, psychiatric patients have to face also with the stigma surrounding mental health [63]. One of the main acknowledged advantages of ChatGPT is the ability to generate sentences similar to a human-like conversation. In medical practice, there are some contexts in which the doctor-patient relationship may complicate the administration of evaluation questionnaires. In particular, in psychiatry, many interfering factors, such as the physician's voice, mood, and environment, can interfere with the patient's assessment. An impersonal interface such as ChatGPT can support the administration of questionnaires with human-like dialogues, but without the human interfering factors [64]. Likewise, there are other contexts where medical knowledge is deficient to detect rare disorders and associated risks. For example, it is well known that patients affected by Charcot-Marie Tooth disease should avoid some drugs, that may accelerate the disease's progression. Unfortunately, many patients remain undiagnosed [65], lacking appropriate management of their condition. The diffusion of friendly instruments such as ChatGPT, after a complete training for rare disorders, might improve the diagnostic ability of general doctors in detecting rare disorders or at least patients that should require a deeper evaluation.

It is expected that, with appropriate training, ChatGPT might be applied to clinical practice. It has been proposed to schedule appointments, collect anamnesis, and write medical records [38]. In particular, a study conducted by Casella and coworkers revealed that ChatGPT can correctly write a medical note for patient admission in an intensive care unit (ICU) based on provided health information (treatments, laboratory results, blood gas, respiratory, and hemodynamic parameters). As expected, the main limitation was the causal relationship between pathological conditions. In this context, authors reported undeniable usefulness of ChatGPT in summarizing medical information using technical languages, but they underlined the need to pay great attention to contents that required medical expertise, such as the identification of causal relationships among conditions [38]. In clinical practice, many significant medical records are very time-consuming. Quickly composing these documents may improve communication among healthcare centers. ChatGPT demonstrated a remarkable ability in compose patient's discharge summaries and operative notes [66]. These documents require expert revision, but the diffusion of these AI model language in medical centers

may improve the time to produce these medical records, providing quick and high-quality transition among healthcare centers [67]. Furthermore, ChatGPT's ability to admit and learn from mistakes is very promising [18]. For example, in the operative note for a patient with age-related macular degeneration, ChatGPT correctly adjusted anesthesia details associated with intraocular ranibizumab injection [67].

#### 4.2. Support in Diagnosis and Differential Diagnosis

In medical practice, AI technologies are entirely used to support the definition of diagnosis, prognosis, and assessment of therapeutic targets (i.e. in oncology it can provide treatment suggestions based on MRI radiomics and ageing-related diseases) [68–70]. Unlike specific AI technologies developed and trained for medical purposes, ChatGPT has been developed without specific medical training. It makes responses to user questions based on internet datasets, without distinguishing between reputable and non-reputable sources. For these reasons, blindly relying on ChatGPT suggestions is dangerous.

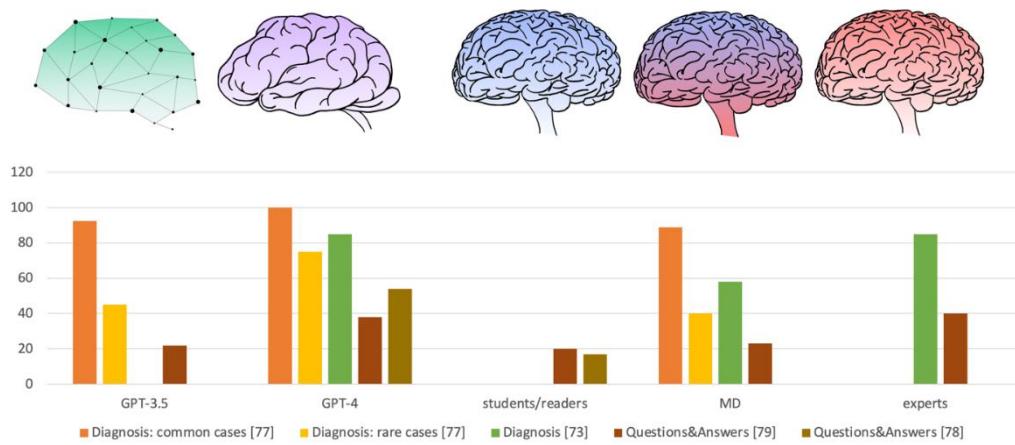
Dynamism is one of the most essential features of ChatGPT and other AI language models. In fact, while diagnostic tools, such as Isabel [20,71,72], contain clinical data for a considerable number of diseases, they are static and cannot evaluate some individual data. Typically, the case presentation is based on a list of signs and symptoms, making the submission of clinical cases inappropriate. The conversation model included in ChatGPT supports the dynamic presentation of the clinical case, improving efficiency and relevance. In a recent study on the accuracy of Isabel to find the correct diagnosis in ophthalmic patients, it did not perform as well as ChatGPT. In particular, Isabel provided the correct diagnosis within the first 10 results in 70% of patients, while ChatGPT reached 100% within the first ten differential [20]. Notably, the chatbot confirmed its results in rare and familial disorders, giving the correct diagnosis in both Behcet's Disease and AMD. Instead, Isabel misdiagnosed both cases, respectively as relapsing polychondritis and uveitis.

Even in infectious diseases, ChatGPT has been evaluated to test its ability in providing diagnostic or therapeutic advice correctly in non-chronic clinical cases. It shows a good but not optimal performance, reaching an average score of 2.8, where the rating spanned from 1 (poor, incorrect advice) to 5 (excellent, fully corresponding with the advice of infectious disease and clinical microbiology specialists) [10].

Interestingly, when compared with specialist and non-specialist physicians, ChatGPT performances were surprising. In a neurology study, 200 synthetic clinical neurological cases were fed into ChatGPT, asking for the most probable and five most probable diagnoses. Results were compared with answers from 12 medical doctors (6 "experts", neurology specialists, and six general medical doctors). The first (most probable) diagnosis given by ChatGPT was correct in 68.5% of cases. This result is surprising because medical doctors group reaches only 57.08% ( $\pm 4.8\%$ ), while the expert group reaches 81.58% ( $\pm 2.34\%$ ). As expected, some clinical cases were misdiagnosed. In particular, 10 cases were classified as "unsolved" because all experts failed to provide the correct diagnosis [73]. Notably, among these cases, there are some genetic and rare disorders. The accuracy of the chatbot in recognizing rare genetic disorders or correctly answering about them is limited. For example, when asking about relationship between mutations in SCN9A and autosomal dominant epilepsy, ChatGPT incorrectly gave positive responses [74–76]. This evidence strongly suggested that ChatGPT may be misleading in evaluating rare disorders, which should also be assessed by a geneticist and/or other specific clinical tools. The ability to detect the correct diagnosis of ChatGPT (GPT-3.5 and GPT-4) was very weak for rare disorders, while it was acceptable for common diseases [77]. Compared with medical doctors, ChatGPT reached the correct diagnosis in first three responses in over 90% of typical cases, which is quite similar to results from medical doctors (in over 90% of typical cases they identified the correct diagnosis in the first two responses). Among rare disorders the performance of both ChatGPT and medical doctors decreased: GPT-3.5 reached 60% within the first ten responses, GPT-4 reached 90% within the first 8-10 responses, and medical doctors solved 30-40% of cases with their first suggested diagnosis, for 2 of them the diagnostic accuracy increases to 50% within the first two suggested diagnoses [77].

The differences significantly increased when ChatGPT diagnostic performances were compared to a less experienced group, as medical-journal readers. A study on clinical case challenges from New England Journal of Medicine revealed that GPT-4 provided correct diagnosis for 22 of 38 clinical cases (57%), whereas the medical-journal readers chose proper diagnosis among six provided options for 36% of cases [78].

ChatGPT is continuously evolving, and the evaluation of ChatGPT4 reliability showed a significant improvement if compared with ChatGPT 3.5. A recent analysis reported a relative safety of output information in more than 90% of responses (91% for GPT-3.5, and 93% for GPT-4), categorized out of the group of “so incorrect as to cause patient harm”. Unfortunately, the concordance between ChatGPT results and physician answers remains low (21% for GPT-3.5 and 41% for GPT-4) [40]. In a more specific context (neurosurgery), GPT-4 confirmed its better performance when compared with GPT-3.5. The percentage of consistent responses, according to guidelines and expert opinions, was 42% for GPT-3.5 and 72% for GPT-4.0 [79] (Figure 2)



**Figure 2.** Comparative Performance Analysis. (Left to right: GPT-3.5, GPT-4, Medical Journal Readers/Students, Medical Doctors (MDs), Experts).

#### 4.2. Support in Treatment Advice

A study on ChatGPT’s reliability in reporting cancer therapies for solid tumors, reaches high scores. In particular, 51 clinical situations and 30 distinct subtypes of solid tumor were posed to ChatGPT, asking for therapies that can be used as first-line treatment. Chatbot results were compared with NCCN (National Comprehensive Cancer Network) guidelines. The accordance among responses was measured by searching ChatGPT suggested therapies among first level therapies listed in NCCN guidelines. In all circumstances, ChatGPT named therapies that may be used as first line treatment for advanced or metastatic solid tumors. One point to consider is that some of the responses included alternate or preliminary drug names (i.e. Blu-667 for pralsetinib). Another important information is that the study evaluates general NCCN guidelines. Instead, recommendations may individually vary among patients [80–82]. In a consequential survey of same field, but with different criteria for chatbot response evaluation, authors revealed that one third of treatment recommendations also included one or more drugs non-concordant with NCCN guidelines. Moreover, ChatGPT recommendations changed depending on the question [83]. Similarly, in neuro-oncology, the ability of ChatGPT in adjuvant therapy decision-making was evaluated for glioma patients. Of them, 80% of patients were diagnosed with glioblastoma, a rare malignant brain tumor [84,85], and 20% has low-grade gliomas. Interestingly, given the patient summary, the chatbot correctly recognized and classified the tumors as glioma, suggesting a tumor type (glioblastoma, grade II or III astrocytoma, etc.). While ChatGPT reported the need to modify treatment according to

patient's preferences and functional status, no alternative therapies were listed, nor alternative diagnosis [86]. Otherwise, treatment suggestions were evaluated positively by a team of experts [86].

In summary, ChatGPT's performance evaluation depended on the extensiveness of the knowledge and their availability in the web. In particular, the assessment of patient functional status was defined moderate, maybe because of the small number of clinical trials available online [20]. In this scenario, it seems that the naïve ChatGPT can support the multidisciplinary activities in neuro-oncology, but it requires a complete training. For these reasons, Guo and coworkers created a trained version of ChatGPT, called neuroGPT-X, and evaluated it compared with naïve ChatGPT and leading neurosurgical experts worldwide. Although its ability to support the neurosurgeons, the human expert's evaluation remains necessary, mainly to ensure safety and reliability of chatbot responses [87].

A recent study compared neurosurgery knowledge among chatbots (GPT-3.5 and 4.0) and neurosurgeons with different seniority levels. Fifty questions about neurosurgical treatments were submitted to chatbot and neurosurgeons. The answers were evaluated by a team of senior neurosurgeons, that judged them as "consistent" or "inconsistent" with recommendations available in guidelines and evidence-based knowledge. The ability of GPT-3.5 was similar to that of neurosurgeons with low seniority, while GPT-4.0 ability was similar to that of neurosurgeons with high seniority [79].

Another study evaluated the reliability of ChatGPT to report potential dangers associated to drug-drug interactions [88]. Juhi and coworkers asked the chatbot "can I take A and B together?", where A and B represent two drug names, and successively "why should I not take A and B together?". They tested 40 drug-drug interaction pairs and only one interaction was unrecognized as dangerous. Interestingly, when the authors submitted the second question to the chatbot ("why should I not take A and B together?"), it corrected the first wrong answer. As expected, answers to the second question were less precise, describing molecular pathways, but frequently with no conclusive facts supporting the drug-drug interaction [88]. In this context, ChatGPT confirmed its ability to recognize mistakes [18].

## 5. Discussion

In this review, medical applications of ChatGPT have been resumed, from the support in drafting medical records to more specialistic purposes, as medical diagnosis, differential diagnosis, and treatment. In all its applications, the chatbot reported good results, still requiring close human supervision. Besides "artificial hallucinations" that may compromise the quality of medical records, many other inaccuracies have been noted. In particular, in the diagnostic path it suffers from the absence of a dialogue, although the language simulates it. The chatbot bases its clinical assessment on information provided without the ability of exploring the clinical context posing further questions. In this scenario GPT is far from replacing human qualities in medical practice [89].

Furthermore, although GPT demonstrated a good even if non optimal performance in recognizing rare disorders, the chatbot reached better results than other diagnostic tools [20]. In this scenario, it is predictable to expect that a short training of GPT might improve its performance even in rare disorders. One of the main problems in diagnosing and treating patients with rare disorders is the lack of expert physicians. In particular, to recognize a patient potentially affected by a rare disorder, it is necessary that a medical doctor identifies clinical signs and symptoms and refers the patient to a specialistic healthcare unit. Unfortunately, the number of medical doctors able to correctly identify these signs and symptoms is limited. It is evident from the significant number of rare disorders that are still undiagnosed [90]. Adequate training in a user-friendly platform such as ChatGPT might improve the ability of medical doctors to recognize signs and symptoms associated with a rare disorder. Similarly, the chatbot may simplify the evaluation of familial history, making easier the detection of familial genetic disorders. ChatGPT interrogation could be a valid support to screen families, identify families requiring a more precise genetic evaluation. To date, the unique survey conducted on potential applications of ChatGPT among genetic counsellors, reported a skeptical attitude. The most common concern about using ChatGPT was the risk of incorrect answers

(82.2%) [91]. As seen in other medical contexts, ChatGPT is used in clinical practice by almost one of three interrogated genetic counsellors for drafting medical documentation (consult notes, result letters, and letters of medical necessity). A tiny percentage of genetic counsellors reported using the chatbot for clinical information on rare disorders (14.1%) or for differential diagnosis (8.6%).

As evident, the ChatGPT consultation cannot substitute for the medical doctor evaluation. One of the most critical shared limits of ChatGPT is the inability to correctly investigate the patient's clinical features. It may seem obvious, but in clinical practice the asking process is fundamental in the diagnostic path. ChatGPT has been tested with clinical cases ready for evaluation. Conversely, in the clinical practice, patients manifest one or more signs/symptoms of disease and often the physician have to ask for other sings/symptoms to complete the clinical picture of the disease [77,93]. The ability to ask the right question is one of the main ChatGPT limits, requiring the medical intervention to translate the clinical phenotype to a case vignette ready for the chatbot evaluation [93].

Nevertheless, it is essential to note that ChatGPT does not claim to be a doctor, nor to replace him. In many cases, the chatbot answers with a disclaimer about "I am not a doctor, but" [19], in other cases it advises the patient to consult professional healthcare for evaluation on the necessity of medication. For example, in an assessment of theoretical psychological cases with sleep disorders ChatGPT suggested a treatment plan based on non-pharmacological interventions [92]. This answer is not only in line with current guidelines [94] but it is also safe for the patients, because the suggestion of a pharmacological treatment was strictly correlated to the necessity of a medical consultation [92].

The immediate application of AI represents a significant technological advancement in the statistical analysis of merged data sets. Our prior experience in biostatistics within the field of ophthalmic genomics indicates that similar outcomes could have been achieved more rapidly and with smaller cohorts of case-control samples [95–97]. This suggests that AI not only enhances efficiency but also reduces the need for extensive sample sizes.

In summary, the evaluation of ChatGPT into the clinical practice of rare disorders demonstrated a good potential. In detail, research and diagnostic applications of ChatGPT as a support of healthcare professionals retrieved excellent results, with similar accuracy for common and rare disorders. In this context, it is well known that rare disorders are less frequent than patients affected by common disorders. Notably, internet information did not follow this tendency. In particular, web spaces about rare disorders are very widespread. Online available sources create chatbot datasets, so rare disorders might be equally represented in these datasets when compared with common disorders. It is a hypothesis that might explain this good accuracy of ChatGPT even in uncommon disease. Anyway, whatever the reasons at the bases of these GPT performances, it is expected that specific training on rare disorders, from clinical presentation, to progression, available treatments, and risks for family members, will improve the ability of the chatbot to support not only medical doctors, but also experts. The widespread adoption and tailored training of ChatGPT could significantly support medical doctors' ability to identify patients with rare disorders, potentially reducing the number of undiagnosed cases. ChatGPT may become a routine tool in clinical practice, not as a substitute for medical evaluation but as an aid in assessing patients. However, it is essential to emphasize that while ChatGPT can offer valuable insights, its recommendations and corrections must be evaluated in conjunction with expert human judgment.

In the evolving landscape of healthcare, there is a pressing need to integrate specialized training programs into the medical education curriculum, aimed at equipping the forthcoming generations of medical professionals with the proficiency required to utilize Artificial Intelligence (AI) programs effectively. These AI programs, when translated as routine clinical tools, have the potential to significantly improve the efficiency, accuracy, and cost-effectiveness of medical services.

A complete familiarity and competence in leveraging AI technologies can ensure that future doctors are well-trained to navigate the complexities of modern healthcare optimizing patient care. This strategic introduction of AI into medical training not only aligns with the technological advancements of our time but also underscores a commitment to advancing the quality and sustainability of healthcare practices.

**Author Contributions:** Conceptualization, S.Z. and E.G.; methodology, S.Z.; validation, C.P., D.M. and G.C.; resources, S.Z. and E.G.; data curation, S.Z.; writing—original draft preparation, S.Z.; writing—review and editing, R.C., C.S., C.C.; visualization, S.Z.; supervision, E.G.; project administration, C.C.; funding acquisition, C.C. and E.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by Ministry of Italian Health, grant number RF19.12.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Xue, V.W.; Lei, P.; Cho, W.C. The potential impact of ChatGPT in clinical and translational medicine. *Clin Transl Med.* **2023**, *13*(3), e1216. doi: 10.1002/ctm2.1216.
2. Stokel-Walker, C.; Van Noorden, R. What ChatGPT and generative AI mean for science. *Nature.* **2023**, *614*(7947), 214–216. doi:10.1038/d41586-023-00340-6
3. No authors listed. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature.* **2023**, *613*(7945), 612. doi:10.1038/d41586-023-00191-1
4. Shen, Y.; Heacock, L.; Elias, J.; et al. ChatGPT and other large language models are double-edged swords. *Radiology.* **2023**, *230163*. doi:10.1148/radiol.230163
5. The Lancet Digital Health. ChatGPT: friend or foe? *Lancet Digit Health.* **2023**, *5*(3), E102. doi:10.1016/S2589-7500(23)00023-7
6. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; et al. Language models are few-shot learners. *Adv Neural Inform Proc Syst.* **2020**, *33*, 1877–901
7. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature.* **2015**, *521*(7553), 436–44
8. Hosseini, M.; Gao, C.A.; Liebovitz, D.M.; Carvalho, A.M.; Ahmad, F.S.; Luo Y; et al. An exploratory survey about using ChatGPT in education, healthcare, and research. *PLoS One.* **2023**, *18*(10), e0292216. doi: 10.1371/journal.pone.0292216.
9. Yadava, O.P. ChatGPT-a foe or an ally? *Indian J Thorac Cardiovasc Surg.* **2023**, *39*(3), 217–221. doi: 10.1007/s12055-023-01507-6.
10. Sarink, M.J.; Bakker, I.L.; Anas, A.A.; Yusuf, E. A study on the performance of ChatGPT in infectious diseases clinical consultation. *Clin Microbiol Infect.* **2023**, *29*(8), 1088–1089. doi: 10.1016/j.cmi.2023.05.017.
11. Bhattacharya, K.; Bhattacharya, A.S.; Bhattacharya, N.; Yagnik, V.D.; Garg, P.; Kumar, S. ChatGPT in Surgical Practice—a New Kid on the Block. *Indian J Surg.* **2023**, *22*, 1–4. doi:10.1007/s12262-023-03727-x
12. Cheng, K.; Li, Z.; He, Y.; Guo, Q.; Lu, Y.; Gu, S.; Wu H. Potential Use of Artificial Intelligence in Infectious Disease: Take ChatGPT as an Example. *Ann Biomed Eng.* **2023**, *51*(6), 1130–1135. doi: 10.1007/s10439-023-03203-3.
13. AdvancedAds. Gamechanger: ChatGPT provides current data. Not limited to before September 2021 Available online: <https://wpadvancedads.com/chatgpt-provides-current-data/#:~:text=Gamechanger%3A%20ChatGPT%20provides%20current%20data,limited%20to%20before%20September%202021&text=Now%20that%20it%20has%20been,to%20use%20it%20more%20efficiently>. (Accessed on 5<sup>th</sup> February 2024)
14. Twitter. OpenAI post dated 27 set 2023. Available online: [https://twitter.com/OpenAI/status/1707077710047216095?t=oeBD2HTJg2HvQeKF6v-MUg&s=1\[%E2%80%A6\]d=IwAR04RwUXxRfjOVXGo4L-15RH2NDt7SC907QbydJlu2jPmZ64H\\_eqVsbrF4](https://twitter.com/OpenAI/status/1707077710047216095?t=oeBD2HTJg2HvQeKF6v-MUg&s=1[%E2%80%A6]d=IwAR04RwUXxRfjOVXGo4L-15RH2NDt7SC907QbydJlu2jPmZ64H_eqVsbrF4) (Accessed on 5<sup>th</sup> February 2024)
15. Dave, T.; Athaluri, S.A.; Singh, S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell.* **2023**, *6*, 1169595. doi: 10.3389/frai.2023.1169595.
16. Nguyen, Y.; Costedoat-Chalumeau, N. Les intelligences artificielles conversationnelles en médecine interne : l'exemple de l'hydroxychloroquine selon ChatGPT [Artificial intelligence and internal medicine: The example of hydroxychloroquine according to ChatGPT]. *Rev Med Interne.* **2023**, *44*(5), 218–226. French. doi: 10.1016/j.revmed.2023.03.017.

17. Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepan˜ o, C.; et al., Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit. Health* **2023**, *2*, e0000198. doi:10.1371/journal.pdig.0000198
18. Lee, P.; Bubeck, S.; Petro, J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med* **2023**, *388*:1233-1239. doi:10.1056/NEJMsr2214184.
19. Darkhabani, M.; Alrifai, M.A.; Elsalti, A.; Dvir, Y.M.; Mahroum, N. ChatGPT and autoimmunity - A new weapon in the battlefield of knowledge. *Autoimmun. Rev.* **2023**, *22*(8):103360. doi:10.1016/j.autrev.2023.103360.
20. Balas, M.; Ing, E.B. Conversational AI Models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel Pro Differential Diagnosis Generator. *JFO Open Ophthalmology*. **2023**, *1*: 100005. doi:10.1016/j.jfop.2023.100005
21. Uz, C.; Umay, E. "Dr ChatGPT": Is it a reliable and useful source for common rheumatic diseases? *Int J Rheum Dis.* **2023**, *26*(7):1343-1349. doi: 10.1111/1756-185X.14749.
22. Benichou, L. ChatGPT. The role of using ChatGPT AI in writing medical scientific articles. *J Stomatol Oral Maxillofac Surg.* **2023**, *124*(5), 101456. doi: 10.1016/j.jormas.2023.101456.
23. Curtis, N.; ChatGPT. To ChatGPT or not to ChatGPT? The Impact of Artificial Intelligence on Academic Publishing. *Pediatr Infect Dis J.* **2023**, *42*(4), 275. doi: 10.1097/INF.0000000000003852.
24. King, M.R.; chatGPT. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cell Mol Bioeng.* **2023**, *16*(1):1-2. doi: 10.1007/s12195-022-00754-8.
25. ChatGPT Generative Pre-trained Transformer; Zhavoronkov, A. Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective. *Oncoscience.* **2022**, *9*, 82-84. doi: 10.18632/oncoscience.571.
26. Mijwil, M.M.; Aljanabi, M.; ChatGPT. Towards Artificial Intelligence-Based Cybersecurity: The Practices and ChatGPT Generated Ways to Combat Cybercrime. *Iraqi Journal for Computer Science and Mathematics*, **2023**, *4*(1), 65-70
27. Aljanabi, M.; Ghazi, M.; Ali, A.H.; Abed, S.A.; ChatGpt. ChatGpt: Open Possibilities. *Iraqi Journal for Computer Science and Mathematics*, **2023**, *4*(1), 62-64
28. O'Connor, S. ChatGPT Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice*, **2023**, *66*, 103537
29. Conroy, G. Scientists used ChatGPT to generate an entire paper from scratch - but is it any good? *Nature*. **2023**, *619*(7970), 443-444. doi: 10.1038/d41586-023-02218-z. PMID: 37419951.
30. Sciencedirect: guide for authors. Available online: <https://www.sciencedirect.com/journal/resources-policy/publish/guide-for-authors> (Accessed on 5<sup>th</sup> February 2024)
31. Cell: guide for authors. Available online: <https://www.cell.com/device/authors> (Accessed on 5<sup>th</sup> February 2024)
32. Elsevier: guide for authors. Available online: <https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier> (Accessed on 5<sup>th</sup> February 2024)
33. Stokel-Walker, C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature*. **2023**, *613*(7945), 620-621. doi: 10.1038/d41586-023-00107-z.
34. Gao, C.A.; Howard, F.M.; Markov, N.S. et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digit. Med.* **2023**, *6*, 75. doi:10.1038/s41746-023-00819-6
35. Wen, J.; Wang W. The future of ChatGPT in academic research and publishing: A commentary for clinical and translational medicine. *Clin Transl Med.* **2023**, *13*(3):e1207. doi: 10.1002/ctm2.1207.
36. Else, H. Abstracts written by ChatGPT fool scientists. *Nature* **2023**, *613*, 423. doi: 10.1038/d41586-023-00056-7
37. Aydin, Ö.; Karaarslan, E. OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. In Ö. Aydin (Ed.), Emerging Computer Technologies 2, 2022, pp. 22-31. İzmir Akademi Derneği.
38. Cascella, M.; Montomoli, J.; Bellini, V.; Bignami, E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *J Med Syst.* **2023**, *47*(1), 33. doi: 10.1007/s10916-023-01925-4.
39. Blanco-González, A.; Cabezón, A.; Seco-González, A.; Conde-Torres, D.; Antelo-Riveiro, P.; Piñeiro, Á.; García-Fandino R. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals (Basel)*. **2023**, *16*(6):891. doi: 10.3390/ph16060891.

40. Mello, M.M.; Guha, N. ChatGPT and Physicians' Malpractice Risk. *JAMA Health Forum*. **2023**, *4*(5):e231938. doi: 10.1001/jamahealthforum.2023.1938.
41. Dunn, C.; Hunter, J.; Steffes, W.; Whitney, Z.; Foss, M.; Mammino, J.; et al. Artificial intelligence-derived dermatology case reports are indistinguishable from those written by humans: A single-blinded observer study. *J Am Acad Dermatol*. **2023**;89(2):388-390. doi: 10.1016/j.jaad.2023.04.005.
42. Goddard, J. Hallucinations in ChatGPT: A Cautionary Tale for Biomedical Researchers. *Am J Med*. **2023**, *136*(11):1059-1060. doi: 10.1016/j.amjmed.2023.06.012.
43. Alkaissi, H.; McFarlane, S.I. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*. **2023**, *15*(2):e35179. doi: 10.7759/cureus.35179.
44. Hirschhorn, R.; Reuser, A.J. Glycogen storage disease type II: acid alpha-glucosidase (acid maltase) deficiency. In: Scriver CR, Beaudet A, Sly WS, Valle D, eds. *The Metabolic and Molecular Bases of Inherited Disease*. New York, NY: McGraw-Hill; 2001:3389-420.
45. Gravel, J.; D'Amours-Gravel, M.; Osmanliu, E. Learning to Fake It: Limited Responses and Fabricated References Provided by ChatGPT for Medical Questions. *Mayo Clinic Proceedings: Digital Health*, **2023**, *1* (3), 226-234, doi: 10.1016/j.mcpdig.2023.05.004.
46. Day, T. A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT, *The Professional Geographer*, **2023**, *75*(6), 1024-1027, doi: 10.1080/00330124.2023.2190373
47. Javid, M.; Reddiboina, M.; Bhandari, M. Emergence of artificial generative intelligence and its potential impact on urology. *Can J Urol*. **2023**, *30*(4), 11588-11598.
48. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; et al. Survey of hallucination in natural language generation. *ACM Comput Surv* **2022**, *55*, 1-38. doi:10.1145/3571730
49. Deveci, C.D.; Baker, J.J.; Sikander, B.; Rosenberg, J. A comparison of cover letters written by ChatGPT-4 or humans. *Dan Med J*. **2023**, *70*(12), A06230412.
50. Cox, L.A. Jr. An AI assistant to help review and improve causal reasoning in epidemiological documents. *Glob Epidemiol*. **2023**, *7*, 100130. doi:10.1016/j.gloepi.2023.100130.
51. Alsadhan, A.; Al-Anezi, F.; Almohanna, A.; Alnaim, N.; Alzahrani, H.; Shinawi, R.; et al. The opportunities and challenges of adopting ChatGPT in medical research. *Front Med (Lausanne)*. **2023**, *10*, 1259640. doi: 10.3389/fmed.2023.1259640.
52. Johnson, S.B.; Parsons, M.; Dorff, T.; et al. Cancer misinformation and harmful information on Facebook and other social media: a brief report. *J Natl Cancer Inst*. **2022**, *114*(7):1036-1039. doi: 10.1093/jnci/djab141.
53. Johnson, S.B.; King, A.J.; Warner, E.L.; Aneja, S.; Kann, B.H.; Bylund, C.L. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectr*. **2023**, *7*(2):pkad015. doi: 10.1093/jncics/pkad015.
54. Kim, J-H. Search for Medical Information and Treatment Options for Musculoskeletal Disorders through an Artificial Intelligence Chatbot: Focusing on Shoulder Impingement Syndrome. Preprint available on <https://www.medrxiv.org/content/10.1101/2022.12.16.22283512v2.full-text> doi: 10.1101/2022.12.16.22283512 (Accessed on 26th February 2024)
55. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. *J Med Internet Res*. **2019**, *21*:e13873. doi: 10.2196/13873
56. Nov, O.; Singh, N.; Mann, D. Putting ChatGPT's Medical Advice to the (Turing) Test: Survey Study. *JMIR Med Educ*. **2023**, *9*:e46939. doi: 10.2196/46939.
57. Nastasi, A.J.; Courtright, K.R.; Halpern, S.D.; Weissman, G.E. A vignette-based evaluation of ChatGPT's ability to provide appropriate and equitable medical advice across care contexts. *Sci Rep*. **2023**, *13*(1), 17885. doi: 10.1038/s41598-023-45223-y.
58. Harskamp, R.E.; De Clercq, L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). *Acta Cardiol*. **2024**, 1-9. doi: 10.1080/00015385.2024.2303528.
59. Johnson, D.; Goodman, R.; Patrinely, J.; Stone, C.; Zimmerman, E.; Donald, R.; et al. Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model. *Res Sq [Preprint]*. **2023**, rs.3.rs-2566942. doi: 10.21203/rs.3.rs-2566942/v1.
60. Fitzpatrick, K.; Darcy, A.; Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health*. **2017**, *4*, e7785. doi: 10.2196/mental.7785

61. Pham, K.; Nabizadeh, A.; Selek, S. Artificial intelligence and chatbots in psychiatry. *Psychiatr Q.* **2022**, *93*, 249–53. doi: 10.1007/s11126-021-09945-w
62. Gaffney, H.; Mansell, W.; Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR Ment Health.* **2019**, *6*, e14166. doi: 10.2196/14166
63. Coombs, N.; Meriwether, W.; Caringi, J.; Newcomer S. Barriers to healthcare access among U.S. adults with mental health challenges: a population-based study. *SSM Popul Health.* **2021**, *15*, 100847. doi: 10.1016/j.ssmph.2021.100847
64. Allen, S. Artificial intelligence and the future of psychiatry. *IEEE Pulse.* **2020**, *11*(3), 2–6. doi:10.1109/MPULS.2020.2993657
65. Vaeth, S.; Andersen, H.; Christensen, R.; Jensen, U.B. A Search for Undiagnosed Charcot-Marie-Tooth Disease Among Patients Registered with Unspecified Polyneuropathy in the Danish National Patient Registry. *Clin Epidemiol.* **2021**, *13*, 113–120. doi: 10.2147/CLEP.S292676.
66. Patel, S.B.; Lam, K. ChatGPT: the future of discharge summaries? *Lancet Digit Health* **2023**, *5*, e107–e108. doi: 10.1016/S2589-7500(23)00021-3.
67. Singh, S.; Djalilian, A.; Ali, M.J. ChatGPT and Ophthalmology: Exploring Its Potential with Discharge Summaries and Operative Notes. *Semin Ophthalmol.* **2023**, *38*(5), 503–507. doi: 10.1080/08820538.2023.2209166.
68. Horvat, N.; Veeraraghavan, H.; Nahas, C.S.R.; et al. Combined artificial intelligence and radiologist model for predicting rectal cancer treatment response from magnetic resonance imaging: an external validation study. *Abdom Radiol (NY).* **2022**, *47*(8), 2770–2782. doi:10.1007/s00261-022-03572-8
69. Pun, F.W.; Leung, G.H.D.; Leung, H.W.; et al. Hallmarks of agingbased dual-purpose disease and age-associated targets predicted using PandaOmics AI-powered discovery engine. *Aging (Albany NY).* **2022**, *14*(6), 2475–2506. 10.18632/aging.203960
70. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD. Evaluating ChatGPT as an adjunct for radiologic decision-making. *medRxiv.* **2023**. doi:10.1101/2023.02.02.23285399
71. Sibbald, M.; Monteiro, S.; Sherbino, J.; LoGiudice, A.; Friedman, C.; Norman G. Should electronic differential diagnosis support be used early or late in the diagnostic process?. A multicentre experimental study of Isabel. *BMJ Qual Safe* **2022**, *31*(6), 426–33.
72. Riches, N.; Panagioti, M.; Alam, R.; Cheraghi-Sohi, S.; Campbell, S.; Esmail, A.; et al. The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis. *PLoS One* **2016**, *11*(3), e0148991.
73. NÓGRÁDI, B.; POLGÁR, T.F.; MESZLÉNYI, V.; KÁDÁR, Z.; HERTELENDY, P.; CSÁTI, A.; et al. ChatGPT M.D.: Is There Any Room for Generative AI in Neurology and Other Medical Areas?. Available at SSRN: <https://ssrn.com/abstract=4372965> or <http://dx.doi.org/10.2139/ssrn.4372965>
74. Boßelmann CM, Leu C, Lal D. Are AI language models such as ChatGPT ready to improve the care of individuals with epilepsy? *Epilepsia.* 2023 May;64(5):1195–1199. doi: 10.1111/epi.17570. Epub 2023 Mar 13. PMID: 36869421.
75. Brunklaus, A. No evidence that SCN9A variants are associated with epilepsy. *Seizure.* **2021**, *91*, 172–3.
76. Curation Results for Gene-Disease Validity. [Accessed on Dec 2023]. Available from: <https://search.clinicalgenome.org/> kb/gene-validity/CGGV:assertion\_72a91ef6-e052-44a4-b14e-6a5ba93393ff-2021-03-09T163649.218Z
77. Mehnert, L.; Gruarin, S.; Vasileva, M.; Knapp, B. Chat GPT as a medical doctor? A diagnostic accuracy study on common and rare diseases. *MedRxiv* **2023**, doi: 10.1101/2023.04.20.23288859
78. Eriksen, A.V.; Möller, S.; Jesper, R. Use of GPT-4 to Diagnose Complex Clinical Cases *NEJM AI* **2023**, *1* (1) doi:10.1056/AIp2300031
79. Liu, J.; Zheng, J.; Cai, X.; Wu, D.; Yin, C. A descriptive study based on the comparison of ChatGPT and evidence-based neurosurgeons. *iScience.* **2023**, *26*(9), 107590. doi: 10.1016/j.isci.2023.107590.
80. Schulte, B. Capacity of ChatGPT to Identify Guideline-Based Treatments for Advanced Solid Tumors. *Cureus.* **2023**;15(4), e37938. doi: 10.7759/cureus.37938.
81. American Society of Clinical Oncology Guidelines. Available online: <https://www.esmo.org/guidelines>. Accessed: December, 2023
82. European Society of Medical Oncology Clinical Practice Guidelines . Available online: <https://www.esmo.org/guidelines>. Accessed: December, 2023

83. Chen, S.; Kann, B.H.; Foote, M.B.; Aerts, H.J.W.L.; Savova, G.K.; Mak, R.H.; Bitterman, D.S. The utility of ChatGPT for cancer treatment information. *medRxiv* **2023**, 03.16.23287316; doi: <https://doi.org/10.1101/2023.03.16.23287316>

84. McGowan, E.; Sanjak, J.; Mathé, E.A.; Zhu, Q. Integrative rare disease biomedical profile-based network supporting drug repurposing or repositioning, a case study of glioblastoma. *Orphanet J Rare Dis.* **2023**, 18(1), 301. doi: 10.1186/s13023-023-02876-2.

85. Glioblastoma. <https://rarediseases.info.nih.gov/diseases/2491/glioblastoma>. Accessed February 21, 2024

86. Haemmerli, J.; Sveikata, L.; Nouri, A.; May, A.; Egervari, K.; Freyschlag, C.; et al. ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board? *BMJ Health Care Inform.* **2023**, 30(1), e100775. doi: 10.1136/bmjhci-2023-100775.

87. Guo, E.; Gupta, M.; Sinha, S.; Rössler, K.; Tatagiba, M.; Akagami, R.; et al. NeuroGPT-X: Towards an Accountable Expert Opinion Tool for Vestibular Schwannoma, *MedRxiv Mendeley Data V1* **2023**, , doi: 10.17632/b9mck42r35.1

88. Juhi, A.; Pipil, N.; Santra, S.; Mondal S.; Behera, J.K.; Mondal, H. The Capability of ChatGPT in Predicting and Explaining Common Drug-Drug Interactions. *Cureus.* **2023**, 15(3), e36272. doi: 10.7759/cureus.36272.

89. Tripathi, M.; Chandra, S.P. ChatGPT: a threat to the natural wisdom from artificial intelligence. *Neurol India.* **2023**, 71(3), 416-7. doi: 10.4103/0028-3886.378687

90. Vaeth, S.; Andersen, H.; Christensen, R.; Jensen, U.B. A Search for Undiagnosed Charcot-Marie-Tooth Disease Among Patients Registered with Unspecified Polyneuropathy in the Danish National Patient Registry. *Clin Epidemiol.* **2021**, 13, 113-120. doi: 10.2147/CLEP.S292676.

91. Ahimaz, P.; Bergner, A.L.; Florido, M.E.; Harkavy, N.; Bhattacharyya, S. Genetic counselors' utilization of ChatGPT in professional practice: A cross-sectional study. *Am J Med Genet A.* **2023** doi: 10.1002/ajmg.a.63493.

92. Dergaa, I.; Fekih-Romdhane, F.; Hallit, S.; Loch, A.A.; Glenn, J.M.; Fessi, M.S.; et al. ChatGPT is not ready yet for use in providing mental health assessment and interventions. *Front Psychiatry.* **2024**, 14, 1277756. doi: 10.3389/fpsyg.2023.1277756.

93. Wa M. Evidence-based clinical practice: asking focused questions (PICO). *Optom Vis Sci.* **2016**, 93, 1187-8. doi: 10.1097/OPX.0000000000000929

94. Rios, P.; Cardoso, R.; Morra, D.; Nincic, V.; Goodarzi, Z.; Farah, B.; et al. Comparative effectiveness and safety of pharmacological and non-pharmacological interventions for insomnia: an overview of reviews. *Syst Rev.* **2019**, 8, 281. doi: 10.1186/s13643-019-1191-x

95. Casella, R.; Strafella, C.; Longo, G.; Ragazzo, M.; Manzo, L.; De Felici, C.; et al. Uncovering genetic and non-genetic biomarkers specific for exudative age-related macular degeneration: significant association of twelve variants. *Oncotarget.* **2017**, 9(8), 7812-7821. doi: 10.18632/oncotarget.23241.

96. Ricci, F.; Zampatti, S.; D'Abbruzzi, F.; Missiroli, F.; Martone, C.; Lepre, T.; et al. Typing of ARMS2 and CFH in age-related macular degeneration: case-control study and assessment of frequency in the Italian population. *Arch Ophthalmol.* **2009**, 127(10), 1368-72. doi: 10.1001/archophthalmol.2009.237.

97. Ricci, F.; Staurenghi, G.; Lepre, T.; Missiroli, F.; Zampatti, S.; Casella, R.; et al. Haplotypes in IL-8 Gene Are Associated to Age-Related Macular Degeneration: A Case-Control Study. *PLoS One.* **2013**, 8(6), e66978. doi: 10.1371/journal.pone.0066978.

**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.