

Review

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

Stefania Zampatti ¹, Cristina Peconi ¹, Domenica Megalizzi ^{1,2}, Giulia Calvino ^{1,2}, Giulia Trastulli ^{1,3}, Raffaella Cascella ^{1,4}, Claudia Strafella ¹, Carlo Caltagirone ⁵ and Emiliano Giardina ^{1,6,*}

¹ Genomic Medicine Laboratory UILDM, IRCCS Santa Lucia Foundation, 00179 Rome, Italy; s.zampatti@hsantalucia.it (S.Z.)

² Department of Science, Roma Tre University, 00146 Rome, Italy

³ Department of System Medicine, Tor Vergata University, 00133 Rome, Italy

⁴ Department of Chemical-Toxicological and Pharmacological Evaluation of Drugs, Catholic University Our Lady of Good Counsel, 1000 Tirana, Albania

⁵ Department of Clinical and Behavioral Neurology, IRCCS Fondazione Santa Lucia, 00179 Rome, Italy; c.caltagirone@hsantalucia.it

⁶ 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 datasets 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



Citation: Zampatti, S.; Peconi, C.; Megalizzi, D.; Calvino, G.; Trastulli, G.; Cascella, R.; Strafella, C.; Caltagirone, C.; Giardina, E. Innovations in Medicine: Exploring ChatGPT's Impact on Rare Disorder Management. *Genes* **2024**, *15*, 421. <https://doi.org/10.3390/genes15040421>

Academic Editor: Stefano Lonardi

Received: 29 February 2024

Revised: 25 March 2024

Accepted: 26 March 2024

Published: 28 March 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) 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. One of the significant advancements in artificial intelligence is generative AI (GAI), which can be characterized as a system capable of creating various media outputs, such as images, text, and other forms, based on human input [1]. Specifically, GAI models are adept at producing diverse media, including text, images, audio, video, and 3D models, in response to user queries. This cutting-edge technology excels at identifying patterns and analyzing existing datasets to produce outputs that are not only realistic but also consistent with the characteristics of the input. GAI models employ a variety of machine learning techniques, such as deep learning (DL), natural language processing (NLP), and neural networks. Notable GAI systems include ChatGPT, Dall-E, Gemini, and Midjourney. In the realm of healthcare, GAI has the potential to enhance various activities; for example,

DALL-E can analyze patient medical images to aid in diagnosis [2]. Similarly, Gemini AI can process images, even with low performance at the moment [3]. Some advanced GAI models are large language models (LLMs) that have been developed to understand and generate human-like languages. The LLMs analyze text using deep learning techniques and generate accurate responses. Among LLMs trained for scientific purposes, there is BioMedLM, a language model developed by Stanford CRFM in partnership with MosaicML. It is specifically trained on biomedical literature (PubMed abstracts and papers) [4] and has demonstrated good performance in biomedicine [5]. Presently, ChatGPT stands out as one of the most prominent LLMs. It uses 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 [6–10]. 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 [11]. Its deep learning basis enables the model to identify patterns in large data volumes, essential for producing accurate predictions and executing specific tasks [12]. 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 achieving broad acceptance and curiosity across the general public [13]. 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 [14,15].

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 [16]. 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 [17]. 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 [18,19]. 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, repetitive questioning at different times might result in different answers with some critical differences in completeness and correctness, as reported below.

Integrating artificial intelligence (AI) into the healthcare sector signals 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 in healthcare and medical research needs 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 [20]. 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 increase in rare disease (Figure 1).

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 has 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 suggest 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 datasets 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.

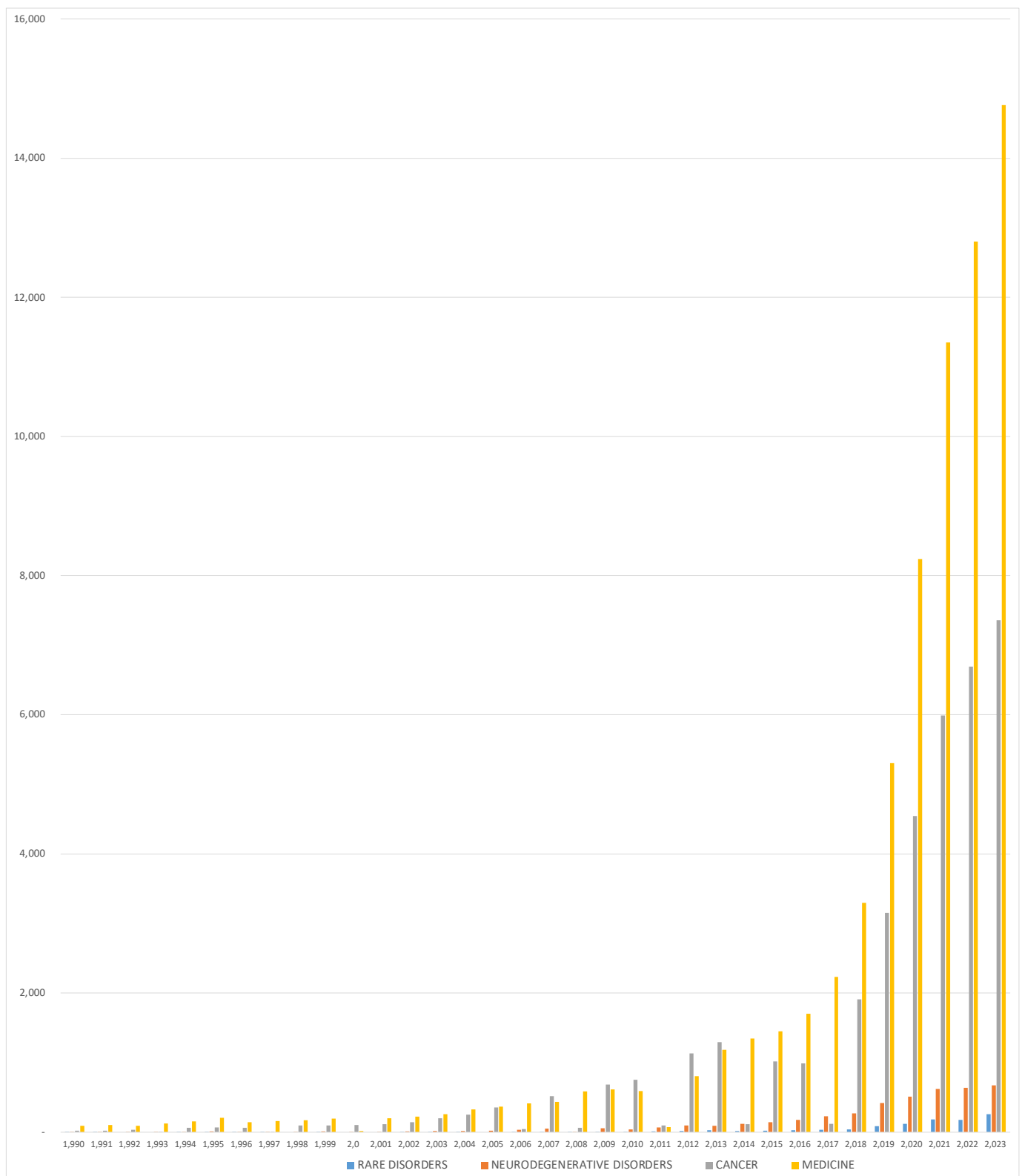


Figure 1. Histogram showing the number of PubMed results per year (“Artificial Intelligence” and, respectively, “rare disorders”, “neurodegenerative disorders”, “cancer”, and “medicine”).

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 [21]. 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 [22,23].

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 resembles the structure and detail expected in standard research projects [24]. 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 confirmed. 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 [25]. 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 [26].

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 12 February 2024), PubMed acknowledges four articles that formally list ChatGPT in the authorship [27–30], while Scopus records three [31–33]. 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 [33].

Support from chatbots was tested as a replacement for human authors. In June 2023, some scientists even produced a paper entirely using ChatGPT [34]. 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 [35–37]. 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 [38].

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, IL, 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 [39,40]. 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 [20,41,42]. 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 [43,44]

Notably, apart from the lack of responsibility for AI systems, some other ethical issues have been highlighted. In fact, it is well known that AI algorithms may be influenced by biases of healthcare data. Some other ethical implications in the use of AI systems concern the lack of transparency, data quality (i.e., the presence of “hallucinations”), and privacy vulnerabilities [45].

In particular, 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 showing a lack of transparency [18,19]. AI will never replace 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 [24,46]

Furthermore, sometimes ChatGPT fabricates 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 the reports. Three key features were reported by reviewers as essential for discrimination: poor description of cutaneous findings, imprecise report of pathophysiology, and inaccurate references [47]. 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, despite 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 [48,49]. 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 [50]. Furthermore, the chatbot provided fabricated references to support its thesis [48,49]. 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 [51]. In a recent study, Gravel and coworkers evaluated 59 references provided by ChatGPT and retrieved that almost two-thirds of them were fabricated.

Interestingly, the fabricated references seemed real at 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 [51]. Interestingly, change in the topic did not improve the truthfulness of references [52,53]. Another described ChatGPT hallucination is in the production of

medical notes, sometimes fabricating patient features (i.e., it generates BMI score, without height and weight data) [23,54].

Notably, prompt engineering enhances the capabilities of pre-trained large language models and reports some methods to prevent LLM hallucinations [55]. In this scenario, some tips have been proposed to reduce the “temperature” of the interaction. In the LLM context, temperature is an indicator of the chatbot’s creativity in its answers [56]. When evaluating scientific fields, the temperature should be low. On the contrary, for creative tasks (i.e., poem generation) the temperature might be increased. Temperature might be changed in a raw API (application programming interface) interaction with LLM. However, ChatGPT can also be guided by providing instructions. For instance, the chatbot will provide more conservative answers when the question asks for a “direct” or “concise” response. Alternately, ChatGPT will provide a more creative or elaborate answer, when the user requests it to “be creative” or to “use imagination”. Similarly, changing the approach (i.e., from “one-shot learning” to “few-shot learning”) can improve chatbot performance [11]. Accurate guidelines to improve the input text and modify the GPT temperature are still lacking. 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 [21–24,57,58]. 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 is 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 by generating manuscripts in multiple languages. In some contexts, English could be an impediment, and ChatGPT can bridge the gap by generating copies of a manuscript in different languages. Similarly, in the conduction of cross-cultural research studies, it may support communication processes. Nevertheless, great attention should be paid to content, as chatbots can generate misleading or inaccurate content with the risk of causing misrepresentation instead of knowledge dissemination [59].

ChatGPT demonstrated a good impact in limiting misinformation derived from the internet on cancer myths. In fact, despite much harmful information available online about cancer [60], the chatbot demonstrated good accuracy in its response to cancer myths [61]. Similarly, in other contexts, ChatGPT was helpful in providing comprehensive information to patients, helping them to understand medical information and treatment options [62]. Undoubtedly, it is advisable to subject ChatGPT to specific training to maintain its ability and prevent the sharing of incorrect information. This 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. Fine-tuning of LLMs, such as ChatGPT, can improve their performance in specific fields [63]. The relevance of fine-tuning might promote its application in medicine, especially in rare disorder management.

The platform’s feasibility is one reason for its widespread diffusion. Other main strengths of ChatGPT are in the form and accessibility of the platform. The user–chatbot interaction is straightforward and 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 using online resources instead of seeking help through face-to-face methods [64]. 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) [65].

ChatGPT shows 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 [66]. The substantial safeness of the chatbot may support the care continuum but confirm its inability to replace medical judgment. As an example, a French study evaluated ChatGPT responses to a virtual patient affected by systemic lupus erythematosus (SLE) and asked for their 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 [21]. Similar issues were noted with cardiovascular conditions, where it performed better on straightforward questions and case vignettes than on complex decision-making scenarios [67].

Likewise, ChatGPT performances on question resolution were confirmed to be 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. The 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) [68]. Interestingly, ChatGPT performances on rare and familial disorders (such 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 are many data available on the web. There were no significant differences according to the difficulty level of questions, except for completeness scores that reached a median of 2.5/3 (mean 2.4, SD 0.7) for complex answers [68].

It is 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 [25]. 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 [26].

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 [69–71]. 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 also face the stigma surrounding mental health [72]. 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 [73]. Likewise, there are other contexts where medical knowledge is deficient in detecting rare disorders and associated risks. For example, it is well known that patients affected by Charcot-Marie Tooth disease should avoid some drugs which may accelerate the disease's progression. Unfortunately, many patients remain undiagnosed [74], lacking appropriate management of their condition. The diffusion of friendly instruments such as ChatGPT, after 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 [43]. In particular, a study conducted by Cascella 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 content that required medical expertise, such as the identification of causal relationships among conditions [43]. 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 composing patients' discharge summaries and operative notes [75]. These documents require expert revision, but the diffusion of these AI models' language in medical centers may improve the time to produce these medical records, providing a quick and high-quality transition among healthcare centers [76]. Furthermore, ChatGPT's ability to admit and learn from mistakes is very promising [23]. For example, in the operative note for a patient with age-related macular degeneration, ChatGPT correctly adjusted anesthesia details associated with intraocular ranibizumab injection [76].

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 aging-related diseases) [77–79]. 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 [25,80,81], 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 in finding 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 differentials [25]. Notably, the latter chatbot confirmed its results in rare and familial disorders, giving the correct diagnosis in both Behcet's Disease and AMD. However, 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) [15].

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 five most probable diagnoses. The results were compared with answers from 12 medical doctors (six “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 the medical doctor group achieved only 57.08% ($\pm 4.8\%$), while the expert group achieved 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 [82]. 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 asked about the relationship between mutations in SCN9A and autosomal dominant epilepsy, ChatGPT incorrectly gave positive responses [83–85]. This evidence strongly suggests 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 of ChatGPT (GPT-3.5 and GPT-4) to detect the correct diagnosis was very weak for rare disorders, while it was acceptable for common diseases [86]. Compared with medical doctors, ChatGPT reached the correct diagnosis in the 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). For 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 two of them, the diagnostic accuracy increased to 50% within the first two suggested diagnoses [86].

The differences significantly increased when ChatGPT diagnostic performances were compared to a less experienced group, such as medical-journal readers. A study on clinical case challenges from *New England Journal of Medicine* revealed that GPT-4 provided the correct diagnosis for 22 of 38 clinical cases (57%), whereas the medical-journal readers chose the proper diagnosis among six provided options for 36% of cases [87].

ChatGPT is continuously evolving, and the evaluation of ChatGPT4 reliability shows 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 a 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) [46]. 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).

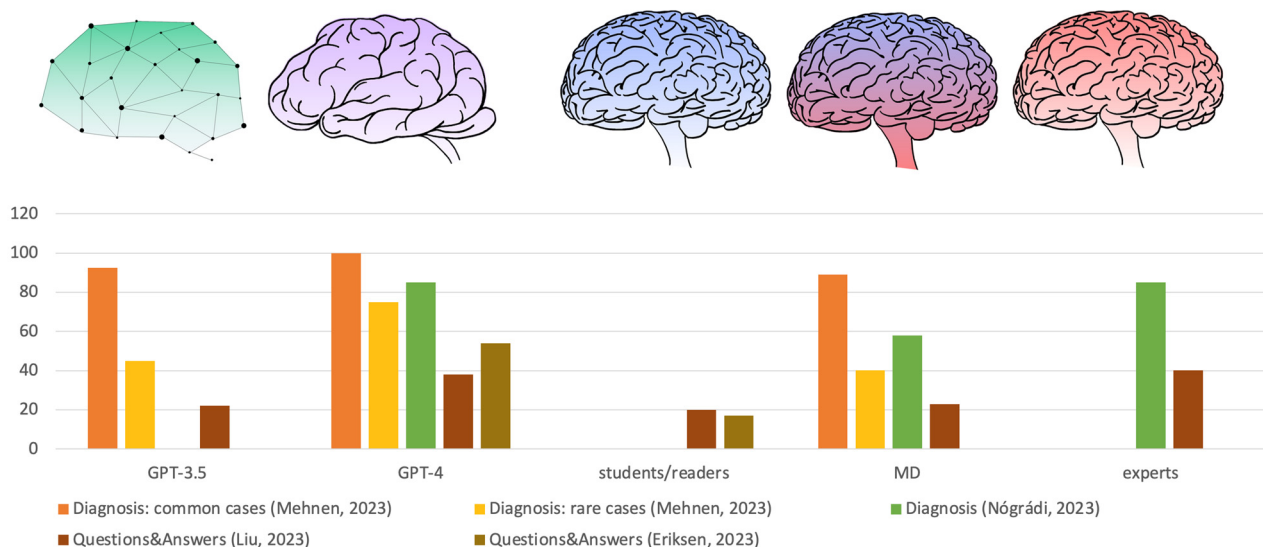


Figure 2. Comparative performance analysis. (Left to right: GPT-3.5, GPT-4, medical-journal readers/students, medical doctors (MDs), experts) [82,86–88].

4.3. Support in Treatment Advice

A study on ChatGPT's reliability in reporting cancer therapies for solid tumors showed 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. The 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. However, recommendations may individually vary among patients [89–91]. In a consequential survey of the same field, but with different criteria for chatbot response evaluation, the 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 [92]. 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 [93,94], and 20% had low-grade gliomas. Interestingly, when 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 the patient's preferences and functional status, no alternative therapies were listed, nor an alternative diagnosis [95]. Otherwise, the treatment suggestions were evaluated positively by a team of experts [95].

In summary, ChatGPT's performance evaluation depended on the extensiveness of the knowledge and their availability on the web. In particular, the assessment of patient functional status was defined as moderate, maybe because of the small number of clinical trials available online [25]. In this scenario, it seems that the naïve ChatGPT can support multidisciplinary activities in neuro-oncology, but it requires complete training. For these reasons, Guo and coworkers created a trained version of ChatGPT, called neuroGPT-X, and evaluated it, comparing with naïve ChatGPT and leading neurosurgical experts worldwide. Despite its ability to support the neurosurgeons, the human expert's evaluation remains necessary, mainly to ensure the safety and reliability of chatbot responses [96].

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 the chatbot and neurosurgeons. The answers were evaluated

by a team of senior neurosurgeons, who 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 [88].

Another study evaluated the reliability of ChatGPT in reporting potential dangers associated with drug–drug interactions [97]. 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 [97]. In this context, ChatGPT confirmed its ability to recognize mistakes [23].

5. Discussion

In this review, medical applications of ChatGPT have been reviewed, from support in drafting medical records to more specialistic purposes, such 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, as well as other LLMs (such as Gemini), is far from replacing human qualities in medical practice [98]. Nonetheless, GPT represents a technology with enormous potential that is rapidly growing. In fact, the new era of AI chatbots has the ability to analyze images. Apart from GPT-4, Gemini (Google’s AI) and LLaMa (Meta’s AI) are promising AI chatbots with image analysis capabilities. To date, this ability seems to be far from application in clinical practice [3], because the evaluation of medical images is complicated by many factors. Nevertheless, the integration of chatbots with existing AI systems, such as Face2Gene, might improve the ability to recognize rare disorders [99]. Furthermore, although GPT demonstrated a good, even if non optimal, performance in recognizing rare disorders, the chatbot achieved better results than other diagnostic tools [25]. 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. This is evident from the significant number of rare disorders that are still undiagnosed [74]. 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 and identify families requiring a more precise genetic evaluation. To date, the only 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%) [100]. 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's 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 has to ask for other signs/symptoms to complete the clinical picture of the disease [86,101]. The ability to ask the right question is one of the main ChatGPT limits, requiring medical intervention to translate the clinical phenotype to a case vignette ready for chatbot evaluation [101].

Nevertheless, it is essential to note that ChatGPT does not claim to be a doctor, nor to replace one. In many cases, the chatbot answers with a disclaimer about "I am not a doctor, but" [24]; 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 [102]. This answer is not only in line with current guidelines [103], 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 [102].

The immediate application of AI represents a significant technological advancement in the statistical analysis of merged datasets. 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 [104–106]. This suggests that AI not only enhances efficiency but also reduces the need for extensive sample sizes.

In summary, the evaluation of ChatGPT in the clinical practice of rare disorders demonstrated a good potential. In detail, research and diagnostic applications of ChatGPT as a support for 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. This is a hypothesis that might explain the high accuracy of ChatGPT even in uncommon diseases. 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.

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., G.T. 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. and 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

- Sai, S.; Gaur, A.; Sai, R.; Chamola, V.; Guizani, M.; Rodrigues, J.J. Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies and Limitations. *IEEE Access* **2024**, *12*, 31078–31106. [CrossRef]
- Marcus, G.; Davis, E.; Aaronson, S. A very preliminary analysis of dall-e 2. *arXiv* **2022**, arXiv:2204.13807.
- Masalkhi, M.; Ong, J.; Waisberg, E.; Lee, A.G. Google DeepMind's gemini AI versus ChatGPT: A comparative analysis in ophthalmology. *Eye* **2024**. [CrossRef] [PubMed]
- Venigalla, A.; Frankle, J.; Carbin, M. Biomedlm: A Domain-Specific Large Language Model for Biomedical Text. MosaicML. Available online: <https://www.mosaicml.com/blog/introducing-pubmed-gpt> (accessed on 23 December 2022).
- Karkera, N.; Acharya, S.; Palaniappan, S.K. Leveraging pre-trained language models for mining microbiome-disease relationships. *BMC Bioinform.* **2023**, *24*, 290. [CrossRef] [PubMed]
- Xue, V.W.; Lei, P.; Cho, W.C. The potential impact of ChatGPT in clinical and translational medicine. *Clin. Transl. Med.* **2023**, *13*, e1216. [CrossRef] [PubMed]
- Stokel-Walker, C.; Van Noorden, R. What ChatGPT and generative AI mean for science. *Nature* **2023**, *614*, 214–216. [CrossRef] [PubMed]
- Anonymous. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* **2023**, *613*, 612. [CrossRef] [PubMed]
- Shen, Y.; Heacock, L.; Elias, J.; Hentel, K.D.; Reig, B.; Shih, G.; Moy, L. ChatGPT and other large language models are double-edged swords. *Radiology* **2023**, *307*, 230163. [CrossRef] [PubMed]
- The Lancet Digital Health. ChatGPT: Friend or foe? *Lancet Digit. Health* **2023**, *5*, E102. [CrossRef]
- Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inform. Proc. Syst.* **2020**, *33*, 1877–1901.
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
- Hosseini, M.; Gao, C.A.; Liebovitz, D.M.; Carvalho, A.M.; Ahmad, F.S.; Luo, Y.; MacDonald, N.; Holmes, K.L.; Kho, A. An exploratory survey about using ChatGPT in education, healthcare, and research. *PLoS ONE* **2023**, *18*, e0292216. [CrossRef]
- Yadava, O.P. ChatGPT—a foe or an ally? *Indian. J. Thorac. Cardiovasc. Surg.* **2023**, *39*, 217–221. [CrossRef] [PubMed]
- 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*, 1088–1089. [CrossRef] [PubMed]
- 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**, *85*, 1346–1349. [CrossRef]
- 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*, 1130–1135. [CrossRef] [PubMed]
- AdvancedAds. Gamechanger: ChatGPT Provides Current Data. Not Limited to before September 2021. Available online: <https://wpadvancedads.com/chatgpt-provides-current-data/#:~:text=Gamechanger:%20ChatGPT%20provides%20current%20data,limited%20to%20before%20September%202021&text=Now%20that%20it%20has%20been,to%20use%20it%20more%20efficiently> (accessed on 5 February 2024).
- 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-15RH2NDt7SC907QbydJIu2jPmZ64H_eqVsb-Rf4](https://twitter.com/OpenAI/status/1707077710047216095?t=oeBD2HTJg2HvQeKF6v-MUG&s=1[%E2%80%A6]d=IwAR04RwUXxRfjOVXGo4L-15RH2NDt7SC907QbydJIu2jPmZ64H_eqVsb-Rf4) (accessed on 5 February 2024).
- 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. [CrossRef]
- 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*, 218–226. (In French) [CrossRef]
- Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepaño, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G.; Maningo, J.; et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit. Health* **2023**, *2*, e0000198. [CrossRef]

23. 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. [[CrossRef](#)]
24. 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*, 103360. [[CrossRef](#)] [[PubMed](#)]
25. 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. [[CrossRef](#)]
26. Uz, C.; Umay, E. “Dr ChatGPT”: Is it a reliable and useful source for common rheumatic diseases? *Int. J. Rheum. Dis.* **2023**, *26*, 1343–1349. [[CrossRef](#)]
27. Benichou, L. ChatGPT. The role of using ChatGPT AI in writing medical scientific articles. *J. Stomatol. Oral. Maxillofac. Surg.* **2023**, *124*, 101456. [[CrossRef](#)] [[PubMed](#)]
28. Curtis, N.; ChatGPT. To ChatGPT or not to ChatGPT? The Impact of Artificial Intelligence on Academic Publishing. *Pediatr. Infect. Dis. J.* **2023**, *42*, 275. [[CrossRef](#)] [[PubMed](#)]
29. King, M.R.; ChatGPT. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cell Mol. Bioeng.* **2023**, *16*, 1–2. [[CrossRef](#)] [[PubMed](#)]
30. 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. [[CrossRef](#)] [[PubMed](#)]
31. Mijwil, M.M.; Aljanabi, M.; ChatGPT. Towards Artificial Intelligence-Based Cybersecurity: The Practices and ChatGPT Generated Ways to Combat Cybercrime. *Iraqi J. Comput. Sci. Math.* **2023**, *4*, 65–70.
32. Aljanabi, M.; Ghazi, M.; Ali, A.H.; Abed, S.A.; ChatGpt. ChatGpt: Open Possibilities. *Iraqi J. Comput. Sci. Math.* **2023**, *4*, 62–64.
33. O’Connor, S. ChatGPT Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Educ. Pract.* **2023**, *66*, 103537. [[CrossRef](#)]
34. Conroy, G. Scientists used ChatGPT to generate an entire paper from scratch—But is it any good? *Nature* **2023**, *619*, 443–444. [[CrossRef](#)] [[PubMed](#)]
35. Sciedirect: Guide for Authors. Available online: <https://www.sciencedirect.com/journal/resources-policy/publish/guide-for-authors> (accessed on 5 February 2024).
36. Cell: Guide for Authors. Available online: <https://www.cell.com/device/authors> (accessed on 5 February 2024).
37. 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 February 2024).
38. Stokel-Walker, C. ChatGPT listed as author on research papers: Many scientists disapprove. *Nature* **2023**, *613*, 620–621. [[CrossRef](#)] [[PubMed](#)]
39. Gao, C.A.; Howard, F.M.; Markov, N.S.; Dyer, E.C.; Ramesh, S.; Luo, Y.; Pearson, A.T. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *Npj Digit. Med.* **2023**, *6*, 75. [[CrossRef](#)] [[PubMed](#)]
40. 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*, e1207. [[CrossRef](#)] [[PubMed](#)]
41. Else, H. Abstracts written by ChatGPT fool scientists. *Nature* **2023**, *613*, 423. [[CrossRef](#)] [[PubMed](#)]
42. Aydın, Ö.; Karaarslan, E. OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. In *Emerging Computer Technologies 2*; Aydın, Ö., Ed.; İzmir Akademi Derneği: Tire, Turkey, 2022; pp. 22–31.
43. 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*, 33. [[CrossRef](#)] [[PubMed](#)]
44. Blanco-González, A.; Cabezón, A.; Seco-González, A.; Conde-Torres, D.; Antelo-Riveiro, P.; Piñeiro, Á.; Garcia-Fandino, R. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals* **2023**, *16*, 891. [[CrossRef](#)] [[PubMed](#)]
45. Jeyaraman, M.; Balaji, S.; Jeyaraman, N.; Yadav, S. Unraveling the Ethical Enigma: Artificial Intelligence in Healthcare. *Cureus* **2023**, *15*, e43262. [[CrossRef](#)]
46. Mello, M.M.; Guha, N. ChatGPT and Physicians’ Malpractice Risk. *JAMA Health Forum.* **2023**, *4*, e231938. [[CrossRef](#)]
47. Dunn, C.; Hunter, J.; Steffes, W.; Whitney, Z.; Foss, M.; Mammino, J.; Leavitt, A.; Hawkins, S.D.; Dane, A.; Yungmann, M.; 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*, 388–390. [[CrossRef](#)]
48. Goddard, J. Hallucinations in ChatGPT: A Cautionary Tale for Biomedical Researchers. *Am. J. Med.* **2023**, *136*, 1059–1060. [[CrossRef](#)] [[PubMed](#)]
49. Alkaissi, H.; McFarlane, S.I. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus* **2023**, *15*, e35179. [[CrossRef](#)] [[PubMed](#)]
50. Hirschhorn, R.; Reuser, A.J. Glycogen storage disease type II: Acid alpha-glucosidase (acid maltase) deficiency. In *The Metabolic and Molecular Bases of Inherited Disease*; Scriver, C.R., Beaudet, A., Sly, W.S., Valle, D., Eds.; McGraw-Hill: New York, NY, USA, 2001; pp. 3389–3420.
51. Gravel, J.; D’Amours-Gravel, M.; Osmanlliu, E. Learning to Fake It: Limited Responses and Fabricated References Provided by ChatGPT for Medical Questions. *Mayo Clin. Proc. Digit. Health* **2023**, *1*, 226–234. [[CrossRef](#)]
52. Day, T. A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT. *Prof. Geogr.* **2023**, *75*, 1024–1027. [[CrossRef](#)]

53. Javid, M.; Reddiboina, M.; Bhandari, M. Emergence of artificial generative intelligence and its potential impact on urology. *Can. J. Urol.* **2023**, *30*, 11588–11598. [PubMed]
54. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.J.; Madotto, A.; Fung, P. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **2022**, *55*, 1–38. [CrossRef]
55. Sahoo, P.; Singh, A.; Saha, S.; Jain, V.; Mondal, S.; Chadha, A. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. *arXiv* **2024**, arXiv:2402.07927v1. [CrossRef]
56. Available online: <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api> (accessed on 5 March 2024).
57. 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*, A06230412. [PubMed]
58. Cox, L.A., Jr. An AI assistant to help review and improve causal reasoning in epidemiological documents. *Glob. Epidemiol.* **2023**, *7*, 100130. [CrossRef]
59. Alsadhan, A.; Al-Anezi, F.; Almohanna, A.; Alnaim, N.; Alzahrani, H.; Shinawi, R.; AboAlsamh, H.; Bakhshwain, A.; Alenazy, M.; Arif, W.; et al. The opportunities and challenges of adopting ChatGPT in medical research. *Front. Med.* **2023**, *10*, 1259640. [CrossRef]
60. Johnson, S.B.; Parsons, M.; Dorff, T.; Moran, M.S.; Ward, J.H.; Cohen, S.A.; Akerley, W.; Bauman, J.; Hubbard, J.; Spratt, D.E.; et al. Cancer misinformation and harmful information on Facebook and other social media: A brief report. *J. Natl. Cancer Inst.* **2022**, *114*, 1036–1039. [CrossRef] [PubMed]
61. 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*, pkad015. [CrossRef] [PubMed]
62. Kim, J.-H. Search for Medical Information and Treatment Options for Musculoskeletal Disorders through an Artificial Intelligence Chatbot: Focusing on Shoulder Impingement Syndrome. Available online: <https://www.medrxiv.org/content/10.1101/2022.12.16.22283512v2.full-text> (accessed on 26 February 2024).
63. Li, Y.; Li, Z.; Zhang, K.; Dan, R.; Jiang, S.; Zhang, Y. ChatDoctor: A Medical Chat Model Fine-Tuned on a Large Language Model Meta-AI (LLaMA) Using Medical Domain Knowledge. *Cureus* **2023**, *15*, e40895. [CrossRef] [PubMed]
64. 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. [CrossRef] [PubMed]
65. Nov, O.; Singh, N.; Mann, D. Putting ChatGPT's Medical Advice to the (Turing) Test: Survey Study. *JMIR Med. Educ.* **2023**, *9*, e46939. [CrossRef]
66. 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*, 17885. [CrossRef]
67. 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. [CrossRef]
68. Johnson, D.; Goodman, R.; Patrinely, J.; Stone, C.; Zimmerman, E.; Donald, R.; Chang, S.; Berkowitz, S.; Finn, A.; Jahangir, E.; et al. Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model. *Res. Sq. [Preprint]* **2023**. [CrossRef]
69. 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. [CrossRef] [PubMed]
70. Pham, K.; Nabizadeh, A.; Selek, S. Artificial intelligence and chatbots in psychiatry. *Psychiatr. Q.* **2022**, *93*, 249–253. [CrossRef]
71. 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. [CrossRef] [PubMed]
72. 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. [CrossRef] [PubMed]
73. Allen, S. Artificial intelligence and the future of psychiatry. *IEEE Pulse* **2020**, *11*, 2–6. [CrossRef] [PubMed]
74. 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. [CrossRef]
75. Patel, S.B.; Lam, K. ChatGPT: The future of discharge summaries? *Lancet Digit. Health* **2023**, *5*, e107–e108. [CrossRef] [PubMed]
76. Singh, S.; Djalilian, A.; Ali, M.J. ChatGPT and Ophthalmology: Exploring Its Potential with Discharge Summaries and Operative Notes. *Semin. Ophthalmol.* **2023**, *38*, 503–507. [CrossRef]
77. Horvat, N.; Veeraraghavan, H.; Nahas, C.S.R.; Bates, D.B.; Ferreira, F.R.; Zheng, J.; Capanu, M.; Fuqua, J.L.; Fernandes, M.C.; Sosa, R.E.; et al. Combined artificial intelligence and radiologist model for predicting rectal cancer treatment response from magnetic resonance imaging: An external validation study. *Abdom. Radiol.* **2022**, *47*, 2770–2782. [CrossRef] [PubMed]
78. Pun, F.W.; Leung, G.H.D.; Leung, H.W.; Liu, B.H.M.; Long, X.; Ozerov, I.V.; Wang, J.; Ren, F.; Aliper, A.; Izumchenko, E.; et al. Hallmarks of aging-based dual-purpose disease and age-associated targets predicted using PandaOmics AI-powered discovery engine. *Aging* **2022**, *14*, 2475–2506. [CrossRef] [PubMed]
79. Rao, A.; Kim, J.; Kamineni, M.; Pang, M.; Lie, W.; Succi, M.D. Evaluating ChatGPT as an adjunct for radiologic decision-making. *MedRxiv* **2023**. [CrossRef]

80. 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. Safte* **2022**, *31*, 426–433. [CrossRef]
81. Riches, N.; Panagiotti, M.; Alam, R.; Cheraghi-Sohi, S.; Campbell, S.; Esmail, A.; Bower, P. The effectiveness of electronic differential diagnoses (DDX) generators: A systematic review and meta-analysis. *PLoS ONE* **2016**, *11*, e0148991. [CrossRef] [PubMed]
82. Nógrádi, B.; Polgár, T.F.; Meszlényi, V.; Kádár, Z.; Hertelendy, P.; Csáti, A.; Szpisjak, L.; Halmi, D.; Erdélyi-Furka, B.; Tóth, M.; et al. ChatGPT M.D.: Is There Any Room for Generative AI in Neurology and Other Medical Areas? Available online: <https://ssrn.com/abstract=4372965> (accessed on 10 January 2024). [CrossRef]
83. Boßelmann, C.M.; Leu, C.; Lal, D. Are AI language models such as ChatGPT ready to improve the care of individuals with epilepsy? *Epilepsia* **2023**, *64*, 1195–1199. [CrossRef] [PubMed]
84. Brunklaus, A. No evidence that SCN9A variants are associated with epilepsy. *Seizure* **2021**, *91*, 172–173. [CrossRef]
85. Curation Results for Gene-Disease Validity. Available online: https://search.clinicalgenome.org/kb/gene-validity/CGGV:assertion_72a91ef6-e052-44a4-b14e-6a5ba93393ff-2021-03-09T163649.218Z (accessed on 19 December 2023).
86. Mehnen, L.; Gruarin, S.; Vasileva, M.; Knapp, B. Chat GPT as a medical doctor? A diagnostic accuracy study on common and rare diseases. *MedRxiv* **2023**. [CrossRef]
87. Eriksen, A.V.; Möller, S.; Jesper, R. Use of GPT-4 to Diagnose Complex Clinical Cases. *NEJM AI* **2023**, *1*, AIp2300031. [CrossRef]
88. 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*, 107590. [CrossRef]
89. Schulte, B. Capacity of ChatGPT to Identify Guideline-Based Treatments for Advanced Solid Tumors. *Cureus* **2023**, *15*, e37938. [CrossRef] [PubMed]
90. American Society of Clinical Oncology Guidelines. Available online: <https://society.asco.org/practice-patients/guidelines> (accessed on 19 December 2023).
91. European Society of Medical Oncology Clinical Practice Guidelines. Available online: <https://www.esmo.org/guidelines> (accessed on 19 December 2023).
92. 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**. [CrossRef]
93. 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*, 301. [CrossRef]
94. Glioblastoma. Available online: <https://rarediseases.info.nih.gov/diseases/2491/glioblastoma> (accessed on 21 February 2024).
95. Haemmerli, J.; Sveikata, L.; Nouri, A.; May, A.; Egervari, K.; Freyschlag, C.; Lobrinus, J.A.; Migliorini, D.; Momjian, S.; Sanda, N.; 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*, e100775. [CrossRef]
96. Guo, E.; Gupta, M.; Sinha, S.; Rössler, K.; Tatagiba, M.; Akagami, R.; El-Mefty, O.; Sugiyama, T.; Stieg, P.; Pickett, G.E.; et al. NeuroGPT-X: Towards an Accountable Expert Opinion Tool for Vestibular Schwannoma. *MedRxiv Mendeley Data V1* **2023**. [CrossRef]
97. 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*, e36272. [CrossRef] [PubMed]
98. Tripathi, M.; Chandra, S.P. ChatGPT: A threat to the natural wisdom from artificial intelligence. *Neurol. India* **2023**, *71*, 416–417. [CrossRef] [PubMed]
99. Carrer, A.; Romaniello, M.G.; Calderara, M.L.; Mariani, M.; Biondi, A.; Selicorni, A. Application of the Face2Gene tool in an Italian dysmorphological pediatric clinic: Retrospective validation and future perspectives. *Am. J. Med. Genet. Part A* **2024**, *194*, e63459. [CrossRef] [PubMed]
100. 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**, *194*, e63493. [CrossRef] [PubMed]
101. Wa, M. Evidence-based clinical practice: Asking focused questions (PICO). *Optom. Vis. Sci.* **2016**, *93*, 1187–1188. [CrossRef]
102. Dergaa, I.; Fekih-Romdhane, F.; Hallit, S.; Loch, A.A.; Glenn, J.M.; Fessi, M.S.; Ben Aissa, M.; Souissi, N.; Guelmami, N.; Swed, S.; et al. ChatGPT is not ready yet for use in providing mental health assessment and interventions. *Front. Psychiatry* **2024**, *14*, 1277756. [CrossRef] [PubMed]
103. Rios, P.; Cardoso, R.; Morra, D.; Nincic, V.; Goodarzi, Z.; Farah, B.; Harricharan, S.; Morin, C.M.; Leech, J.; Straus, S.E.; et al. Comparative effectiveness and safety of pharmacological and non-pharmacological interventions for insomnia: An overview of reviews. *Syst. Rev.* **2019**, *8*, 281. [CrossRef] [PubMed]
104. Cascella, R.; Strafella, C.; Longo, G.; Ragazzo, M.; Manzo, L.; De Felici, C.; Errichiello, V.; Caputo, V.; Viola, F.; Eandi, C.M.; et al. Uncovering genetic and non-genetic biomarkers specific for exudative age-related macular degeneration: Significant association of twelve variants. *Oncotarget* **2017**, *9*, 7812–7821. [CrossRef]

105. Ricci, F.; Zampatti, S.; D'Abbruzzi, F.; Missiroli, F.; Martone, C.; Lepre, T.; Pietrangeli, I.; Sinibaldi, C.; Peconi, C.; Novelli, G.; 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*, 1368–1372. [[CrossRef](#)]
106. Ricci, F.; Staurengli, G.; Lepre, T.; Missiroli, F.; Zampatti, S.; Cascella, R.; Borgiani, P.; Marsella, L.T.; Eandi, C.M.; Cusumano, A.; et al. Haplotypes in IL-8 Gene Are Associated to Age-Related Macular Degeneration: A Case-Control Study. *PLoS ONE* **2013**, *8*, e66978. [[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.