Turkish Archives of Pediatrics

REVIEW
DOI: 10.5152/TurkArchPediatr.2024.24002

Artificial Intelligence in Pediatrics: Learning to Walk Together

Kaan Can Demirbaş[1], Mehmet Yıldız[2], Seha Saygılı[3], Nur Canpolat[3], Özgür Kasapçopur[2]

[1]İstanbul University-Cerrahpaşa, Cerrahpaşa Faculty of Medicine, İstanbul, Turkey
[2]Department of Pediatric Rheumatology, İstanbul University-Cerrahpaşa, Cerrahpaşa Faculty of Medicine, İstanbul, Turkey
[3]Department of Pediatric Nephrology, İstanbul University-Cerrahpaşa, Cerrahpaşa Faculty of Medicine, İstanbul, Turkey

ABSTRACT

In this era of rapidly advancing technology, artificial intelligence (AI) has emerged as a transformative force, even being called the Fourth Industrial Revolution, along with gene editing and robotics. While it has undoubtedly become an increasingly important part of our daily lives, it must be recognized that it is not an additional tool, but rather a complex concept that poses a variety of challenges. AI, with considerable potential, has found its place in both medical care and clinical research. Within the vast field of pediatrics, it stands out as a particularly promising advancement. As pediatricians, we are indeed witnessing the impactful integration of AI-based applications into our daily clinical practice and research efforts. These tools are being used for simple to more complex tasks such as diagnosing clinically challenging conditions, predicting disease outcomes, creating treatment plans, educating both patients and healthcare professionals, and generating accurate medical records or scientific papers. In conclusion, the multifaceted applications of AI in pediatrics will increase efficiency and improve the quality of healthcare and research. However, there are certain risks and threats accompanying this advancement including the biases that may contribute to health disparities and, inaccuracies. Therefore, it is crucial to recognize and address the technical, ethical, and legal challenges as well as explore the benefits in both clinical and research fields.

Keywords: Pediatrics, artificial intelligence, machine learning, clinical decision-making

INTRODUCTION

The role of computers and medical informatics in healthcare has been a subject of discussion for decades.[1] In this rapidly advancing era, artificial intelligence (AI) has emerged as a revolutionary force, even defined as the Fourth Industrial Revolution along with gene editing and robotics.[2] Rather than being an ordinary additional tool, it is a complex broad concept that has the potential to reshape the structure and establishment of modern medicine and drastically empower both pediatricians and patients/parents while bringing some significant clinical, technical, ethical, and legal challenges as expected from any advancement.[3,4]

AI-based systems have the capacity to perform tasks that would typically require human intelligence such as understanding language, learning, self-training, decision-making, and problem-solving. With these capabilities, theoretically, AI tools can analyze enormous amounts of written, visual, or auditory datasets and train themselves. They can be employed to diagnose and evaluate clinically challenging presentations, predict outcomes and prognoses, create management plans according to the most up-to-date knowledge, educate patients and medical professionals, and produce accurate medical records or even scientific papers.[5] The potential of AI-based tools is particularly thrilling given the mounting pressures on healthcare systems and clinicians due to increasing demand, complexity of cases, and limited resources. These tools have the potential to reduce the need for a larger workforce and enhance productivity and efficiency in childcare. However, AI-based applications go together with significant risk factors regarding the quality, completeness, and adequacy of the data; safety and security; and most importantly responsibility attribution and governance

Corresponding author:
Kaan Can Demirbaş
✉ demirbaskaancan@gmail.com
Received: January 3, 2024
Accepted: February 2, 2024
Publication Date: March 1, 2024

Cite this article as: Demirbaş KC, Yıldız M, Saygılı S, Canpolat N, Kasapçopur Ö. Artificial intelligence in pediatrics: Learning to walk together. *Turk Arch Pediatr.* 2024;59(2):121-130

of the automated technology. To move further beyond theory and implement these tools in real-life practice, risks and challenges associated with AI must be considered and studied extensively. Rather than being an object of AI, medicine itself should be the subject and use AI-based technologies according to its own principles, rules, and regulations.[6]

This review is written in this scope, serving as an introduction for pediatricians to the vast subject analyzing the current state of the art of its applications to clinical practice and research in pediatrics. The first section briefly introduces the concept of AI to clinicians, highlighting its history and subset definitions. The subsequent section emphasizes the importance of data, analysis, and prediction in pediatrics and the role of AI in this process. The third part of the article provides a conclusion, focusing on large language models and their role as a bridge between humans and AI, as well as exploring their clinical application.

BACK TO BASICS—WHAT IS AI?

Commonly, artificial intelligence (AI) is defined as a set of approaches that can perform tasks that typically require human intelligence such as information perception, language recognition, decision-making, task creation, and problem-solving.[7] It is a rapidly improving field of computer science and its products and effects have already reclaimed their places in other various fields and sectors such as economics, education, energy, and manufacturing. Currently, AI comprises different types of models and approaches that are intertwined with each other and might be confusing. Essential definitions regarding this area are given in the *Glossary box* (Supplementary Table 1) and different models are summarized in Figure 1.

To be able to comprehend the current state-of-the-art AI applications and their capabilities (i.e., their mechanism of action) looking back to its history and development is important. The term "artificial intelligence" was coined by John McCarthy in 1956.[8] In the beginning, researchers' focus was on developing systems that could implement "symbolic reasoning" in which expert-coded strict rules were used to address a defined, circumscribed problem. In these expert systems, rules (such as treatment algorithms, and tax laws) were coded into the program, and upon answering a set of questions by users, the software could provide a result or an answer using "if-then" statements.[9]

A set of examples of this "good old fashion AI" in medicine are the MYCIN system which was designed to advise the proper antibiotic treatment for infections based on user-entered patient information; the CASNET network applied for glaucoma management; and INTERNIST-I, a general consultation system where multiple diagnoses could be achieved according to clinician entered information.[10-12] Although these applications drew significant attention at the time, they failed to achieve widespread adaptation. The major disadvantage was that they were labor-intensive as they needed a team of experts to manually enter a long, up-to-date list of rules and information. Furthermore, to establish and sustain this system effectively, it was vital that the rules and information that had been inputted into it were known and remained stable. This challenge is notably pronounced in the field of medicine, where data and knowledge undergo significant and rapid transformations on a daily basis.[13,14] Hence, during the 1980s and 1990s, AI research shifted toward machine learning (ML) and neural networks to allow a machine to learn from data.[15,16] The increasing predominance of ML methods in AI today has led to 2 terms being used interchangeably.

Machine learning is a subset of AI that has the capability to adapt and learn repetitively by applying statistical models to

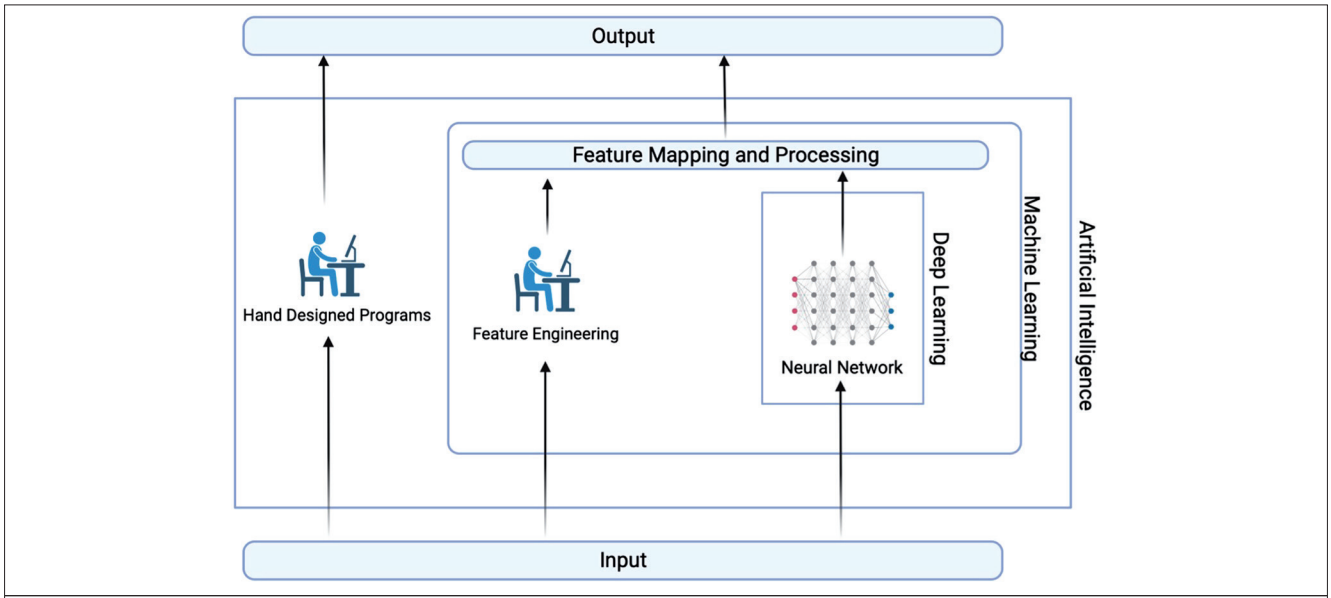

Figure 1. Hierarchy and interaction between main AI-based applications and methods. As explained before, expert systems are examples of hand-designed programs. Moving into the future, machine learning applications that enable machines to learn by themselves without minimum intervention have become more common in use. Deep learning, a subset of machine learning, uses artificial neural networks (ANNs) resembling the connections of neurons in the human brain to process the input in multiple layers and extract progressively higher features from it. It implements a higher learning function in AI-based systems. Created with BioRender.com.

identify patterns in each dataset to draw a conclusion. Although ML models can be categorized into different classes based on their learning approaches, definitions and details of these are out of the scope of this review.[17] It is sufficient to know that the primary objective of ML applications is to build a mathematical mapping from an input to reach an output (Figure 1). For this mapping, different methods labeled as "algorithms" are used.[18] In particular, artificial neural networks (ANN) represent a powerful class of algorithms within ML, inspired by the structure and functioning of the human brain. These networks, composed of interconnected nodes resembling neurons, enable the modeling of complex, uncovered relationships in data and are important aspects of deep learning (DL) models. With implemented ANN structures, DL represents a higher functional capacity for more complex tasks as a subset of ML due to its neural layers. Until now, DL algorithms have been used in different tasks such as drug interaction prediction, malignancy type analysis, classification of radiological images, and stratification of patients.[19-21] The hierarchy of AI, ML, and DL is shown in Figure 1 adapted for a better visualization.[22]

Natural language processing (NLP) is another intriguing and effective subfield of AI that focuses on the interaction between computational systems and humans through natural language processes.[23] It can be considered as a bridge of communication between humans and machines. A pivotal development in the field of NLP has been the development of large language models (LLMs). LLMs are AI systems that are trained using billions of words derived from different sources. They typically use neural network architectures to represent complicated associative relationships between words in the text-based training dataset.[24] They possess broad applications including text generation, translation, content summary, rewriting, classification, categorization, and even sentiment analysis.[15] The 2 cardinal features of LLMs are their ability to learn useful patterns in large amounts of unlabeled data via self-supervision and their capacity to be fine-tuned via the user's prompts (instructions) to generate responses aligned with the user's expectations.[25] Beyond these, some LLMs' training involves "reinforcement learning from human feedback" meaning that these systems also learn from direct human interaction and responses to tune themselves better. Interaction between the user and the programs occurs through a set of prompts and outputs. Prompts are instructions and the basic knowledge that the user gives to the system, after which the system produces a response/output. With these features, one can communicate with machines in a human-like manner in a chosen context, in a chosen manner for a chosen subject. Well-known examples of generalized (i.e., not specifically trained for medical use) LLMs are Google's Bard®, OpenAI's ChatGPT®, and Meta's LLaMA®.

CRACKING THE CODE—PREDICTIVE MODELING AND MACHINE LEARNING

The development of data science and the application of statistics to medicine has been one of the crucial developments of the last century since these approaches are fundamental for the assimilation of prediction and precision to daily clinical practice.[26] However, the ways we use to analyze the data and produce the evidence are gradually evolving and expanding at a rate that the human mind finds impossible to process.[27] In pediatrics and adolescent medicine, the use of ML applications has increased drastically in the previous decades. The first example of a computational system being used in pediatrics was from 1984, when researchers applied a rule-based medical diagnosis system (explained above) called SHELP, to diagnose inborn errors of metabolism.[28] Following this, research on ML and pediatrics has focused on more advanced algorithms such as deep learning, natural language processing, and regression analysis. Notably, data sources for these studies encompass electronic healthcare records (EHR), information obtained from various investigations, and omics data (including genomic, proteomic, and metabolomic data).[27] Due to their characteristics, datasets gained through these sources are categorized as Big Data. As per definition, "big data" is a high volume, high velocity, and high variety information asset that demands sustainable innovative forms of information processing for better insight and decision making.[29] For a dataset to be classified as big data, there is no specified amount limit but rather it is the massive complexity of data that demands advanced technologies like ML and DL to analyze it, making them "big data."[27,30]

While a multitude of studies delve into these areas, it is more viable to focus on some specific studies that do not only elucidate the duality of AI but also provide insights into factors crucial for the effective and secure implementation of these technologies.

Diagnosis and Predictive Risk Analysis

In a recent study, researchers developed an AI-based NLP model that can process free text from physicians' notes in the EHR to accurately predict the primary diagnosis in the pediatric population. The prepared model had been trained using 101.6 million data points, gathered from approximately 1.4 million pediatric patient encounters in a single tertiary children's hospital. After training, logistic regression classifiers were used to establish a hierarchical diagnostic system in the model.[31] Upon evaluation across a spectrum of clinical scenarios, the model demonstrated robust performance, proving its efficacy not only in diagnosing mild and prevalent conditions such as acute sinusitis but also in accurately identifying more critical diagnoses, including meningitis, encephalitis, and asthma exacerbation. Notably, in a comparative analysis assessing the model against 5 distinct physician cohorts, stratified based on their varying levels of expertise and experience, the AI model convincingly outperformed junior pediatricians.

This study was among others to show that AI-based systems that encompass advanced algorithms can be trained with EHR for predictive analysis. Another important example of these comes from neonatal intensive care units (NICU) and data produced from observational records. It was estimated that a well-functioning NICU generates approximately 1 terabyte of data per bed per year.[30,32] Hence, numerous studies have been conducted to implement ML algorithms to detect the risks and/or outcomes of specific neonatal entities.[29,33] In one study where artificial neural networks were used to predict diagnostic risk for neonatal sepsis, the implemented model outperformed physicians and traditional models in place for the diagnosis of sepsis, with a sensitivity and specificity of 93% and

80% respectively.[34] In another one, authors described an ANN architecture to generate a personalized necrotizing enterocolitis risk score incorporating different risk factors including intestinal microbiota of the patients. In this study, the authors were successful in detecting NEC development 8 days prior to clinical onset of the disease with a sensitivity and specificity of 86% and 90% respectively.[35] Different models have been also applied to diagnose other important reasons of morbidity and mortality such as retinopathy of prematurity and bronchopulmonary dysplasia which showed promising results.[36,37]

Notably, EHRs are not the only data source that an AI system can be trained with. Visual data also plays a crucial role in training AI-based systems within the realm of medicine, enabling the development of advanced diagnostic and analytical tools. By leveraging vast datasets of medical images such as x-rays, MRI scans, histopathology slides, and even facial photographs, AI algorithms can be trained to recognize patterns, anomalies, and subtle nuances that might elude the human eye. In pediatric radiology, deep learning methods have been used vastly for a long time to identify and grade a range of conditions including rickets, colitis, hydronephrosis, respiratory infections, and intracranial pathologies.[38-40] Besides these, AI-based models that were developed for the analysis of histopathological slides and medical imaging systems achieved high sensitivity and specificity across different malignancies including acute lymphoid leukemia and central nervous system tumors.[41,42] An important example of the use of visual data comes also from a multinational study where the researchers used deep neural network structures to screen the facial phenotypes of children with suspected genetic syndromes.[43] In this study, facial photographs of children with confirmed genetic syndromes were used to train the system, and the model was evaluated on a patient and healthy control group that included 2800 children. According to their findings, the developed model achieved 88% accuracy for the detection of genetic syndromes, with a sensitivity and specificity of 90% and 86% respectively, accuracy being higher in the White population.

As stated, and exemplified, the implementation of AI-based systems, especially the ones that incorporate advanced methods, has been studied vastly for diagnosis and prediction analysis in a wide range of conditions, including ones as simple as acute respiratory tract infections and more complex diagnoses; encountered in different pediatric sub-specialties. In emergency departments, these systems can be used to assist triage procedures, predict diagnosis, and prioritize patients who need urgent medical attention. This can speed up the workflow, reduce the burden and cost of triage, reduce patient waiting times and, most importantly, help clinicians manage their time, effort, and limited resources. In outpatient clinics, AI tools can be used to diagnose patients with complex or rare conditions to avoid under- or overdiagnosis. In inpatient settings and intensive care units, they can be used to provide a more careful, targeted, and allocated approach to critically ill patients to improve survival.

Beyond Diagnosis and Prediction

Moving beyond diagnosis, AI and machine learning applications might have profound contributions to the advancement of precision medicine and the facilitation of clinical trials in pediatric healthcare. "Big data" in pediatric research does not only come from observations and collections from daily clinical practice but also from translational research that delves into the world of "omics."[44] In the human body, there are thousands of proteins, protein-coding genes, mRNAs, miRNAs, and other bio-entities whose analysis with computational and laboratory methods form genomic, epigenomic, proteomic, transcriptomic, and metabolomic studies.[45] While traditional methods of imaging, pathologic and genetic tests allow pediatricians to diagnose and manage patients, for us to be able to exert precision medicine principles and draw a personalized phenotype, omics studies are crucial. In pediatric medicine, omics studies can be particularly valuable for identifying genetic predispositions to certain diseases, predicting disease progression, and selecting the most effective and least harmful treatment options for young patients. Furthermore, integrating omics data with clinical information can aid in the identification of biomarkers that can be used for early detection, prognosis, and monitoring of pediatric diseases. This comprehensive approach not only enhances diagnostic accuracy but also facilitates the development of targeted interventions that consider an individual's unique genetic and molecular profile, ultimately leading to more effective and tailored healthcare for pediatric patients.[46] However, when examining a singular kind of omics data, the scope is largely restricted to the detection of variances, or at most, correlations between 1 or 2 forms of biological entities. This limitation confines the results to reactive processes rather than causal occurrences.[47] Thus, interconnectivity and interdependence of various bio-entities demand a holistic approach utilizing a far more exhaustive and comprehensive application and integration of multi-omics data. Because of this characteristic of the omics, in the biological context, they can be also classified as "big data."[45,48]

As happens with big data in healthcare, the integration of advanced AI algorithms holds great promise for pediatricians and researchers. These innovations allow for a more personalized strategy and development of novel drugs, prognostic markers, and therapies. This approach can help physicians manage their patients, especially in subspecialties where physicians deal with entities with complex underlying mechanisms, such as malignancies, autoimmune diseases, neurodevelopmental pathologies, and genetic syndromes.[27,49] For example, in a remarkable study, researchers developed a multi-omics late integration (MOLI) method based on deep neural networks. This model used genomic data as input to predict the response of specific cancer types to certain oncological therapies. In the developed model, they achieved a high accuracy in predicting the response of certain malignancies to selective targeted therapies.[50]

In the research context, another area of interest for AI-based models is their applications to clinical trial processes. From expediting patient recruitment and selection through data-driven patient matching algorithms to enhancing the accuracy of participant monitoring and adverse event detection, AI can streamline the often demanding and time-consuming trial phases. Moreover, AI-driven predictive analytics and real-time data analysis might offer valuable insights for optimizing trial design, leading to more efficient protocols and improved decision-making. Its ability to decipher complex patterns within clinical data sets also can contribute to the identification

of potential biomarkers, aiding in the stratification of patient cohorts and facilitating personalized treatment approaches.[51] As the adaptation of AI-based tools into clinical research and daily practice increases, some guidelines have already been published regarding their use and reports of use in the research context.[52,53]

Considerations and Pitfalls

It is beyond doubt that, even just with their capability to analyze and summarize EHRs and vast amounts of clinical data, these systems will bring an enormous reduction in workflow and a drastic improvement in patient care and health system performance. However, the examples presented so far merely scratch the surface. Our focus has been on studies employing the most effective AI models, yielding the most promising outcomes but many of these studies were centered on specific contexts. Throughout the globe, different regulatory authorities use highly detailed and restrictive guidelines for the development of medical devices or novel drugs and their implementation into routine clinical practice. For both, before development and after implementation, continuous surveillance is required to be sure about their feasibility, efficiency, and safety.[54] AI solutions for healthcare differ from drugs or medical devices in that they are designed to affect human decision-making.[55] It is as important as to perform the same criticized approach. It is crucial to recognize that the very features that make AI a potent tool also expose it to limitations and constraints.[56] From this perspective, one must delve into the intricacies of AI including the significance of data quality, the used model, transparency and fairness of the system, governance of its outcomes, and multidisciplinary approach to its implementation, each within its own context.[57]

As most of the AI systems are built on secondary data sources, it is important to realize that *what has been put in, defines what comes out*.[57] In the scope of the data used to train the AI model, the scale of the data, its generalizability, and completeness are important factors. Machine learning systems are as good as the data they are trained with. It has been shown that as the amount and quality of the data increase, the output and performance of the system increase as well.[58,59] High-resolution, large datasets are ideal for ML applications. The importance of the scale of the dataset can be seen in studies and AI-based models where an extremely vast amount of data has been used to train the model, such as ChatGPT where billions of data were used, and in the study done by Liang et al[31] where more than 100 million data points were used and achieved a remarkable outcome. However, as children account for a limited proportion of healthcare sources, datasets in pediatric care are frequently smaller. Even advanced ML models may not be helpful in offering a significant advantage over classical methods to draw conclusions from small datasets.[60] On the other hand, as most of the AI models are trained to detect patterns in the data which might be prespecified or not, in some models false-positive results might increase as well unless rigorous procedures to assess the reproducibility of findings are incorporated.[56]

Apart from the scale of the given dataset, its inclusiveness and completeness are other crucial factors. An ML model can inherit biases and unrepresentativeness that exist in the training dataset.[61] A crucial example of this can be seen in the study where the authors trained a deep neural network model with the facial photographs of children with genetic syndromes.[43] Although researchers in this study showed an extra effort to represent children from other ethnic backgrounds, due to a lack of data from these populations, the model eventually performed worse in diagnosing children with certain ethnic backgrounds, compared to its performance in the white and Hispanic populations. In another study, the authors found unintentional discrimination against black patients in an AI-informed algorithm that is applied to millions of people in the USA to identify patients who were at high risk of incurring substantial healthcare costs.[62] These examples illustrate both the need for human experts in the loop and the need to carefully specify objective functions for training and education.[56] ML outputs can only be trusted if the data is trusted.[57]

If considerations regarding the model and used datasets are met, another important pitfall for developed AI models is their capacity to be generalized and reproduced. One of the most classical examples of this pitfall is the unsuccessful external validation results of developed predictive models, which stems mostly from the lack of interoperability and homogeneity of EHRs or other datasets used in different hospitals and facilities.[63] This disadvantage might be pronounced especially in predictive models used for conditions where syntax and record-keeping procedures show significant differences, such as developed AI models for the prediction of child abuse cases.[64]

Currently, the consortium emphasizes the integration of human oversight to ensure the safe integration of these systems in both healthcare and research settings. However, there remains a significant gap in providing guidance and establishing regulations to address the issues of accountability and ethical dilemmas that may emerge from the deployment of such systems. As pediatricians lack formal education in *advanced* data science, computational model development, and AI, to be able to get the best out of our research on developed models, it is imperative for us to collaborate with data scientists. For a larger scale implementation of AI in healthcare, collaborations involving medical professionals, lawmakers, computer and data scientists, bioethicists, and public representatives are crucial for the safe, accessible, and effective implementation of AI in healthcare.

SIMPLIFYING THE COMPLEX—POTENTIAL OF LARGE LANGUAGE MODELS

Large language models (LLMs), AI algorithms that can understand and produce language and speech, are among the latest and most advanced developments in this field. The most notable, openly available LLM types include OpenAI's GPT models, Google's recently released Bard, and Meta's LLaMA. Each Large Language Model has its own differences regarding their training, computational demand, economic cost, and capacity; however, these are out of the scope of this review.[24] Most of the studies and analyses regarding the use of LLMs in medicine have been conducted over OpenAI's products. OpenAI released the GPT-3 model in 2020. It was trained with over 175 billion parameters, sources being mostly internet texts and social media content. In March 2023, with support from

Microsoft, OpenAI released GPT-4, which is a multimodal language model. Although the amount of the parameters that were used to train the model has not been disclosed, it is expected to be over 1 trillion.[65] It has the capacity to analyze both textual and visual data (hence, multimodal), in contrast to GPT-3 which can only analyze and produce text. Furthermore, according to the OpenAI's overview, it can perform "advanced data analyses." Currently, OpenAI's ChatGPT models have over 1,5 billion visitors monthly.[66]

Even though they are not specifically trained using private or publicly available medical data including EHRs, opportunities offered by LLMs have drawn significant attention in the field of medicine in both clinical practice and research settings. In fact, as of August 2023, there were over a thousand research papers indexed in PubMed, only related to ChatGPT itself.[67] The applications of ChatGPT in medicine can be broadly categorized into clinical, research-related, and educational domains.[24] To be able to discuss the advantages, available tools, and disadvantages of LLMs, we chose to classify their applications by this categorization. In this section, we will delve into how LLMs can affect pediatricians in their clinical practices and researchers during their studies, including manuscript writing and peer-review process.

Diagnosis and Pediatric Care Settings

While it has gathered attention for its success in achieving passing grades in the United States Medical Licensing Examinations more than one time, ChatGPT's performance in handling patient queries, sub-specialty board examinations, and diagnosing specific clinical contexts has yielded mixed results in terms of feasibility and accuracy.[68] In a study focusing on pediatric urology cases, researchers assessed ChatGPT's responses to frequently asked questions gathered from various online sources. The outputs generated by the language model were generally satisfactory and aligned with current medical guidelines.[69] Similarly, in another study, ChatGPT was tasked with answering frequently asked questions encountered in pediatric cases, such as fever management, appropriate antipyretic dosages, and identification of red flag symptoms. The outputs were deemed moderately accurate and consistent.[70] Furthermore, in another study it was found that ChatGPT outputs to a general set of medical questions gathered from a public online forum were higher in quality and empathy compared to physician answers.[71] It should be noted that, especially in pediatric cases, diagnosis and patient management are highly complicated, individualistic, and long processes for which pediatricians spend years to be competent. Although there are several studies where outputs of ChatGPT were rated as "adequate" or "accurate" for simple clinical settings, in more complex cases feasibility of ChatGPT decreases significantly. As an example, ChatGPT has been shown to fail or produce suboptimal outputs in neonatal board examinations, answering complex questions for cardiovascular disease management and oncology cases.[72-74] Furthermore, even in the studies where ChatGPT outputs were deemed as "accurate" or "adequate," the rate of correct responses may not reflect the scale of accuracy needed in real-life practices. Assuming that ChatGPT can be used as a diagnostic or clinical decision-support tool in pediatrics merely because it exhibited proficiency in licensing examinations or demonstrated empathy in certain clinical queries is as absurd as assuming a compass can predict a hurricane's path solely based on its ability to point north. Such an assumption neglects the nuanced complexities inherent in medical decision-making, the dynamic nature of patient care, and the indispensable role of real-time clinical judgment, which hinges on multifaceted patient histories and comprehensive physical examinations. Until now there is not much study specifically conducted on the feasibility and accuracy of ChatGPT or any related LLMs for pediatric scenarios. However, from the studies that analyzed the outputs of ChatGPT on some clinical questions gathered from internet sources or produced by researchers, certain disadvantages can be identified that highlight these issues. These are summarized in Table 1.

The first obvious disadvantage of ChatGPT comes from the recency and content of the dataset that had been used to train the system. GPT 3,5 and GPT 4 were trained mostly using text generated up to September 2021.[24] Expectedly, the output created by the program is unable to transfer the most up-to-date information and knowledge, a major disadvantage that creates a big pitfall in pediatric practice, a specialty that changes frequently. Whether it is a minor shift in medical terminology or major updates like novel therapies or medications, this

Table 1. Major Limitations of ChatGPT as an Information Source in Medicine

Limitation	Description
Accuracy	- Models are not trained to understand the human language per se but rather they learn the probabilistic association between words. Hence, outputs do not always reflect the content accurately. - The dataset used to train GPT models is gathered from online sources, which include unconfirmed and unvalidated resources as well. - Outputs of the model contain "fabricated facts" irrelevant to the input.
Recency	- The dataset used to train GPT models does not include content after September 2021. - GPT-4 is said to be able to browse the internet however accuracy and functionality is not validated.
Transparency	- A major problem with any advanced AI model is the lack of transparency. It is not always clear how the system creates an output, using which associations, calculations, and algorithms (known as "black box" issues). - Lack of transparency brings other problems regarding accountability, reproducibility, and explainability.
Ethical concerns	- Ethical issues regarding the use of any AI model in medicine have not been made clear yet. - Lack of accountability, responsibility, governance; risks regarding privacy and security; potential biases, and discriminations should be kept in mind.

AI, artificial Intelligence; GPT, Generative Pre-trained Transformer.

limitation poses potential risks for public users. Furthermore, the content of the dataset that was used to train the system included publicly available information that was not shaped specifically for medical use, without cross-validation or validation of the outputs. Despite the extensive data used in training GPT-4, the lack of diverse, high-quality text within the dataset can lead to inaccuracies in the outputs, a situation which was previously described as "garbage in, garbage out" in the field of computer science and reflected in the studies that analyzed ChatGPT's output for different clinical reasons. Google has announced its new LLM called Med-PaLM, tailored for medical use. It has been shown that their model can also achieve a high score on USMLE. However, this model is still under development and has not been publicly available yet.[75]

A third and major disadvantage related to ChatGPT is incoherency and "fact fabrication." LLMs are not trained to understand language as humans do. They have the capacity to "learn" the statistical association between words as humans use them. They predict which word completes the sentence or the phrase.[76] Their ability to produce text in the context of a given prompt is truly remarkable. However, it has been observed that these models can produce correct sounding, coherently phrased but incorrect outputs, an incidence known as "fact fabrication" or "hallucination." As an example, in one study authors tasked ChatGPT to write a medical note after reading the transcript of a physician–patient outcome. Although the text generated by the system was clear and seemed coherent, it also contained fabricated BMI levels while in the original transcript there was no information on weight or height.[5] It should be also noted that interaction between the LLM and the user can stimulate to system's output to be better. As the prompt or instructions provided by the user increases in quality and clearance, the output also increases in accuracy and these prompts are essential for the better function of the LLMs.

Potential inaccuracies, lack of validation, nonspecific training, and issues regarding the recency of the dataset certainly warrant a cautious approach to the usage of ChatGPT as a diagnostic or clinical decision-support tool for the time being as these critical processes demand a reliable and precise approach, which currently may not align with ChatGPT's capabilities. However, it should not be overlooked that ChatGPT exhibits a strong competence in tasks where specialist knowledge is not required and user prompts are well-engineered to execute simple tasks such as assimilation, summarization, and rephrasing of given information.[24] This capability opens the door to utilizing the system for streamlining routine administrative tasks in pediatric care, including automated documentation, the creation of discharge or summary letters, and medical notetaking. This capability can improve workflow, fasten patient care, and decrease the burden on physicians which can create positive effects even on burn-out rates and efficiency.[5,77] Nevertheless, the utilization of personal data and the safeguarding of personal information remains a crucial concern. Currently, the training dataset of GPT-4 comprises publicly accessible personal information, presenting the potential for misuse and conflicting with legal rights, including the 'right to be forgotten.'[78] Consequently, constructing the prompts provided to ChatGPT should be done meticulously, prioritizing the protection and safety of patients without including any personal, identifiable data.

Research and Publishing

In the scope of pediatric research, the ability of AI to transform and fasten translational research and clinical trials has been discussed briefly in the previous sections. In this section, we will focus on the advantages and roles LLMs might have in publication and review processes.

In the scope of pediatric research, the potential of AI, particularly large language models (LLMs), to transform and expedite translational research and clinical trials has gained attention. It is evident that LLMs can effectively be utilized at various stages of the research and publishing process. Notably, their capacity to suggest research topics and hypotheses tailored to specific pediatric health concerns holds promise for researchers seeking innovative paths for investigation. However, as current systems are not trained with the most up-to-date knowledge, outputs may come short of this. Beyond hypothesis generation, LLMs can also significantly contribute to the development of robust methodologies and study designs, thus

Table 2. Examples of Academic AI Tools that Incorporate Large Language Models*

LLM	Definition
Scite	A web-based research management tool that helps researchers organize their literature, notes, and data. It has been developed by analyzing over 25 million full-text scientific articles and has a database of more than 800 million classified citation statements. It is defined as a 'smart citation index' and the model is based on Large Language Models and deep learning architectures.
ResearchRabbit	A publication discovery tool supported by AI. It lets researchers discover publications related to each other with the help of visualization maps and lists relevant publications according to different features.
Elicit	It is an AI-powered academic search engine that can streamline the process of finding, summarizing, and understanding research articles. It uses language models to automate workflow. Ideal for evidence synthesis, text extraction, and review process.
Perplexity.ai	It is an AI tool that combines web searches to produce ready-made answers to specific questions whilst citing the sources used. Best suited to identify sources. It uses OpenAI's GPT model to function.
Consensus	It is an AI-powered academic search engine that can streamline the process of finding, summarizing, and understanding research articles. It can produce specific outputs to structured questions such as 'Can iron deficiency anemia cause growth failure?'

AI, artificial intelligence; GPT, Generative Pre-trained Transformer.
*It should be noted that the delivery of these examples does not mean that authors advise or condemn their usage. They are given only for theoretical and comprehensive reasons.

streamlining the initial stages of research planning. With the capability to recommend suitable statistical methods and even generate codes for statistical analysis software such as R and Python, LLMs can potentially enhance the efficiency and accuracy of data analysis in pediatric studies. Moreover, during the manuscript writing phase, LLMs like ChatGPT can serve as invaluable aids, facilitating tasks such as translation, editing, and proofreading. By leveraging their language processing capabilities, LLMs can assist in refining the overall quality of scientific writing, ensuring clear and coherent communication of complex research findings. Considering the 2.5 times higher rejection rates faced by non-native English speakers in the peer-review process, implementing this function can become instrumental in fostering greater inclusivity and equity in scientific contributions.[79] Furthermore, the integration of LLMs in the peer review process has the potential to streamline the evaluation of scientific manuscripts. Their capacity to identify grammatical errors and inconsistencies, coupled with their ability to provide constructive feedback for improving the overall clarity and coherence of the content, can accelerate the review process and enhance the quality of published pediatric research. Beyond this, there are other LLMs that help authors to fasten literature scanning and citation processes summarized in Table 2.[80,81]

Scientists cannot ignore the fact that LLMs are more than just grammatical tools that have been implemented in common writing platforms for years. ChatGPT itself is capable of drafting an entire article with remarkable coherence. However, it should be noted that both the advantages and the challenges discussed in this section are solely theoretical. Whether or not they should be used for publication is still a question that remains to be answered. Ethical issues related to accountability, plagiarism, and the governance of AI tools and LLMs are essential issues that need to be addressed by the consortia of the scientific community. Fortunately, scientific journals have acted quickly to stop ChatGPT or other LLMs from qualifying as an author, as they cannot provide the accountability required for authorship and should be treated as another tool to assist humans in their work.[82,83] Furthermore, most of the journals have incorporated disclosure materials into their editorial policies on the use of AI tools.[52] In a recent study, authors examined the disclosure statements of 300 journals and found that 59% of them had specific statements on the use of AI tools. Of these, only a small number of them prohibited the use of these models for any purpose during the publication process, including the prestigious journals like *Science* and *Nature*.[84] Expectedly, those that do not prohibit their use expect authors to clearly state the manner, reason, and extent of the use of AI models. Currently, there are several tools available online assumed to have the capability to detect AI-generated texts and contents to a scale. However, like ChatGPT itself, these tools lack extensive validation of their capabilities and are not mentioned here.

CONCLUSION

As pediatricians, we are already witnessing the transformative power of AI-based applications in our daily clinical practice and research efforts. This rapid progress is set to revolutionize various aspects of pediatric healthcare, reshaping the way we diagnose, treat, create, and innovate. The integration of machine learning and deep learning algorithms has demonstrated great success in certain clinical settings for defined outcomes. New drug developments, better diagnostic and treatment approaches, and translational research opportunities are on the horizon. For each day, LLMs from different companies become more effective, accurate, and comprehensive. It becomes clear that collaborative efforts involving diverse professionals, including data scientists, engineers, bioethicists, and patients themselves, are crucial for realizing the full benefits of AI in pediatric care. However, recognizing and addressing the challenges associated with AI in this domain will require continued research, scrutiny, and interdisciplinary contributions.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – K.C.D., M.Y., S.S.; Design – K.C.D., M.Y., S.S.; Supervision – M.Y., S.S., N.C., Ö.K.; Data Collection and/or Processing – K.C.D., M.Y., S.S.; Analysis and/or Interpretation – K.C.D., M.Y., S.S., N.C., Ö.K.; Literature Search – K.C.D., M.Y., S.S.; Writing – K.C.D., M.Y., S.S., N.C.; Critical Review – M.Y., S.S., N.C., Ö.K.

Declaration of Interests: Özgür Kasapçapur is the editor-in chief, Nur Canpolat is a deputy editor, Seha Saygılı and Mehmet Yıldız are social media editors at the Turkish Archives of Pediatrics, however, their involvement in the peer review process was solely as an author and not as a reviewer. Kaan Can Demirbaş has no conflict of interest to declare.

Funding: This study received no funding.

REFERENCES

1. Lindberg DA. Medical informatics/computers in medicine. *JAMA*. 1986;256(15):2120-2122. [CrossRef]
2. Oosthuizen RM. The fourth Industrial Revolution – smart technology, artificial intelligence, robotics and algorithms: industrial psychologists in future workplaces. *Front Artif Intell*. 2022;5:913168. [CrossRef]
3. Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. *JAMA*. 2023;330(4):317-318. [CrossRef]
4. Saygılı S, Yıldız M. The ebb and flow of social media for researchers. *Turk Arch Pediatr*. 2023;58(5):456-457. [CrossRef]
5. 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(13):1233-1239. [CrossRef]
6. Naik N, Hameed BMZ, Shetty DK, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? *Front Surg*. 2022;9:862322. [CrossRef]
7. Amisha MP, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Fam Med Prim Care*. 2019;8(7):2328-2331. [CrossRef]
8. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. *Minim Invasive Ther Allied Technol*. 2019;28(2):73-81. [CrossRef]
9. Holman JG, Cookson MJ. Expert systems for medical applications. *J Med Eng Technol*. 1987;11(4):151-159. [CrossRef]
10. van Melle W. MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *Int J Man Mach Stud*. 1978;10(3):313-322. [CrossRef]
11. Miller RA, Pople HE, Jr, Myers JD. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med*. 1982;307(8):468-476. [CrossRef]

12. Weiss SM, Kulikowski CA, Amarel S, Safir A. A model-based method for computer-aided medical decision-making. *Artif Intell*. 1978;11(1-2):145-172. [CrossRef]
13. Gennatas ED, Chen JH. *Artificial Intelligence in Medicine: Technical Basis and Clinical Applications*. Cambridge: Academic Press; 2021:3-18.
14. Chen JH, Alagappan M, Goldstein MK, Asch SM, Altman RB. Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets. *Int J Med Inform*. 2017;102:71-79. [CrossRef]
15. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23(1):689. [CrossRef]
16. Buchanan BG. Can machine learning offer anything to expert systems? *Mach Learn*. 1989;4(3-4):251-254. [CrossRef]
17. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. *Supervised and Unsupervised Learning for Data Science*. Cham: Springer International Publishing; 2020:3-21.
18. Pettit RW, Fullem R, Cheng C, Amos CI. Artificial intelligence, machine learning, and deep learning for clinical outcome prediction. *Emerg Top Life Sci*. 2021;5(6):729-745. [CrossRef]
19. Kaur M, Gianey HK, Singh D, Sabharwal M. Multi-objective differential evolution based random forest for e-health applications. *Mod Phys Lett B*. 2019;33(5):1950022. [CrossRef]
20. Kumar Shukla P, Kumar Shukla P, Sharma P, et al. Efficient prediction of drug-drug interaction using deep learning models. *IET Syst Biol*. 2020;14(4):211-216. [CrossRef]
21. Rezaee K, Jeon G, Khosravi MR, Attar HH, Sabzevari A. Deep learning-based microarray cancer classification and ensemble gene selection approach. *IET Syst Biol*. 2022;16(3-4):120-131. [CrossRef]
22. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform*. 2017;18(5):851-869. [CrossRef]
23. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. *Trends Anaesth Crit Care*. 2021;38:4-9. [CrossRef]
24. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-1940. [CrossRef]
25. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. *JAMA*. 2023;330(9):866-869. [CrossRef]
26. Looking back on the millennium in medicine. *N Engl J Med*. 2000;342(1):42-49. [CrossRef]
27. Misra SC, Mukhopadhyay K. Data harnessing to nurture the human mind for a tailored approach to the child. *Pediatr Res*. 2023;93(2):357-365. [CrossRef]
28. Sugiyama K, Hasegawa Y. Computer assisted medical diagnosis system for inborn errors of metabolism. *Jpn J Electron Biol Eng*. 1984;22:942-943.
29. Malhotra A, Molloy EJ, Bearer CF, Mulkey SB. Emerging role of artificial intelligence, big data analysis and precision medicine in pediatrics. *Pediatr Res*. 2023;93(2):281-283. [CrossRef]
30. Hoodbhoy Z, Masroor Jeelani S, Aziz A, et al. Machine learning for child and adolescent health: a systematic review. *Pediatrics*. 2021;147(1). [CrossRef]
31. Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med*. 2019;25(3):433-438. [CrossRef]
32. Khazaei H, Mench-Bressan N, McGregor C, Pugh JE. Health informatics for neonatal intensive care units: an analytical modeling perspective. *IEEE J Transl Eng Health Med*. 2015;3:3000109. [CrossRef]
33. McAdams RM, Kaur R, Sun Y, Bindra H, Cho SJ, Singh H. Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: a systematic review. *J Perinatol*. 2022;42(12):1561-1575. [CrossRef]
34. Helguera-Repetto AC, Soto-Ramírez MD, Villavicencio-Carrisoza O, et al. Neonatal sepsis diagnosis decision-making based on artificial neural networks. *Front Pediatr*. 2020;8:525. [CrossRef]
35. Lin YC, Salleb-Aouissi A, Hooven TA. Interpretable prediction of necrotizing enterocolitis from machine learning analysis of premature infant stool microbiota. *BMC Bioinformatics*. 2022;23(1):104. [CrossRef]
36. Verder H, Heiring C, Ramanathan R, et al. Bronchopulmonary dysplasia predicted at birth by artificial intelligence. *Acta Paediatr*. 2021;110(2):503-509. [CrossRef]
37. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136(7):803-810. [CrossRef]
38. Chen KC, Yu HR, Chen WS, et al. Diagnosis of common pulmonary diseases in children by X-ray images and deep learning. *Sci Rep*. 2020;10(1):17374. [CrossRef]
39. Summers RM. Deep learning lends a hand to pediatric radiology. *Radiology*. 2018;287(1):323-325. [CrossRef]
40. Meda KC, Milla SS, Rostad BS. Artificial intelligence research within reach: an object detection model to identify rickets on pediatric wrist radiographs. *Pediatr Radiol*. 2021;51(5):782-791. [CrossRef]
41. Zhou M, Wu K, Yu L, et al. Development and evaluation of a leukemia diagnosis system using deep learning in real clinical scenarios. *Front Pediatr*. 2021;9:693676. [CrossRef]
42. Khammad V, Otero JJ, Cabello Izquierdo Y, et al. *Application of Machine Learning Algorithms for the Diagnosis of Primary Brain Tumors*. American Society of Clinical Oncology; 2020.
43. Porras AR, Rosenbaum K, Tor-Diez C, Summar M, Linguraru MG. Development and evaluation of a machine learning-based point-of-care screening tool for genetic syndromes in children: a multinational retrospective study. *Lancet Digit Health*. 2021;3(10):e635-e643. [CrossRef]
44. Molloy EJ, Bearer CF. Translational research is all-encompassing and lets everyone be a researcher. *Pediatr Res*. 2021;90(1):2-3. [CrossRef]
45. Biswas N, Chakrabarti S. Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer [Review]. *Front Oncol*. 2020;10:588221. [CrossRef]
46. Neumann E, Schreeck F, Herberg J, et al. How paediatric drug development and use could benefit from OMICs: a c4c expert group white paper. *Br J Clin Pharmacol*. 2022;88(12):5017-5033. [CrossRef]
47. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18(1):83. [CrossRef]
48. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinformatics Biol Insights*. 2020;14:1177932219899051. [CrossRef]
49. Li C, Sullivan RE, Zhu D, Hicks SD. Putting the "mi" in omics: discovering miRNA biomarkers for pediatric precision care. *Pediatr Res*. 2023;93(2):316-323. [CrossRef]
50. Sharifi-Noghabi H, Zolotareva O, Collins CC, Ester M. MOLI: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics*. 2019;35(14):i501-i509. [CrossRef]
51. Cascini F, Beccia F, Causio FA, Melnyk A, Zaino A, Ricciardi W. Scoping review of the current landscape of AI-based applications in clinical trials. *Front Public Health*. 2022;10:949377. [CrossRef]
52. Flanagin A, Kendall-Taylor J, Bibbins-Domingo K. Guidance for authors, peer reviewers, and editors on use of AI, language models, and chatbots. *JAMA*. 2023;330(8):702-703. [CrossRef]
53. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-1374. [CrossRef]
54. Faris O, Shuren J. An FDA viewpoint on unique considerations for medical-device clinical trials. *N Engl J Med*. 2017;376(14):1350-1357. [CrossRef]

55. Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. *JAMIA Open*. 2020;3(3):326-331. [CrossRef]
56. Hunter DJ, Holmes C. Where medical statistics meets artificial intelligence. *N Engl J Med*. 2023;389(13):1211-1219. [CrossRef]
57. Cutillo CM, Sharma KR, Foschini L, et al. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. *npj Digit Med*. 2020;3(1):47. [CrossRef]
58. Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. *IEEE Intell Syst*. 2009;24(2):8-12. [CrossRef]
59. Klie A, Tsui BY, Mollah S, et al. Increasing metadata coverage of SRA BioSample entries using deep learning–based named entity recognition. *Database (Oxford)*. 2021;2021. [CrossRef]
60. Panesar SS, D'Souza RN, Yeh FC, Fernandez-Miranda JC. Machine learning versus logistic regression methods for 2-year mortality prognostication in a small, heterogeneous glioma database. *World Neurosurg X*. 2019;2:100012. [CrossRef]
61. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA*. 2019;322(24):2377-2378. [CrossRef]
62. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. [CrossRef]
63. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med*. 2021;181(8):1065-1070. [CrossRef]
64. Annapragada AV, Donaruma-Kwoh MM, Annapragada AV, Starosolski ZA. A natural language processing and deep learning approach to identify child abuse from pediatric electronic medical records. *PLoS One*. 2021;16(2):e0247404. [CrossRef]
65. Venerito V, Bilgin E, Iannone F, Kiraz S. AI am a rheumatologist: a practical primer to large language models for rheumatologists. *Rheumatol (Oxf Engl)*. 2023;62(10):3256-3260. [CrossRef]
66. OpenAI. ChatGPT, Overview. Available at: https://openai.com/chatgpt. (Accessed 28 August, 2023).
67. Temsah MH, Altamimi I, Jamal A, Alhasan K, Al-Eyadhy A. ChatGPT surpasses 1000 publications on PubMed: envisioning the road ahead. *Cureus*. 2023;15(9):e44769. [CrossRef]
68. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digit Health*. 2023;2(2):e0000198. [CrossRef]
69. Caglar U, Yildiz O, Meric A, et al. Evaluating the performance of ChatGPT in answering questions related to pediatric urology. *J Pediatr Urol*. 2023. [CrossRef]
70. Kao HJ, Chien TW, Wang WC, Chou W, Chow JC. Assessing ChatGPT's capacity for clinical decision support in pediatrics: a comparative study with pediatricians using KIDMAP of Rasch analysis. *Med (Baltim)*. 2023;102(25):e34068. [CrossRef]
71. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social Media Forum. *JAMA Intern Med*. 2023;183(6):589-596. [CrossRef]
72. Beam K, Sharma P, Kumar B, et al. Performance of a large language model on practice questions for the neonatal board examination. *JAMA Pediatr*. 2023;177(9):977-979. [CrossRef]
73. Cocci A, Pezzoli M, Lo Re M, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. *Prostate Cancer Prostatic Dis*. 2023.
74. Nastasi, A.J., Courtright, K.R., Halpern, S.D. et al. A vignette-based evaluation of ChatGPT's ability to provide appropriate and equitable medical advice across care contexts. *Sci Rep* 13, 17885 (2023). [CrossRef]
75. Google. Med-PaLM. Available at: https://sites.research.google/med-palm/. (Accessed November 10, 2023).
76. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst*. 2020;33:1877-1901.
77. Nayak A, Alkaitis MS, Nayak K, Nikolov M, Weinfurt KP, Schulman K. Comparison of history of present illness summaries generated by a chatbot and senior internal medicine residents. *JAMA Intern Med*. 2023;183(9):1026-1027. [CrossRef]
78. OpenAI. GPT-4 system card. Available at: https://cdn.openai.com/papers/gpt-4-system-card.pdf. (Accessed 7th of November, 2023)
79. Amano T, Ramírez-Castañeda V, Berdejo-Espinola V, et al. The manifold costs of being a non-native English speaker in science. *PLoS Biol*. 2023;21(7):e3002184. [CrossRef]
80. Consensus. AI search engine for research. Available at: https://consensus.app/search/. (Accessed 7 November, 2023)
81. Scite. AI for Research. Available at: https://scite.ai. (Accessed 7 November, 2023)
82. Gaggioli A. Ethics: disclose use of AI in scientific manuscripts. *Nature*. 2023;614(7948):413. [CrossRef]
83. Thorp HH. ChatGPT is fun, but not an author. *Science*. 2023;379(6630):313-313. [CrossRef]
84. Lund BD, Naheem KT. Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals. *Learn Publ*. 2024;37(1):13-21. [CrossRef]

Supplementary Table 1. Glossary Box	
Artificial intelligence	Refers to the development of computer systems that can perform tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, and natural language understanding. It covers machine learning, deep learning and natural language processing under its umbrella.
Machine learning	It is a subset of artificial intelligence that focuses on enabling computers to learn from data. Instead of being explicitly programmed, ML algorithms use statistical techniques to improve their performance on a specific task over time. According to the techniques used in the model, it can be classified into different models.
Artificial neural network	An artificial neural network is a computational model inspired by the structure and function of the human brain. It consists of interconnected nodes (neurons) organized in layers, allowing the network to learn complex patterns and relationships from data. They serve as the foundation for more advanced machine learning techniques, such as deep learning, which involves neural networks with many hidden layers, enabling them to automatically learn hierarchical representations of data.
Deep learning	Deep Learning is an advenced subfield of Machine Learning that involves neural networks with many layers (deep neural networks). It has been particularly successful in tasks such as image and speech recognition, and it allows systems to automatically learn hierarchical representations of data.
Natural language processing	Natural Language Processing (NLP) is a branch of AI that focuses on the interaction between computers and human languages. NLP enables computers to understand, interpret, and generate human language, facilitating communication between machines and humans.
Large language models	Large Language Models are advanced AI models that are trained on massive datasets to understand and generate human-like language. They have applications in various NLP tasks and can exhibit a high level of language understanding and generation.
GPT	Generative Pre-trained Transformer (GPT) is a type of large language model developed by OpenAI. It utilizes a transformer architecture and is pre-trained on diverse datasets, allowing it to perform various natural language processing tasks, such as text completion, translation, and summarization. Chat-GPT is a variant of GPT model fine-tuned specifically for generating conversational speech. GPT models are applied to different LLM variants for different (including scientific research) purposes.
BARD	Building AutoML with Reinforcement Learning (BARD) is a large language model developed by Google AI. It is trained on a massive dataset of text and code. It has been empowered with a new AI model called the 'Gemini' that can also exert image synthesis and production. BARD has the same essential features as an LLM such as natural language processing, translation, summarization etc.
Prompt	In the scope of artificial intelligence and large language models, prompt refers to the input text or query that is provided to the model to guide its response. The prompt essentially sets the initial context for the model's generation process and can have a significant impact on the output.

AI, Artificial Intelligence; GPT, Generative Pre-trained Transformer; LLM, Large Language Model; NLP, Natural Language Processing.