



REVIEW

Open Access

Virtual medicine: medical AI in human health and diseases

Chen Zhang^{1,2,3,4†}, Fu-Xiao Wang^{1,2,3,4†}, Xuan Tang^{1,2,3,4†}, Ji-Long Li^{1,2,3,4}, Ning Ding^{1,2,3,4}, Yang Hong^{1,2,3,4}, Pei-Ran Song^{1,2,4*}, Long Bai^{1,2,4,5*} and Jia-Can Su^{1,2,4,6*}

Abstract

The evolution of medicine has progressed through distinct ages: from empirical observation and evidence-based practice to the current era of precision medicine. However, traditional healthcare paradigms remain constrained by data fragmentation, scalability limits, and the overwhelming complexity of multi-omics integration. In the rapid explosion of artificial intelligence (AI), a transformative paradigm is emerging. This review introduces the concept of “Virtual Medicine”, which is defined as a comprehensive ecosystem of AI-empowered medical practice that transcends physical limitations. This review systematically summarizes the technological foundations, historical evolution, and core applications of AI in medicine, including electronic health records (EHRs) analysis, medical imaging, multimodal diagnostics, drug discovery, precision oncology, intelligent surgery, and clinical decision support systems. It further highlights the role of medical AI in health management, public health surveillance, and healthcare delivery in resource-limited settings. Special attention is given to the transformative emergence of large language models (LLMs), such as medical large language models (MedLLM) and generative pre-trained transformer (GPT) architectures, emphasizing their potential to revolutionize virtual medical interaction, clinical reasoning, and documentation. Despite these advances, significant challenges remain regarding model transparency, data bias, fairness, and patient privacy. Overcoming these limitations necessitates standardized evaluation frameworks, interpretable algorithm designs, and strengthened privacy protections. Ultimately, these efforts aim to foster a trustworthy and equitable future for virtual medicine.

Key words Medical artificial intelligence (AI), Clinical decision support, Virtual Medicine, Precision medicine, Large language models

Background

Faced with an aging global population, a growing burden of chronic diseases, and unequal access to healthcare resources, traditional healthcare systems are increasingly struggling in terms of efficiency, diagnostic accuracy, and resource allocation [1,2]. These constraints highlight the urgency of a paradigm shift toward “Virtual Medicine”, a digitalized ecosystem that integrates advanced technologies to optimize disease prevention, diagnosis, and treatment (Fig. 1).

Over the past decades, artificial intelligence (AI) has evolved from a conceptual framework to a practical engine for medical innovation. Advances in machine learning (ML) and natural language processing have catalyzed applications in areas such as medical imaging, clinical prediction, drug

development, and decision support [3]. Medical AI systems now assist in identifying subclinical lesions, triaging high-risk patients, and tailoring personalized treatment plans, surpassing conventional performance in several diagnostic tasks and accelerating clinical workflows [4,5]. A study on diabetic retinopathy screening in Africa has achieved an automatic interpretation sensitivity of more than 97%, providing accurate diagnosis for residents under poor medical conditions [6]. Alongside these advances, critical challenges persist. Model opacity undermines interpretability and clinical trust, biased training datasets risk exacerbating health disparities, and the reliance on patient data raises profound concerns over privacy and data security [7]. These issues require rigorous validation, ethical safeguards, and systematic model governance to ensure safe, transparent, and equitable deployment [8].

Several works have discussed the role of AI in medicine. For instance, Rajpurkar *et al.* [3] summarized technical progress and validation studies, particularly in imaging and molecular applications, while He *et al.* [8] focused on the

[†]Chen Zhang, Fu-Xiao Wang, and Xuan Tang contributed equally to this work
^{*}Correspondence: Pei-Ran Song, songpeiran@163.com; Long Bai, bailong@shu.edu.cn; Jia-Can Su, drsujiacan@163.com
¹Organoid Research Center, Institute of Translational Medicine, Shanghai University, Shanghai 200444, China
Full list of author information is available at the end of the article

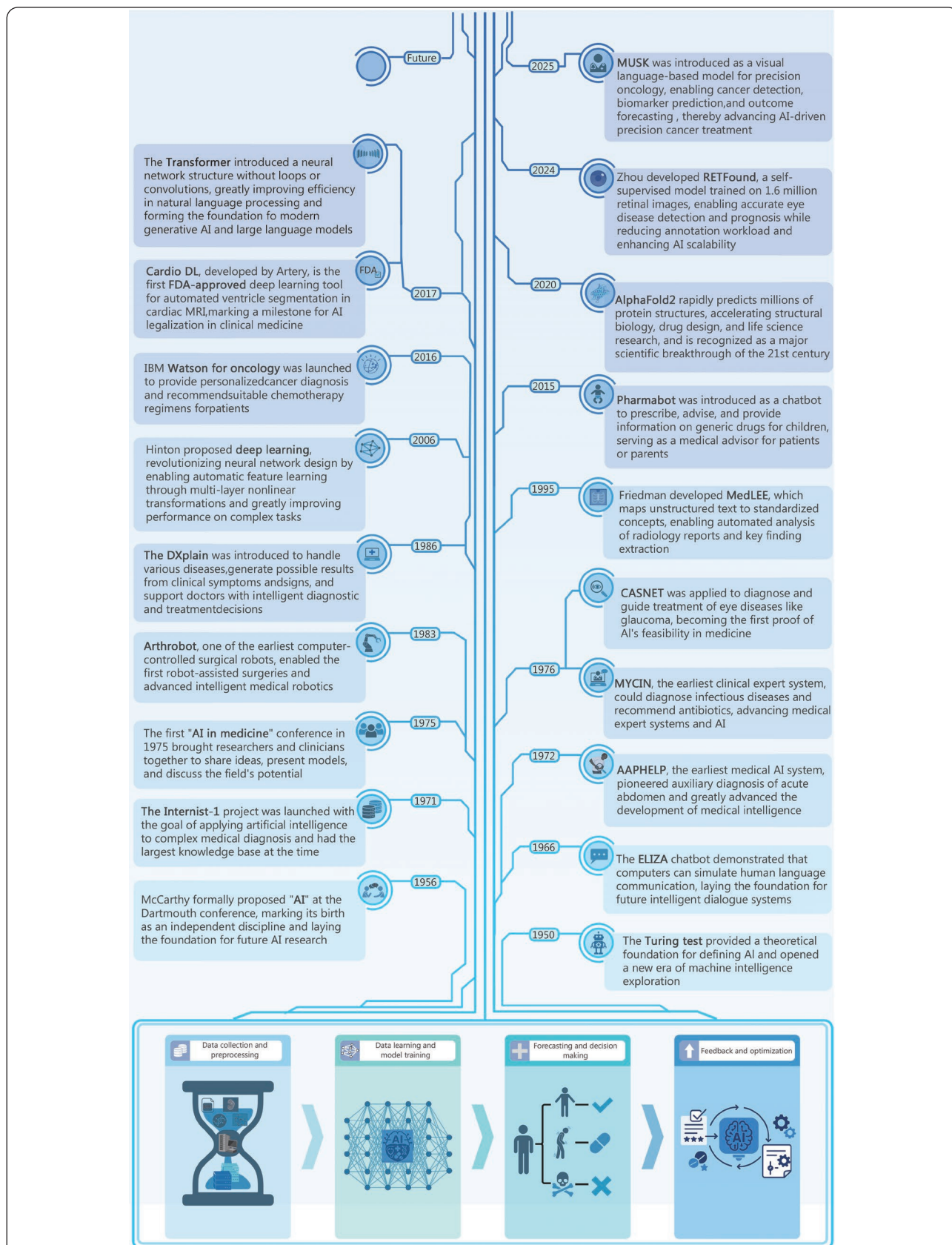


Fig. 2 The historical development and decision-making principles of medical artificial intelligence (AI).

DL. Deep learning; FDA. U.S. Food and Drug Administration; MRI. Magnetic resonance imaging; IBM. International business machines corporation; MedLEE. Medical language extraction and encoding system

marking the birth of modern AI research [10]. However, explorations into AI began even earlier. In 1943, American psychologists McCulloch and mathematician Pitts proposed the mathematical model of binary neurons, inspired by the workings of biological neurons, thus laying the groundwork for neural network theory [14]. Alan Turing introduced the Turing Test in 1950, which remains a classic benchmark for assessing AI.

Building on this foundational work, AI experienced rapid development. Nine years after the Dartmouth conference, the chatbot ELIZA was introduced as an early exploration in natural language processing, paving the way for future medical consultation systems [15]. Released in 1976, the MYCIN system represented the first implementation of an AI-based expert system in medicine [16]. Developed at Stanford University, MYCIN focused on diagnosing infectious diseases and selecting appropriate antibiotics. By inputting patient information into a computer, MYCIN could use logical inference to recommend optimal antibiotic regimens, showcasing the clinical potential of AI for the first time [16]. Similarly, DXplain, introduced in 1986, was another influential medical expert system based on probabilistic models [17]. It integrated information across multiple medical fields to provide broad differential diagnoses for clinical symptoms and signs, thus supporting clinicians in decision-making [17].

With the advent of the 21st century, AI technology experienced significant expansion due to breakthroughs in deep learning (DL) [18], extending its medical applications beyond diagnostic support to areas such as medical image analysis, personalized therapy, and public health management [19]. DL theory, proposed by Hinton *et al.* [20] in 2006, revolutionized neural network architecture by enabling automatic feature extraction through multi-layer nonlinear transformations, greatly enhancing model capacity for complex tasks.

In 2015, the pharmabot conversational chatbot was born, trained to prescribe, advise, and provide information about drugs for children. They introduced an application that can act as a medical advisor to patients or parents who are confused about generic drugs [21]. In 2017, Cardio DL, developed by Artery, became the first U.S. Food and Drug Administration (FDA)-approved DL-based medical application, automating cardiac magnetic resonance imaging (MRI) segmentation with accuracy on par with experienced clinicians [22,23]. This regulatory milestone sets the stage for the legitimization of AI in clinical medicine. That same year, the introduction of the transformer architecture created a structure that does not require loops and convolutions,

greatly improving the efficiency and performance in tasks such as Natural Language Processing (NLP), and laying the foundation for the development of modern GPT and LLMs [24]. In 2020, AlphaFold2 made a landmark breakthrough in protein structure prediction, solving the long-standing protein folding problem and accelerating the development of peptide drugs and structural biology [25]. In 2024, Zhou *et al.* [26] developed RETFound, which can efficiently perform eye disease detection and prognosis tasks, reduce the burden of expert annotation, and enhance the scalability and generalization of AI in retinal imaging and broader clinical applications. In 2025, the MUSK model was introduced as a visual language-based model for precision oncology [26]. MUSK is capable of performing cancer detection, molecular biomarker prediction, and predicting clinical outcomes such as recurrence, prognosis, and immunotherapy response, thereby significantly improving the accuracy and universality of AI in precision cancer treatment [27].

Over these decades, through continuous algorithmic innovation and advances in sensing technologies, AI has gradually attained sophisticated autonomous learning and decision-making capabilities, driving the development of modern medicine towards personalization and intelligence [28].

The basic decision-making process of medical AI

The decision-making process of AI in medicine typically encompasses 4 main stages: data collection and preprocessing, data learning and model training, prediction and decision-making, and feedback and optimization [29].

Medical data is often sourced from a variety of records, including Electronic Health Records (EHRs), medical images, genomic sequences, and real-time monitoring signals [30-33]. Such data is characterized by its multi-source and diverse nature, with both structured data and unstructured data coexisting. Given the variability in data type and quality, preprocessing is crucial to enable effective AI analysis. During this phase, missing and unusual values are addressed through data cleaning, image denoising techniques are applied to reduce interference in medical imaging, normalization is achieved via standardization of units and formats, and feature engineering extracts key indicators, converting raw data into a computable form [34].

Data learning and model training constitute the core steps in transforming medical data into AI models. Using appropriate algorithms, AI systems can identify disease patterns and risk factors from preprocessed medical data. For example, in disease diagnosis, medical AI applies ML algorithms such as support vector machines (SVMs), regression analysis,

or decision trees to discover relationships between clinical features and symptoms from large patient groups, thereby generating predictive outcomes for clinicians [35,36]. In this process, datasets are divided into training and validation sets; the model repeatedly adjusts its parameters to minimize prediction error and achieve optimal fitting. Medical AI may use unsupervised and reinforcement learning to further refine its decision-making, enabling effective prediction and decision support even when encountering novel or previously unseen cases [37,38].

Prediction and decision-making form the core of medical AI applications. Trained models infer outcomes from new data and support clinical decision-making. For instance, an AI system may analyze digital mammography images to predict the malignancy probability of a tumor [39], or anticipate a patient's drug response based on genomic information and medical history [40]. Throughout this process, medical AI decisions incorporate confidence estimation and explainability techniques to ensure clinical reliability, and may synthesize results from multiple models or rule-based systems to deliver high-confidence, personalized recommendations [41].

Feedback and optimization are essential to the continuous improvement of AI systems. After clinical validation, differences between AI recommendations and real-world outcomes are fed back into the model, triggering retraining or parameter adjustment to iteratively refine the system. As more data accumulates, medical AI models progressively enhance their decision quality, matching more closely with actual clinical needs [42].

Through these stages, medical AI establishes a closed-loop system from raw data to clinical decision-making. The key strength lies in data-driven optimization, translating statistical patterns into actionable medical knowledge that supports clinicians in achieving greater accuracy and efficiency in diagnosis and treatment.

Advantages of medical AI

As the volume of medical data continues to expand, AI demonstrates multiple advantages within healthcare. Medical AI can rapidly process and analyze massive datasets, extracting hidden patterns through DL algorithms to provide a scientific basis for disease diagnosis and precision therapy [43]. The efficiency and accuracy of medical AI reduce errors associated with manual operations, thereby improving the reliability and safety of medical workflows [44]. Medical AI excels in real-time monitoring, disease progression prediction, and therapeutic optimization, offering clinicians more precise and personalized treatment support [45].

With continuous learning and self-improvement, medical AI systems refine their models through the accumulation of clinical cases, maintaining a leading edge in diagnostic and therapeutic capabilities. Unlike traditional experience-based approaches, medical AI utilizes objective clinical data, enabling rapid advancement in both diagnostic speed and accuracy. The automation and standardization afforded by medical AI allow it to assist clinicians in repetitive and highly regulated tasks such as image interpretation and laboratory result analysis, helping to reduce the burden on doctors [46].

In the context of population aging and unequal distribution of medical resources, medical AI not only alleviates the workload of healthcare professionals but also optimizes the patient care experience [47]. Through efficient data analysis, ongoing technological improvement, and process automation, medical AI is setting the stage for the future of medicine and is set to become an indispensable driving force in healthcare.

The application of medical AI in disease diagnosis **AI-based disease diagnosis using EHRs**

EHRs are digital versions of patients' medical histories maintained over time, including comprehensive clinical details including patient information, diagnoses, medications, treatment plans, immunizations, allergies, radiology images, and laboratory results [48]. They offer substantial value in healthcare delivery, ensuring continuity of patient care and reducing medical errors by enabling patient information sharing among healthcare providers [49].

Integrating EHRs with AI enables earlier prediction and identification of diseases that are difficult to diagnose. Utilizing ML and NLP techniques, AI algorithms can analyze structured and unstructured data within EHRs, identifying subtle clinical patterns indicative of early-stage diseases [50]. For instance, AI algorithms using NLP on unstructured clinical notes significantly enhance early detection of sepsis, a condition known for subtle early symptoms but rapid progression. The SERA algorithm, utilizing structured and unstructured EHR data, has demonstrated a high predictive accuracy 12 h before clinical diagnosis, substantially surpassing traditional clinical assessments. This algorithm notably improved sensitivity and reduced false positives, critical factors in timely sepsis management [51].

In oncology, medical AI integrated with EHRs data facilitates continuous pan-cancer prognostication, significantly improving early cancer detection and personalized management strategies. For example, the MEDomics platform incorporates multimodal health data from EHR systems to create individualized patient profiles [52]. These profiles have

successfully predicted hospitalizations and emergency visits in cancer patients by analyzing structured clinical variables and unstructured clinical notes using NLP. This approach has led to clear improvements in patient outcomes, including reduced emergency visits and hospital admissions, through earlier identification and management of complications [53].

Pediatric medicine also benefits substantially from AI-enhanced EHR analysis. AI algorithms have shown considerable effectiveness in diagnosing complex pediatric conditions, where early clinical signs might not be overtly discernible. These models integrate multiple clinical variables, enhancing diagnostic precision and facilitating earlier therapeutic interventions, which can markedly improve long-term pediatric health outcomes [54].

Despite these advances, using EHRs data for predictive analytics encounters inherent challenges. Biases originating from the healthcare process, such as variations in clinician documentation practices, incomplete records, and inconsistent coding of clinical information, pose significant obstacles [55]. Additionally, missing data points and inaccuracies in unstructured notes can reduce predictive reliability [56]. To mitigate these issues, employing advanced data optimization techniques, including robust data imputation methods, NLP-driven error correction, and standardized data entry protocols, is crucial. The rigorous validation of EHR-derived predictive models across multiple institutions further ensures broad applicability and reliability.

Future directions should emphasize enhancing data governance, adopting standardized ontologies, and fostering collaborative, multi-institutional data sharing. These strategies will significantly refine EHRs' data quality, advancing AI-driven predictive capabilities and fostering robust, real-time clinical decision support systems [57].

AI-based disease diagnosis using medical imaging

During diagnosis and treatment, clinicians rely heavily on imaging modalities such as X-rays, MRI, computed tomography (CT), and electrocardiograms (ECGs) to visualize internal anatomical and physiological changes [58]. These images provide crucial, intuitive insights into disease progression. However, the large amount of imaging data, combined with subtle early-stage anomalies that may be missed by human observers, presents significant diagnostic challenges and workload burdens.

To address these limitations, AI technologies, especially convolutional neural networks (CNNs) and advanced image preprocessing, are being applied to medical imaging [59]. These algorithms exhibit high efficiency, sensitivity, and

standardization, often matching or even surpassing expert radiologists in diagnostic performance. For instance, in breast cancer screening, a DL model using mammography and digital breast tomosynthesis outperformed 5 full-time breast imaging specialists, boosting sensitivity by 14% [60]. Similarly, AI-enhanced CT analysis has demonstrated high accuracy in diagnosing COVID-19 pneumonia, even in early or atypical cases, and provided predictive insights into disease severity and progression [61].

Medical AI is also revolutionizing cardiovascular diagnostics. In ECG analysis, DL models trained on 90,000 ECG-echocardiogram pairs identified asymptomatic left ventricular dysfunction, equaling the accuracy of standard echocardiographic methods [62]. Likewise, Cardiac magnetic resonance imaging (CMR) has been accelerated by medical AI to automatically screen for and diagnose multiple cardiovascular conditions with a performance that in some cases exceeds that of experienced cardiologists [63].

These advancements are not standalone, and hybrid medical AI workflows are proving especially effective. For example, decision-referral systems in breast cancer screening allow AI to automatically triage high-certainty cases, while uncertain evaluations are directed to radiologists. This collaboration improved both sensitivity and specificity beyond what either could achieve alone [64]. Moreover, AI can also repair artifacts and damage in defective three-dimensional (3D) biomedical imaging and reconstruct it, truly restoring tissue structure and cell distribution, greatly improving the spatial continuity and quantitative analysis accuracy of 3D reconstruction [65].

The above examples demonstrate the significant advantages of medical AI in imaging diagnostics. It can efficiently process large-scale datasets, enhance the sensitivity and accuracy of early abnormality detection, and achieve diagnostic performance comparable to or even surpassing that of human experts across multiple disease contexts. In addition, AI systems are capable of repairing defective images, thereby improving diagnostic reliability. However, notable limitations remain. The training of AI models for image recognition relies heavily on large, high-quality annotated datasets, and inconsistencies in annotation standards hinder clinical translation and broader deployment of medical AI.

AI-based disease diagnosis using multimodal data

Relying on one type of data, whether textual or visual, to diagnose complex diseases is not sufficient for comprehensive healthcare solutions. Medical diagnoses are inherently multifactorial and often require the integration of various data sources for accurate predictions. Therefore, multimodal

medical AI models, which combine information from multiple sources including clinical laboratory results, medical imaging, genomics, and patient demographics, provide a more complete method for disease diagnosis and prediction [66,67].

Multimodal AI has also demonstrated significant value in the diagnosis of dementia. Differentiating between various types of dementia, such as Alzheimer's disease (AD), vascular dementia, and Lewy body dementia, remains a major challenge due to the overlapping symptoms. Multimodal medical AI models with the ability to analyze neuroimaging, genetic markers, neuropsychological testing, and clinical history can diagnose multiple causes of dementia. For example, models trained using multimodal datasets of neuroimaging and clinical datasets can accurately differentiate between AD and vascular dementia, showing significant improvement over single-modality approaches [68]. Additionally, the Medical Multimodal Multitask Foundation Model (M3FM) can be used for lung cancer screening, and clinical teams apply M3FM for comprehensive risk assessment. The model integrates CT images, smoking records, family cancer history, and laboratory test results to assess lung cancer risk [69]. Furthermore, the M3FM extends to other related tasks, such as cardiovascular disease (CVD) mortality risk prediction and incidental finding detection [69].

The advantage of such multimodal models is that they can adapt to diverse data sets, thereby providing more accurate and personalized predictions [70]. While traditional methods may overlook key features due to the limitations of unimodal inputs, multimodal models mimic the diagnostic methods of human clinicians. Multimodal medical AI models can not only improve diagnostic accuracy but also assist doctors in making better decisions [71]. Improvements in these models mark the future of medical diagnostics. Their ability to combine diverse datasets allows for a more robust and individualized understanding of health, thus advancing precision medicine.

The applications of medical AI in disease treatment

AI-enabled drug discovery and repurposing

AI plays a crucial role in predicting disease occurrence and enhancing therapeutic strategies. Drug therapy remains a foundation of clinical treatment, yet identifying effective drugs for complex diseases is often slow and expensive [72]. Therefore, accelerating the drug discovery process is an urgent need. One of the most influential areas where AI contributes is in target screening and early drug discovery. Medical AI can process large-scale biological data to identify disease-relevant targets and predict ligand-receptor interactions [73]. DL algorithms are used to evaluate molecular properties and

binding affinities, improving the efficiency of virtual screening. Additionally, AlphaFold has revolutionized protein structure prediction by generating high-confidence 3D models of proteins, including those without experimental structures [74]. These predicted structures allow researchers to perform structure-based drug design at scale, supporting faster identification of candidate molecules and reducing reliance on expensive laboratory-based assays.

Medical AI also facilitates the discovery of new structures of antibiotics, addressing the global issue of antibiotic resistance. Traditional antibiotics are losing efficacy due to widespread overuse, and new structural classes are urgently needed [75]. Recent studies have applied graph neural networks (GNNs) to explore chemical spaces containing millions of compounds [76,77]. These models identify substructures associated with antimicrobial activity and low cytotoxicity. In a notable example, explainable AI was used to screen over 12 million compounds and predict new antibiotic candidates. The medical AI successfully identified structural classes selective against drug-resistant pathogens, demonstrating activity in both *in vitro* and *in vivo* models [78].

Another critical application of medical AI lies in drug repurposing. Repurposing existing drugs for new indications can reduce development time and cost, offering a viable solution for diseases with limited therapeutic options [79]. The TxGNN model demonstrates this approach, using a knowledge graph to learn relationships between drugs, targets, and diseases. It supports zero-shot predictions and can identify drug candidates for both treated and untreated diseases. In addition, the model is able to generate interpretable results that are consistent with expert clinical decisions [80].

In short, medical AI is transforming drug development and disease treatment. It accelerates the discovery of targets and candidate molecules, speeds up the development of new antibiotics, and streamlines effective drug repurposing processes.

AI-enabled precision medicine

Precision medicine is an evolving healthcare approach where therapies are customized for each person according to the specific features of their condition, including molecular profiles [81]. This field has become increasingly significant in oncology, where the complexity of cancer demands personalized therapies to improve outcomes. The core principle of precision medicine is to customize therapy guided by an individual's genomic profile and the unique properties of their tumor, and other factors, thus optimizing treatment effectiveness and minimizing side effects [82]. In cancer care, this approach

holds the capacity to revolutionize how oncologists diagnose and treat the disease by focusing on the molecular and genetic basis of cancer rather than the traditional one-size-fits-all model. However, the successful application of precision medicine hinges on the integration and analysis of massive and complex datasets from multiple domains, such as genomics, transcriptomics, proteomics, and metabolomics [83,84]. These omics technologies, clinical information, and molecular biomarkers can provide a comprehensive picture of the tumor and its microenvironment and provide important information about the disease's response to treatment.

Manual evaluation of these datasets is time-consuming, error-prone, and expensive. The lack of standardized procedures can also cause inconsistent results and hinder the scalability of precision medicine. This is where medical AI steps in, offering remarkable capabilities for processing and analyzing large datasets. AI's ability to discover connections between data from multiple perspectives can accelerate the development of platforms that support not only diagnostic confidence but also personalized treatment approaches. DL and ML algorithms are particularly good at processing the massive multi-omics datasets generated by cancer research [85].

AI medical models show promise in early cancer detection using multi-omics analysis. The CancerSEEK test combines genomic and protein biomarkers from blood tests to detect a variety of cancers. The technology is highly sensitive and can be used to detect ovarian, liver, and pancreatic cancers, sometimes even before symptoms appear [86]. Medical AI is also being used to enhance radiogenomics, combining radiological imaging and genomic data to enhance cancer identification and prediction of disease progression [87]. The application of medical AI in multi-omics datasets is not limited to diagnostic prediction. Patient responses to treatment can be predicted by AI systems, allowing the best treatment to be selected. Medical AI has been used to predict immunotherapy responses based on proteomic datasets, enabling personalized immunotherapy treatment [88].

AI-enhanced surgery task

Surgery is the foundation of modern medicine. However, as the complexity and risks of surgery increase, it is essential to improve surgical precision and achieve optimal efficacy and safety. In recent years, AI has increasingly penetrated the field of surgery, striving to revolutionize medical practice [89].

AI's involvement in surgery is multifaceted, beginning with its ability to assess surgical competence. Studies have shown that a surgeon's skill level directly reflects patient outcomes [90,91]. Medical AI can assess surgical outcomes in real time

based on robotic surgery data [92]. This unbiased assessment is critical, as traditional evaluations often rely on expert assessments, which are both subjective and variable. ML algorithms can efficiently analyze large amounts of surgical data, extracting details that might be overlooked, enhancing clinical outcome predictions, and informing postoperative care choices [93]. By reviewing surgical techniques, postoperative outcomes, and recovery, medical AI can support future treatment planning, resulting in more personalized care for patients.

Beyond assessment, AI can also play a role directly in surgical procedures, through the use of robotics. The da Vinci Surgical System stands out as a widely implemented technology in this field, which has been adopted across numerous surgical disciplines [94]. These robotic systems allow surgeons to perform minimally invasive surgeries with greater precision, using robotic arms controlled from a console [95]. AI-based systems can help doctors fine-tune procedures, provide real-time feedback, and improve accuracy [96]. This smart assistance is particularly beneficial in complex procedures where precision is critical, such as in prostatectomies or cardiovascular surgeries [97].

Medical AI will be used to enhance training. By reviewing surgeries, medical AI can assess surgeons' technical proficiency, helping them improve their skills. In some systems, DL is used to identify and correct technical errors such as improper suturing techniques during live surgeries. These systems not only promote continuous learning but also ensure that surgeons maintain proficiency throughout their careers [98].

Future innovations could lead to fully autonomous robotic surgeries, where AI moves beyond mere assistance to perform certain tasks independently, guided by vast data and ML algorithms. However, the integration of medical AI in surgery should be accompanied by rigorous evaluation frameworks, such as the IDEAL framework, to ensure safety, effectiveness, and ethical application [99].

In conclusion, medical AI is transforming surgery by enhancing precision, improving training, and evaluating surgeon performance. Its integration into robotics is transforming surgical intervention with better patient outcomes as well as a more efficient surgical method [100].

The applications of medical AI in health management and disease prevention

AI-powered health management

AI not only serves a key function in the identification and management of illnesses but also enables continuous health monitoring beyond the clinic [101]. By analyzing patient

behavior, heart rate or audio data collected from wearable devices, medical AI can detect hidden disease-related patterns, enabling ongoing disease monitoring and management [102]. In the following, we will describe how wearable technology combined with medical AI is transforming predictive health monitoring through multimodal data collection and intelligent analysis.

Wearable devices vary widely in form and function, ranging from wrist-worn sensors and smartwatches to full-body suits equipped with multiple inertial measurement units (IMUs). These devices can continuously capture physiological and biomechanical signals, including heart rate, skin temperature, movement kinematics, and electrodermal activity. For instance, in a study of Friedreich's ataxia (FA), a full-body motion capture suit embedded with 17 inertial sensors was used to monitor patients during clinical assessments such as the 8-meter walk and 9-hole peg tests [103]. These digital behavioral features, when analyzed by ML algorithms, allowed for the prediction of disease progression with significantly higher precision than traditional clinical scales. In another example involving Duchenne muscular dystrophy (DMD), researchers used a wearable 17-sensor bodysuit to capture daily-life movement data over a 12-month period. ML models were then applied to define behavioral "fingerprints" that accurately tracked disease progression and therapeutic responses [104]. These digital biomarkers offered a more objective and personalized method of assessment compared with conventional clinical tests.

A particularly promising area is the use of medical AI to detect and monitor Parkinson's disease (PD) through nocturnal breathing signals. A study has demonstrated that a single night of breathing data can be analyzed using medical AI models to detect PD with high accuracy and estimate disease severity based on the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [105]. Notably, these data can be collected either through wearable breathing belts or non-contact devices mounted on walls, which use low-power radio waves to measure chest and abdominal motion. This contactless setup allows for long-term monitoring without disrupting the patient's sleep environment, promoting better adherence and improved data quality [106,107].

As wearable technologies continue to evolve, the focus is shifting toward highly integrated, inconspicuous, and user-friendly designs [108,109]. Future systems are expected to achieve "invisible" wearability, minimizing user discomfort while maximizing data quality. This trend will further enable AI to utilize continuous, real-world data for personalized, predictive, and preventive healthcare, opening new frontiers in digital medicine.

AI-powered public health

AI is increasingly applied to both personal and public health, particularly in infectious disease control [110]. This application of medical AI expands beyond individual diagnostics and treatments to include a broader range involving population-wide surveillance, risk management, and healthcare resource optimization.

In the early stages of disease outbreaks, AI technologies significantly improve surveillance and early-warning capabilities [111]. For instance, HealthMap utilizes NLP and ML algorithms to scan global web content continuously, detecting and classifying disease-related information [112]. This AI-driven system was notably successful in identifying the emergence of COVID-19 through early detection of pneumonia clusters of unknown etiology, providing crucial early alerts to public health authorities [113]. Once a disease outbreak is identified, medical AI also plays a critical role in population risk assessment. AI-powered predictive analytics, such as ML models using syndromic surveillance data, effectively assess risk and predict the severity of infections like dengue fever and COVID-19, enabling targeted public health interventions and optimized resource allocation [114].

Additionally, medical AI accelerates the development of antiviral drugs through structural modeling and simulation. ML methodologies, including graph neural networks and generative models, have been instrumental in virtually screening extensive compound libraries [115]. These approaches enable the rapid identification of potential antiviral candidates during pandemics such as COVID-19 [110]. These techniques allow researchers to rapidly target new molecular structures, shortening the timeline from discovery to clinical testing.

In conclusion, medical AI can enhance various areas of public health, including disease surveillance, risk assessment, antiviral drug development, and resource allocation [116], significantly transforming global capacity for health emergency preparedness and intervention.

AI-powered medical assistance

In many low- and middle-income countries (LMICs), limited medical resources significantly hinder the ability to deliver adequate healthcare. These limitations manifest as insufficient funding, outdated medical infrastructure, shortages of healthcare professionals, and limited access to comprehensive health data necessary for effective patient management [117].

To address these constraints, traditional models of sending medical teams from high-income countries (HICs) have often been adopted. However, such approaches are unsustainable,

costly, and typically short-term solutions. Conversely, AI-assisted remote medical services have emerged as more viable, sustainable, and efficient alternatives, using advanced healthcare expertise from developed nations without physically relocating healthcare professionals [118].

The effectiveness of medical AI in LMIC healthcare contexts has been demonstrated in several practical implementations. A clinical validation study in Zambia exemplifies the impact of AI for screening diabetic retinopathy, a major cause of preventable blindness [6]. In this study, a DL-based AI model initially trained on data from a high-income setting (Singapore) was applied to Zambia's population-based diabetic retinopathy screening program. The medical AI demonstrated robust diagnostic performance, with a sensitivity of 92.25% and a specificity of 89.04% for detecting referable diabetic retinopathy. It was particularly effective in identifying vision-threatening cases, achieving a sensitivity of 99.42%. Such outcomes highlight AI's potential for large-scale implementation even in resource-constrained African regions, significantly contributing to early diagnosis and prevention of diabetic blindness.

Moreover, the integration of AI in antenatal care through digital health registries has also illustrated significant benefits in LMICs. For example, in Palestine, a controlled trial of an eRegistry incorporating clinical decision support significantly improved adherence to guidelines for screening and managing key antenatal conditions such as anemia, hypertension, and diabetes, as compared with traditional paper-based systems [119]. Although this did not immediately translate into improved health outcomes, the improved process adherence clearly demonstrated the potential of digital clinical decision-support systems to enhance the quality and consistency of healthcare services in LMICs, aligning closely with World Health Organization (WHO) guidelines.

Despite these promising applications, challenges remain when transferring medical AI technologies developed in high-income countries to LMIC contexts. Research conducted in Vietnam involving the use of a United Kingdom (UK)-developed medical AI model for COVID-19 triage illustrated this concern [117]. While this medical AI model initially showed strong performance in its original UK hospital setting, direct application to Vietnamese hospitals without adjustments led to significantly reduced performance [117]. Factors such as differences in population characteristics, variations in clinical practice, and differences in healthcare infrastructure were significant barriers to effective generalization of the model. This emphasizes the necessity of tailored AI solutions adapted to the specific contexts and resources of LMICs rather than a

direct application of models developed in distinctly different healthcare environments.

An additional systematic scoping review reinforced these observations, noting substantial barriers to successful medical AI adoption in LMICs, including limited availability of high-quality data, mixed impacts on workflow integration, poor user-friendliness, and insufficient adaptation to local contexts [120]. This underscores that the successful implementation of medical AI in LMIC healthcare systems requires thorough contextualization and careful consideration of local data quality, infrastructure capabilities, and human factors.

Large language models in healthcare

The AI models mentioned above have performed well in the medical field, but most current medical AI systems usually rely on data labeled in specific fields for training and focus on solving specific tasks. This specialized training model limits the model's application flexibility in cross-domain or general medical scenarios, making it difficult to generalize and deploy on a large scale in actual clinical environments.

Generative pre-trained transformer in healthcare

Unlike AI systems that work in fixed fields, GPT is gradually emerging as a general model. GPT is a LLM that uses massive text data for self-supervised learning pre-training. It can learn the intrinsic patterns and knowledge of language by training on a large-scale unlabeled text corpus, and then adapt to specific tasks through fine-tuning or directly apply to multiple tasks without fine-tuning. Its core technical route is based on the self-attention mechanism and transformer architecture, which enables it to flexibly process and generate coherent text information. This general and flexible training method is significantly different from traditional medical-specific AI systems and is not limited by structured data and fixed tasks [121].

Distinct from other medical AI systems that require certain professional knowledge to be used effectively, GPT is more oriented to the general public, greatly reducing the threshold for users to obtain information. This is due to its interactive, conversational question-and-answer method. Users do not need to master complex medical terms or have a medical background. They can interact with the GPT model through simple natural language to obtain relevant health advice, medical information, or answers to questions [122].

Nowadays, the range of applications of GPT in the medical field has expanded quickly and become more advanced [123]. A large number of studies have verified the application effect and potential of GPT in various scenarios such as clinical text summary generation, medical record

analysis, doctor-patient communication assistance, and disease diagnosis decision support [124-126]. For example, in the clinical text summary task, studies have shown that with the assistance of the GPT model, the completeness and accuracy of clinical text summaries can reach or even exceed the level of medical experts, significantly improving the efficiency of medical documents [127,128]. In addition, GPT has also shown excellent results in processing medical records in different languages. For example, in a medical record analysis study involving English, Spanish, and Italian, GPT-4 achieved an overall accuracy of 79% for 14 medical questions, demonstrating its great potential for automated analysis of multilingual medical texts [129]. At the same time, in open clinical management reasoning tasks, doctors using GPT assistance performed significantly better. A randomized controlled trial found that compared with traditional resources, GPT-4 assisted doctors' clinical decision-making scores increased by an average of 6.5% ($P < 0.001$), improving multiple capabilities such as diagnosis, management, and situational judgment [130]. Recently, the DeepSeek model developed by a Chinese team achieved an accuracy rate of 89.2% in the United States medical licensing examination (USMLE), exceeding GPT-4 (84.5%), and in terms of the completeness of diagnostic steps, DeepSeek scored 4.7 points, higher than GPT-4's 4.2 points. In addition, DeepSeek also performed very well in the diagnosis task of 125 standardized patient cases, especially in the treatment recommendation task, where its performance was comparable to or even better than the GPT-4 model [131]. These results further demonstrate that the performance of open source models such as DeepSeek in clinical decision-making tasks has reached or even exceeded the most advanced proprietary large models [132].

MedLLM in healthcare

MedLLM is developed based on LLM, and through fine-tuning it in the medical field and optimizing the underlying model, it can better play a role in clinical practice [133]. It is pre-trained in multiple stages through medical literature and EHRs data, and is specially optimized to enhance the diagnostic reasoning process similar to that of doctors' practice. The model uses self-guided learning and unified preference alignment strategies, closely aligns with clinical standards and hierarchical structures such as International Classification of Diseases (ICD-10) classification, and demonstrates a strong ability to diagnose diseases across multiple professional fields. Likewise, BiomedGPT is a general visual-language basic model that focuses on processing medical multimodal data. It performs well in medical imaging report generation

and imaging question-and-answer tasks, and is particularly suitable for medical imaging analysis [134]. In addition, the AI consultation system (SSPEC) customized for the hospital reflects the outstanding contribution of LLM in reducing the burden of nurses and improving the doctor-patient communication experience [135]. In another study, the human papilloma virus (HPV) vaccination chatbot intervention for parents of junior high school girls significantly increased the vaccination rate, with the vaccination or appointment rate in the intervention group reaching 7.1%, while the control group was only 1.8%, and significantly increased the frequency of consultation between parents and medical professionals (49.1% vs. 17.6%, $P < 0.001$), reflecting the potential of LLM in promoting public health [136].

At present, more and more LLMs in specific fields have been launched (Table 1) [133,137-150], such as organoids-GPT (O-GPT), a vertical large model focusing on organoid research. Organoids are "mini organs" with functions formed in 3D culture *in vitro* through the self-organization of stem cells, and play an important role in drug development, disease research, and regenerative medicine [151]. The release of O-GPT remedies the lack of application of vertical large models in the organoids research, providing researchers with intelligent knowledge support and efficient research tools. The advent of LLMs has significantly enhanced the efficiency of related research and promoted the in-depth application and innovation of AI technology.

With the continuous breakthroughs in technologies such as DL, multimodal fusion, and large models, AI is being applied to various fields of medicine, whether it is disease screening, diagnosis, treatment decision-making, or health management, public health, drug research and development, etc., it has shown unprecedented innovative vitality (Fig. 3).

Limitations of LLMs

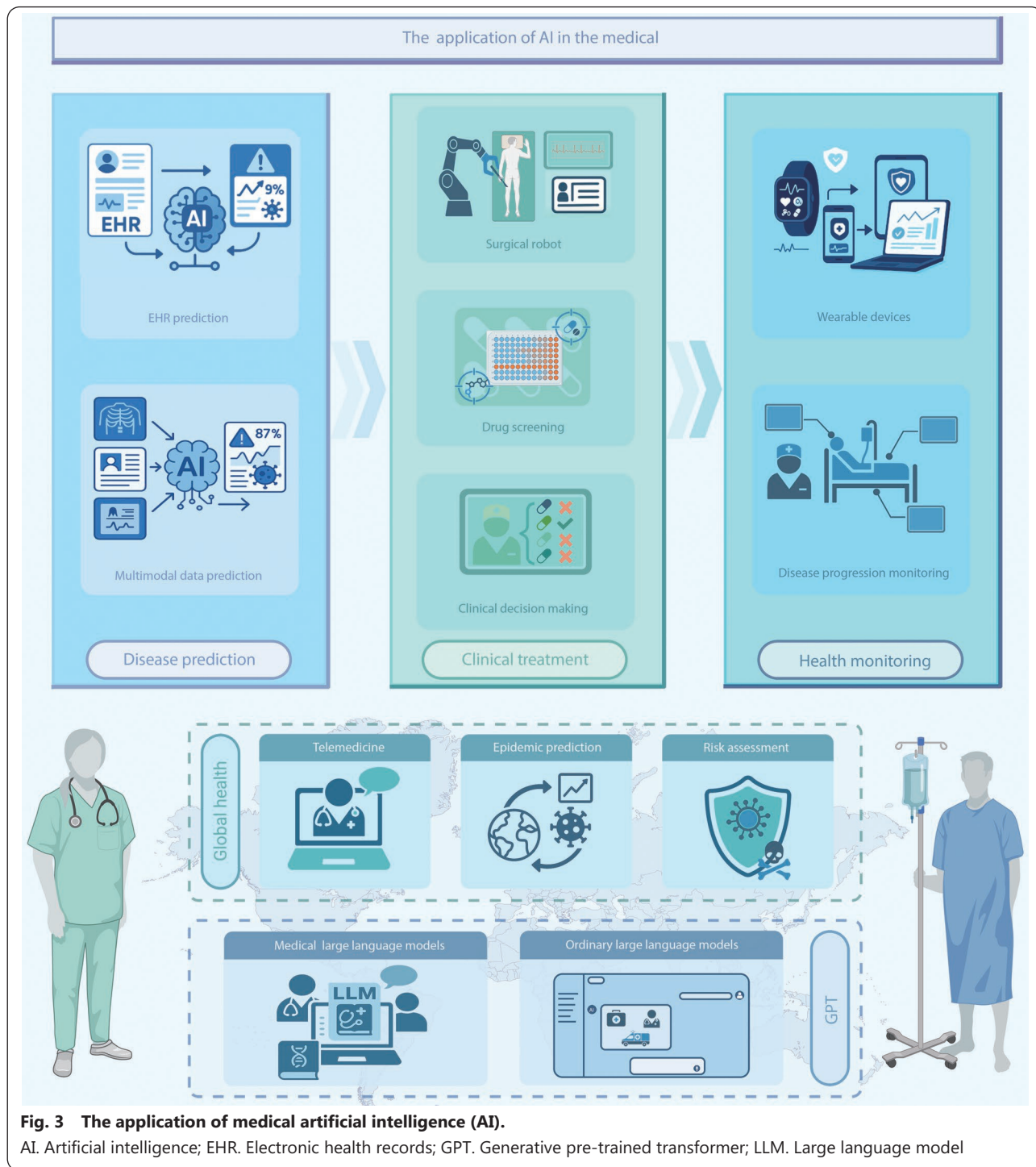
Although LLMs have potential in the medical field, their existing problems cannot be ignored. LLMs have obvious racial, income, and gender biases in the medical decision-making process, such as recommending different medical interventions for different socioeconomic groups. This bias may worsen the unfair distribution of medical resources [152,153]. In addition, although LLMs perform well on static medical record data, their performance in actual patient conversations and dynamic clinical decision-making scenarios has declined significantly [154].

LLM also faces the risk of data contamination. Alber et al. [155] have shown that only a very small amount of erroneous information (0.001%) is enough to mislead LLM's

Table 1 The latest medical large language models

Year	Name	Training data	Evaluation data	Application	Clinical relevance	Shortage	Reference
2023	GatorTronGPT	Real clinical text	Medical NLP benchmark tasks	Human-like clinical text generation	Medium	Lack of clinical safety verification	[137]
2023	Med-MLLM	CXR, CT, PubMed, MIMIC-CXR/III	COVID-19 multimodal dataset	Rapid multimodal disease support	High	Limited clinical interpretability	[138]
2024	DeepDR-LLM	NDSP	CPSSDRM	Diabetes/retinopathy screening assistant	High	Narrow scope of application	[139]
2024	PathChat	Pathological images	Pathological cases	Pathology vision-language AI assistant	Medium	Insufficient independent diagnostic capabilities	[140]
2024	MEDIC	Pharmacist labeling instructions	Pharmacy real environment test	LLM prevents pharmacy errors	High	Limited applicability to complex drug combinations	[141]
2024	OncoLLM	Oncology patient EHR	Doctor's proofreading	Automated clinical trial matching	Medium	Lower accuracy than doctors	[142]
2025	BrainGPT	Brain CT images	CQ500 dataset	3D brain CT reporting	Medium	Applicability limitations	[143]
2025	MedFound-DX-PA	MedCorpus, MedDX-FT	MedDX-Bench	Generalist disease diagnosis LLM	Medium	High computing power consumption	[133]
2025	Med-PaLM2	MultiMedQA	MultiMedQA	Expert-level medical Q&A	High	Occasional hallucinations and wrong answers	[144]
2025	COMPOSER-LLM	EHR from UCSD	Patient visit rate	Early sepsis prediction support	High	Strong reliance on EHR integrity	[145]
2025	Woollie	Real-world oncology data	Real-world oncology data	Cancer progression prediction LLM	Medium	Training data comes from a single source	[146]
2025	MMedIns-Llama 3	MedS-Ins	MedS-Bench	Versatile clinical medical LLM	Medium	Lack of real hospital application verification	[147]
2025	MedFound	MedCorpus	MedDX-Bench	Generalist diagnostic reasoning LLM	High	High computing power consumption	[148]
2025	EyeCLIP	Multimodal ophthalmic imaging	APTOS2019, MESSIDOR2	Multimodal ophthalmology disease diagnosis	High	Data mainly comes from Chinese patients	[149]
2025	ThyGPT	Ultrasound imaging of thyroid nodules	Diagnostic accuracy	Interactive thyroid nodule diagnosis	High	Insufficient identification of disease subtypes	[150]

3D. Three-dimensional; AI. Artificial intelligence; COVID-19. Coronavirus disease 2019; CT. Computed tomography; CXR. Chest X-ray; EHR. Electronic health record; LLM. Large language model; MLLM. Multimodal large language model; NDSP. NICHENG Diabetes Screening Project; NLP. Natural language processing; PA. Patient assessment; Q&A. Question and answering; UCSD. University of California San Diego



performance in medical tasks, thus posing a potential threat to patient health. Furthermore, LLM generally has the problem of “hallucination”, that is, generating information that does not match the facts, which is particularly dangerous in the medical field [156].

To avoid these problems, standardized evaluation systems such as TRIPOD-LLM and CRAFT-MD reporting specifications have been proposed and gradually widely

adopted [157]. These standardized reports provide clear evaluation criteria, including requirements in multiple aspects such as model transparency, data quality, bias identification, and clinical applicability [158]. Based on these evaluations, the current mainstream LLM still needs to be further optimized in actual clinical applications to ensure its safety, fairness, and reliability.

Despite LLM shows great potential in the medical field, in

order to realize its true clinical value, it remains important to monitor the bias, data pollution, and hallucination problems in the model, and promote its safety through strict and standardized evaluation standards.

Challenges and limitations of medical AI

While medical AI has significantly accelerated the progress of modern medicine, a number of challenges have emerged as a result of algorithmic limitations, incomplete data, and insufficient regulatory frameworks. The most notable concerns include bias and fairness, algorithmic interpretability and transparency, as well as data security and privacy [158,159] (Fig. 4).

Bias in medical AI

Bias in medical AI refers to systematic unfairness that arises during diagnosis, prediction, or decision-making processes due to factors such as data imbalance, model design, or broader social structures. This bias often manifests as significant variations in predictive performance across different racial, gender, age, or socioeconomic groups [160]. As a result, certain populations may receive lower-quality care, further worsening existing health disparities. For example, studies have revealed that mainstream health risk assessment

algorithms in the United States often underestimate the health risks of Black patients during resource allocation [161,162]. This occurs because these algorithms incorporate “healthcare costs” as a proxy variable, which, in turn, undervalues the needs of minority patients, even when they have more severe conditions, leading to insufficient access to medical support [163]. The primary cause lies in structural inequalities embedded in healthcare spending. Due to multiple socioeconomic barriers, minority groups such as Black patients typically have reduced access to healthcare resources. During training, AI algorithms “learn” and further reinforce or intensify these inequities [164]. Beyond racial bias, gender bias is also serious in medical AI. Many algorithms are developed without adequate consideration of gender-specific factors, resulting in the underrepresentation or misdiagnosis of women, lesbian, gay, bisexual, transgender, and queer/questioning (LGBTQ) individuals, and transgender patients [165]. For instance, in medical imaging and psychiatric AI applications, the rate of missed diagnoses is higher for women and minority groups, and recommendations are often less accurate, directly impacting the fairness and effectiveness of healthcare [166].

To address these challenges, the medical community and AI developers are actively exploring various solutions. There

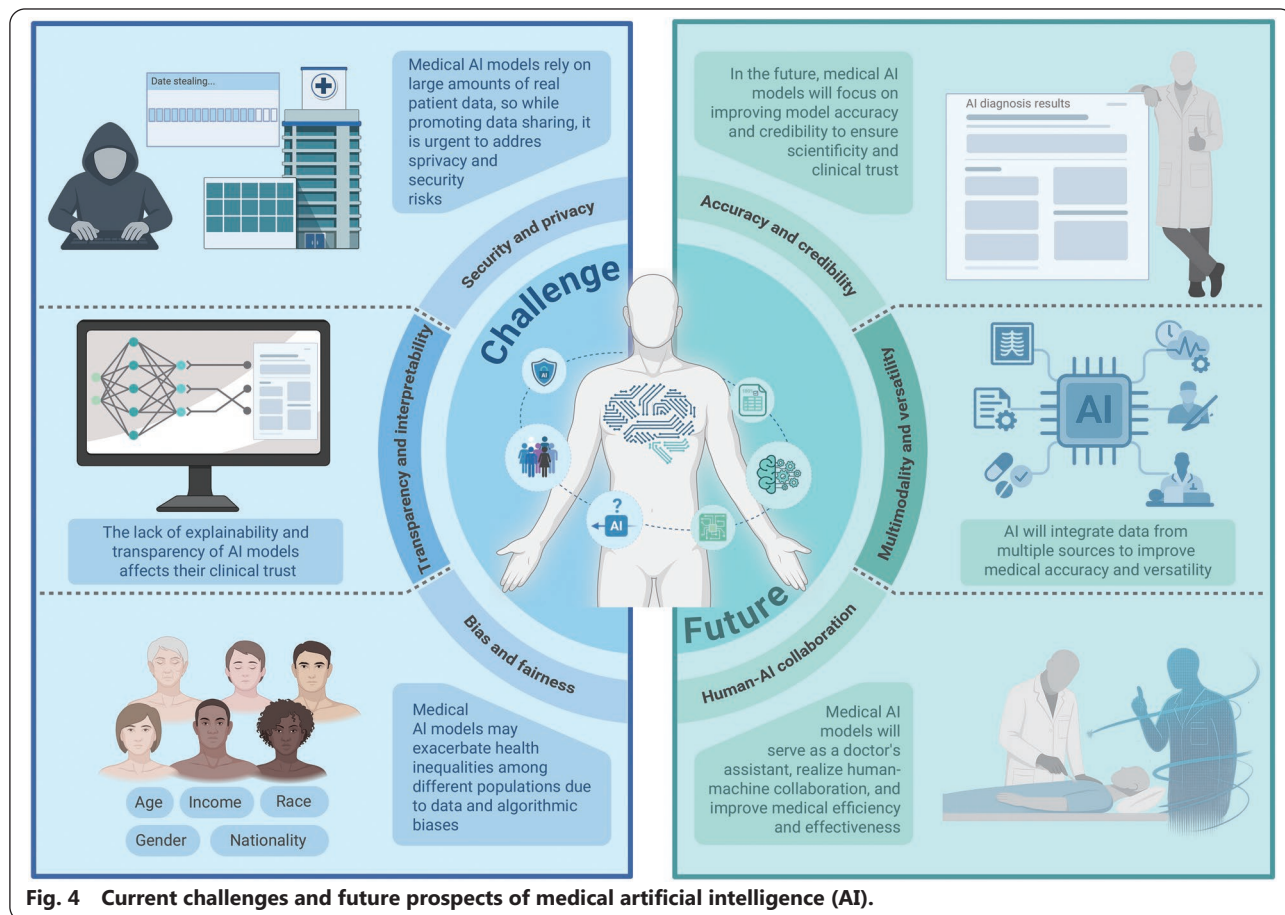


Fig. 4 Current challenges and future prospects of medical artificial intelligence (AI).

is a need to improve the diversity and representativeness of datasets at the source, ensuring adequate inclusion of minority populations, different genders, and varied social backgrounds [167]. In recent years, organizations such as the National Institutes of Health (NIH) have promoted the development of multicenter, multiethnic high-quality medical datasets, which help alleviate data bias [163]. Additionally, algorithm development should focus on the appropriate selection of objective functions, avoiding the use of easily accessible but structurally flawed proxy variables in place of actual health needs. During model evaluation and deployment, fairness assessment metrics should be incorporated to monitor algorithmic performance across different populations and subgroups, enabling timely identification and correction of potential inequities [168]. Additionally, technical approaches such as resampling, adversarial training, and domain adaptation can be used to reduce bias in model outputs.

Interpretability, transparency, and the “black box” in medical AI

In the evolution of medical AI, the concepts of “interpretability”, “transparency”, and the “black box” phenomenon are closely interrelated [169]. Interpretability refers to the extent to which an AI model can clearly explain to human users the reasons behind its predictions or decisions, enabling clinicians, patients, and regulators to understand the model’s rationale. Transparency emphasizes the openness and traceability of the system’s architecture, algorithmic details, and data processing workflows, allowing external experts or users to examine the inner operations [170,171]. In contrast, the “black box” describes highly complex AI models, particularly DL networks, whose internal logic is largely unclear [172].

With the widespread adoption of DL and other complex models in medicine, model capacity and predictive accuracy have improved markedly. However, this progress has also resulted in a pronounced “black box” effect, reducing both interpretability and transparency. The black box phenomenon directly impacts the clinical acceptance and trustworthiness of medical AI. Clinicians and patients are often more concerned with understanding why a model produces a specific diagnosis or recommendation, not just the outcome itself. If a medical AI cannot justify its decisions, clinicians may be reluctant to use it, especially when predictions conflict with established medical knowledge or influence critical clinical decisions [173,174].

To overcome these barriers, researchers are increasingly incorporating explainable AI (XAI) frameworks into medical model design [175]. By utilizing additive explanation algorithms such as SHAP, it is possible to attribute each

prediction to specific input features in real time, identifying the most influential variables that shaped the model’s output [172]. This allows clinicians to know the model’s logic, supervise the decision-making process, and make informed judgments.

Improving transparency also involves open-sourcing algorithm code, fully disclosing model training procedures, data processing details, and decision thresholds, and continuously assessing model performance and bias in clinical practice [176,177]. Only through these measures can AI systems escape the constraints of the “black box”, build trust among clinicians and patients, and achieve standardized application in healthcare settings.

Data security and privacy in medical AI

In the current era of deep integration between AI and healthcare, access to real patient data has become the foundation for training and validating medical AI models. Whether for automated disease diagnosis, treatment recommendation, or health risk prediction, high-quality and diverse clinical datasets are crucial for continuous model improvement [178]. To maximize generalizability and clinical applicability, data sharing across multiple institutions is often required [179]. However, this process inevitably raises significant concerns regarding patient privacy and data security. Patient data breaches can lead to a range of security risks and erode public confidence in both healthcare systems and emerging AI technologies [180]. This loss of trust may ultimately hinder the practical implementation of AI in clinical settings.

Researchers have developed multiple technical solutions to address these concerns. For end-to-end privacy protection, the PriMIA framework integrates differential privacy, encrypted aggregation, and secure multiparty computation, ensuring that data remains local throughout the model training process [181]. This effectively guards against re-identification attacks and model theft, while maintaining control over proprietary models, providing a robust security foundation for collaborative, multi-institutional medical AI. In the context of medical imaging data sharing, “digital masking” technology leverages 3D reconstruction and DL to permanently remove facial biometric features, retaining only disease-relevant information [182]. This greatly reduces the risk of identity leakage, increases patient acceptance of data sharing, and offers a novel paradigm for the compliant use of medical imaging data. Federated learning, as a leading approach for distributed collaborative training, allows models to be trained locally at each institution with only model parameters exchanged, never raw data. This enables diverse data sharing across regions and institutions while effectively protecting privacy, significantly

improving model fairness, robustness, and generalizability, and supporting the establishment of multicenter medical AI systems [183,184].

Legal and regulatory issues in medical AI

With the widespread application of AI in healthcare, ensuring its safe, effective, and compliant integration remains a critical challenge. Currently, there is no unified standard for the approval systems and legal frameworks for medical AI across different countries and regions, which complicates global implementation and dissemination. The European Union's AI Act is the world's first comprehensive regulatory framework for AI technology, specifically emphasizing high-risk categories for medical AI applications [185]. It mandates rigorous approval processes for AI systems involved in life and health, ensuring their transparency, safety, and compliance. For the shortcomings, AI Act simplifies "trustworthiness" into "risk acceptability", relying on expert conformity assessments to determine whether AI systems can be deployed. However, this technocratic approach neglects the public's actual perception of trust [186].

The General Data Protection Regulation (GDPR) imposes strict requirements for data privacy in AI applications, particularly in healthcare. GDPR emphasizes informed consent, data minimization, data security, and traceability of data [187]. This means that AI systems processing patient health data must ensure transparency and privacy protection while requiring organizations to comply with regulations during data collection, storage, processing, and sharing. GDPR presents a challenge for cross-border AI applications, as it restricts the free flow of patient data, which is critical for global data sharing in medical AI applications [188].

In the United States, the FDA regulates AI in medical devices typically through the Software as a Medical Device (SaMD) pathway, which requires AI models to demonstrate their accuracy, reliability, and clinical effectiveness [189]. However, due to the "continuous learning" nature of AI models, the FDA faces challenges in addressing dynamic algorithm changes, and existing approval processes may not be well-suited to the rapid development of AI technology. Additionally, data privacy laws such as Health Insurance Portability and Accountability Act (HIPAA) impose strict requirements on AI applications in healthcare, particularly regarding the protection and security of patient data [190].

Evaluating medical AI goes beyond assessing algorithmic performance. It also requires considering clinical effectiveness, representativeness of data sources, fairness, interpretability, and compatibility with existing healthcare systems [191]. However,

there is no unified standard for evaluating the comprehensive performance of AI systems, resulting in significant variations in evaluation outcomes across regions and institutions, further complicating cross-national and cross-regional applications.

Therefore, the future should focus on promoting a globally unified AI evaluation standard, particularly in areas such as data privacy protection, algorithm interpretability, and fairness assessments. This will ensure that medical AI can be effectively applied within different legal and regulatory frameworks and truly benefit global patients.

Prospects and future directions

To improve the performance and efficiency of medical AI, it is essential to recognize that even the slightest error or "hallucination" in a clinical context can deprive patients of life-saving opportunities or even cause medical accidents [192]. Therefore, future medical AI model optimization should not focus solely on performance metrics, but also safeguard the fundamental principle of "safety and reliability", prioritizing predictive accuracy and risk control. Medical AI should adopt a more conservative approach. In the face of uncertainty, the system should defer to human intervention rather than issue unreliable conclusions. This strategy helps minimize medical error rates, manage potential risks, and uphold patient safety.

The widespread adoption of medical AI in clinical practice is contingent upon the standardization of medical data sources [8]. Establishing high-quality standards for medical data is essential. These standards include strictly regulating data collection, annotation, storage, and sharing procedures, as well as improving representativeness and accuracy [193]. Only on the foundation of standardized and regulated data can medical AI models adapt to diverse clinical scenarios and achieve greater generalizability and fairness.

Furthermore, ensuring model transparency is essential [173]. Medical AI systems must avoid functioning as inscrutable "black boxes"; each stage of model reasoning should be both traceable and interpretable, thereby enabling clinicians to review, interrogate, and, when necessary, revise the system's outputs. Integrating explainable algorithms and developing transparent, logical, and causally sound reasoning processes can help strengthen physician and patient trust in medical AI. Meanwhile, the industry should accelerate the standardization and disclosure of model evaluation criteria. It is essential to establish unified and scientific performance assessment systems that consider accuracy, safety, generalizability, fairness, and interpretability. Such standards are critical to ensuring the clinical reliability of medical AI.

Despite the rapid development of medical AI, it should

neither replace nor attempt to replace professional clinicians at present. AI is better suited to serve as a supportive tool for physicians, particularly in tasks that are highly repetitive, workflow-driven, or data-intensive, reducing workload and improving efficiency and accuracy [194]. Numerous studies show that human-AI collaborative diagnostic strategies consistently outperform either human experts or AI models alone [195,196]. By leveraging their complementary strengths, diagnostic performance and patient outcomes are significantly enhanced. Therefore, the future design of medical AI should emphasize deep integration with clinical teams, enabling “human-machine co-management” and collaborative decision-making to maximize the efficiency of medical resource utilization.

Moreover, the development of medical AI should progress beyond single-task functionalities toward fully integrated, multimodal, and end-to-end intelligent systems. Ideally, medical AI models will combine the interactive capabilities of LLMs for natural dialogue with patients, the analytical power of computer vision for interpreting medical images and pathology slides, and the capacity for evidence-based diagnostic reasoning and personalized treatment planning. In surgical settings, AI may even assist with certain procedures, reducing clinician burden and enhancing both safety and precision. In the longer term, the introduction of digital models, such as virtual organs, is expected to further innovate the virtual medical system [197,198]. These high-fidelity digital biological models will enable us to realize risk-free personalized disease simulation and treatment prediction in the computer before any physical intervention.

In conclusion, medical AI represents a next-generation productivity tool with substantial transformative potential. It offers solutions to enduring challenges such as inequitable healthcare access, facilitates the implementation of precision medicine and intelligent health strategies, and supports the provision of comprehensive care for diverse populations, positioning itself as a central driver of global medical advancement. Achieving this goal requires that AI systems are always committed to protecting patient safety and clinical value. It also requires continuous improvement in model performance and ongoing strengthening of regulatory and ethical frameworks. Medical AI can be effectively and safely integrated into mainstream clinical practice only through adherence to these principles, thereby advancing human health in a more intelligent and equitable manner.

Abbreviations

3D: Three-dimensional

AI: Artificial intelligence
EHRs: Electronic health records
FDA: U.S. Food and Drug Administration
GDPR: General Data Protection Regulation
GPT: Generative pre-trained transformer
LMICs: Low- and middle-income countries
LLMs: Large language models
M3FM: Medical multimodal multitask foundation model
MedLLM: Medical large language models
ML: Machine learning
NLP: Natural language processing

Acknowledgements

We would like to acknowledge the use of the following software for figure design and plotting: Biorender (www.biorender.com). The language polishing for this article was done by Springer Nature Editing Service.

Authors' contributions

LB, FXW, and XT conceptualized the review. CZ wrote and edited the manuscript; CZ, JLL, ND, and YH collected published papers. LB, JCS, and PRS revised the manuscript. The illustrations in this review were created with Biorender.com by CZ. LB and JCS provided funding support. All authors reviewed and approved the final version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (82230071, 32471396, 82427809), the Shanghai Committee of Science and Technology (23141900600), the Shanghai Clinical Research Plan (SHDC2023CRT013), the Shanghai Municipal Demonstration Project for Innovative Medical Device Applications (23SH505700), and the Young Elite Scientist Sponsorship Program by China Association for Science and Technology (YESS20230049).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Organoid Research Center, Institute of Translational Medicine, Shanghai University, Shanghai 200444, China. ²MedEng-X Institutes, Shanghai University, Shanghai 200444, China. ³School of Medicine, Shanghai University, Shanghai 200444, China. ⁴National Center for Translational Medicine (Shanghai) SHU Branch, Shanghai University, Shanghai 200444, China. ⁵Wenzhou Institute of Shanghai University, Wenzhou 325000, Zhejiang, China. ⁶Department of Orthopaedics, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China.

References

1. Safiri S, Asghari KM, Sullman MJ. The global burden of diseases and injuries among older adults. *Int J Aging*. 2023;1:e16.
2. Pettifor JM, Saloojee H. Unequal access: the prohibitive costs of new drugs in low- and middle-income countries. *Nat Rev Nephrol*. 2025;21(5):281-2.
3. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-8.
4. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health*. 2020;2(9):e486-8.
5. Xu Q, Zhan X, Zhou Z, Li Y, Xie P, Zhang S, et al. AI-based analysis of CT images for rapid triage of COVID-19 patients. *NPJ Digit Med*. 2021;4(1):75.
6. Bellemo V, Lim ZW, Lim G, Nguyen QD, Xie Y, Yip MY, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digit Health*. 2019;1(1):e35-44.
7. Kim C, Gadgil SU, Lee SI. Transparency of medical artificial intelligence systems. *Nat Rev Bioeng*. 2026;4:11-29.
8. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30-6.
9. Maynard AD. Navigating the fourth industrial revolution. *Nat Nanotechnol* 2015;10(12):1005-6.
10. Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, et al. Artificial intelligence: a powerful paradigm for scientific research. *Innovation*. 2021;2(4):100179.
11. Cai Y, Cai YQ, Tang LY, Wang YH, Gong M, Jing TC, et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review. *BMC Med*. 2024;22(1):56.
12. Shikino K, Shimizu T, Otsuka Y, Tago M, Takahashi H, Watari T, et al. Evaluation of ChatGPT-generated differential diagnosis for common diseases with atypical presentation: descriptive research. *JMIR Med Edu*. 2024;10:e58758.
13. Xu H, Usuyama N, Bagga J, Zhang S, Rao R, Naumann T, et al. A whole-slide foundation model for digital pathology from real-world data. *Nature*. 2024;630(8015):181-8.
14. Agrawal T, Choudhary P. Segmentation and classification on chest radiography: a systematic survey. *Vis Comput*. 2023;39(3):875-913.
15. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM*. 1966;9(1):36-45.
16. Shortliffe EH. Mycin: a knowledge-based computer program applied to infectious diseases. *Proc Annu Symp Comput Appl Med Care*. 1977;5:66-9.
17. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain: an evolving diagnostic decision-support system. *JAMA*. 1987;258(1):67-74.
18. Deng L. Artificial intelligence in the rising wave of deep learning: the historical path and future outlook. *IEEE Signal Process Mag*. 2018;35(1):180-77.
19. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. *World J Gastroenterol*. 2019;25(14):1666.
20. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;313(5786):504-7.
21. Comendador BEV, Francisco BMB, Medenilla JS, Mae S. Pharmabot: a pediatric generic medicine consultant chatbot. *J Autom Control Eng*. 2015;3(2):137-40.
22. Hamilton D, Kohli SS, McBeth P, Moore R, Hamilton K, Kirkpatrick AW. Low Earth orbit communication satellites: a positively disruptive technology that could change the delivery of healthcare in rural and northern Canada. *J Med Internet Res*. 2025;27:e46113.
23. Arterys Cardio DL Cloud MRI Analytics Software Receives FDA Clearance. Diagnostic and interventional cardiology (DAIC), Arlington Heights. 2017. <https://www.dicardiology.com/product/arterys-cardio-dl-cloud-mri-analytics-software-receives-fda-clearance>. Accessed 13 Jan, 2026.
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30.
25. Yang Z, Zeng X, Zhao Y, Chen R. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct Tar*. 2023;8(1):115.
26. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. *Nature*. 2023;622(7981):156-63.
27. Xiang J, Wang X, Zhang X, Xi Y, Eweje F, Chen Y, et al. A vision-language foundation model for precision oncology. *Nature*. 2025;638(8051):769-78.
28. Ni J, You Y, Wu X, Chen X, Wang J, Li Y. Performance evaluation of deep learning for the detection and segmentation of thyroid nodules: systematic review and meta-analysis. *J Med Internet Res*. 2025;27:e73516.
29. Gao Y, Chen Y, Wang M, Wu J, Kim Y, Zhou K, et al. Optimising the paradigms of human AI collaborative clinical coding. *NPJ Digit Med*. 2024;7(1):368.
30. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395-405.
31. Hu C, Xia Y, Zheng Z, Cao M, Zheng G, Chen S, et al. AI-based large-scale screening of gastric cancer from noncontrast CT imaging. *Nat Med*. 2025:1-9.
32. Ma D, Pang J, Gotway MB, Liang J. A fully open AI foundation model applied to chest radiography. *Nature*. 2025;643(8071):488-98.
33. Ulloa-Cerna AE, Jing L, Pfeifer JM, Raghunath S, Ruhl JA, Rocha DB, et al. rECHOmmend: an ECG-based machine learning approach for identifying patients at increased risk of undiagnosed structural heart disease detectable by echocardiography. *Circulation*. 2022;146(1):36-47.
34. Miller DD. The medical AI insurgency: what physicians must know about data to practice with intelligent machines. *NPJ Digit Med*. 2019;2(1):62.
35. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. *JAMA Intern Med*. 2019;179(3):293-4.
36. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol*. 2017;69(21):2657-64.
37. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-30.
38. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349(6245):255-60.
39. Ng AY, Oberije CJ, Ambrózay É, Szabó E, Serfözö O, Karpati E, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. *Nat Med*.

- 2023;29(12):3044-9.
40. Adir O, Poley M, Chen G, Froim S, Krinsky N, Shklover J, *et al.* Integrating artificial intelligence and nanotechnology for precision cancer medicine. *Adv Mater.* 2020;32(13):1901989.
 41. Chanda T, Hauser K, Hobelsberger S, Bucher TC, Garcia CN, Wies C, *et al.* Dermatologist-like explainable AI enhances trust and confidence in diagnosing melanoma. *Nat Commun.* 2024; 15(1):524.
 42. Nimri R, Battelino T, Laffel LM, Slover RH, Schatz D, Weinzimer SA, *et al.* Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. *Nat Med.* 2020;26(9):1380-4.
 43. Liu GS, Fereydooni S, Lee MC, Polkampally S, Huynh J, Kuchibhotla S, *et al.* Scoping review of deep learning research illuminates artificial intelligence chasm in otolaryngology-head and neck surgery. *NPJ Digit Med.* 2025;8(1):265.
 44. Guermazi A, Tannoury C, Kompel AJ, Murakami AM, Ducarouge A, Gillibert A, *et al.* Improving radiographic fracture recognition performance and efficiency using artificial intelligence. *Radiology.* 2022;302(3):627-36.
 45. Baumgartner C, Brislinger D. Transforming precision medicine: the potential of the clinical artificial intelligent single-cell framework. *Clin Transl Med.* 2025;15(1):e70096.
 46. Chen M, Wang Y, Wang Q, Shi J, Wang H, Ye Z, *et al.* Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation. *NPJ Digit Med.* 2024; 7(1):349.
 47. Phongpreecha T, Ghanem M, Reiss JD, Oskotsky T, Mataraso SJ, De Francesco D, *et al.* AI-guided precision parenteral nutrition for neonatal intensive care units. *Nat Med.* 2025;31(6):1882-94.
 48. Tang AS, Woldemariam SR, Miramontes S, Norgeot B, Oskotsky TT, Sirota M. Harnessing EHR data for health research. *Nat Med.* 2024; 30(7):1847-55.
 49. Saleh S, El Arnaout N, Abdouni L, Jammoul Z, Hachach N, Dasgupta A. Sijilli: a scalable model of cloud-based electronic health records for migrating populations in low-resource settings. *J Med Internet Res.* 2020;22(8):e18183.
 50. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, *et al.* Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *Br Med J.* 2015;350:h1885.
 51. Goh KH, Wang L, Yeow AYK, Poh H, Li K, Yeow JLL, *et al.* Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun.* 2021;12(1):711.
 52. Deasy JO, Stetson PD. A platform for continuous learning in oncology. *Nat Cancer.* 2021;2(7):675-6.
 53. Morin O, Vallières M, Braunstein S, Ginart JB, Upadhaya T, Woodruff HC, *et al.* An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication. *Nat Cancer.* 2021;2(7):709-22.
 54. Liang H, Tsui BY, Ni H, Valentim CC, Baxter SL, Liu G, *et al.* Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med.* 2019;25(3):433-8.
 55. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *Br Med J.* 2018;361.
 56. Tong L, Shi W, Isgut M, Zhong Y, Lais P, Gloster L, *et al.* Integrating multi-omics data with EHR for precision medicine using advanced artificial intelligence. *IEEE Rev Biomed Eng* 2023;17:80-97.
 57. Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. *Ann Rheum Dis.* 2023;82(3):306-11.
 58. James AP, Dasarathy BV. Medical image fusion: a survey of the state of the art. *Inf Fusion.* 2014;19:4-19.
 59. Sjoding MW, Taylor D, Motyka J, Lee E, Claar D, McSparron JJ, *et al.* Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. *Lancet Digit Health.* 2021;3(6):e340-8.
 60. Lotter W, Diab AR, Haslam B, Kim JG, Grisot G, Wu E, *et al.* Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med.* 2021;27(2):244-9.
 61. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, *et al.* Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell.* 2020;181(6):1423-33.e11.
 62. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, *et al.* Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med.* 2019; 25(1):70-4.
 63. Wang YR, Yang K, Wen Y, Wang P, Hu Y, Lai Y, *et al.* Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging. *Nat Med.* 2024; 30(5):1471-80.
 64. Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digit Health.* 2022;4(7): e507-19.
 65. Joshi S, Forjaz A, Han KS, Shen Y, Queiroga V, Selaru FA, *et al.* Interpol-AI: deep learning-based optical flow interpolation and restoration of biomedical images for improved 3D tissue mapping. *Nat Methods.* 2025;22(7):1556-67.
 66. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med.* 2022;28(9):1773-84.
 67. Rao VM, Hla M, Moor M, Adithan S, Kwak S, Topol EJ, *et al.* Multimodal generative AI for medical image interpretation. *Nature.* 2025;639(8056):888-96.
 68. Xue C, Kowshik SS, Lteif D, Puducheri S, Jasodanand VH, Zhou OT, *et al.* AI-based differential diagnosis of dementia etiologies on multimodal data. *Nat Med.* 2024;30(10):2977-89.
 69. Niu C, Lyu Q, Carothers CD, Kaviani P, Tan J, Yan P, *et al.* Medical multimodal multitask foundation model for lung cancer screening. *Nat Commun.* 2025;16(1):1523.
 70. Keyl J, Keyl P, Montavon G, Hosch R, Brehmer A, Mochmann L, *et al.* Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nat Cancer.* 2025;6(2):307-22.
 71. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, *et al.* Foundation models for generalist medical artificial intelligence. *Nature.* 2023;616(7956):259-65.
 72. Zhang K, Yang X, Wang Y, Yu Y, Huang N, Li G, *et al.* Artificial intelligence in drug development. *Nat Med.* 2025;31(1):45-59.
 73. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature.* 2023;616(7958):673-85.
 74. Brotzakis ZF, Zhang S, Murtada MH, Vendruscolo M. AlphaFold prediction of structural ensembles of disordered proteins. *Nat Commun.* 2025;16(1):1632.
 75. Cook MA, Wright GD. The past, present, and future of antibiotics. *Sci Transl Med.* 2022;14(657):eabo7793.

76. Krishnan A, Valeri JA, Jin W, Donghia NM, Sieben L, Luttens A, et al. A generative deep learning approach to de novo antibiotic design. *Cell*. 2025;188(21):5962-79.
77. Lavecchia A. Deep learning helps fight against antibiotic resistance. *Cell Host Microbe*. 2025;33(10):1641-2.
78. Wong F, Zheng EJ, Valeri JA, Donghia NM, Anahtar MN, Omori S, et al. Discovery of a structural class of antibiotics with explainable deep learning. *Nature*. 2024;626(7997):177-85.
79. Tanoli Z, Fernández-Torras A, Özcan UO, Kushnir A, Nader KM, Gadiya Y, et al. Computational drug repurposing: approaches, evaluation of in silico resources and case studies. *Nat Rev Drug Discovery*. 2025;24(7):521-42.
80. Huang K, Chandak P, Wang Q, Havaladar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. *Nat Med*. 2024;30(12):3601-13.
81. Ashley EA. Towards precision medicine. *Nat Rev Genet*. 2016;17(9):507-22.
82. Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. *Nat Rev Cancer* 2015;15(12):747-56.
83. Sandhu C, Qureshi A, Emili A. Panomics for precision medicine. *Trends Mol Med*. 2018;24(1):85-101.
84. Jiang YL, Zhao G, Wang SH, Li N. Leveraging artificial intelligence for clinical decision support in personalized standard regimen recommendation for cancer. *Mil Med Res*. 2025;12(1):31.
85. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. *Cancer Discov*. 2021;11(4):900-15.
86. Killock D. CancerSEEK and destroy—a blood test for early cancer detection. *Nat Rev Clin Oncol*. 2018;15(3):133.
87. Guo Y, Li T, Gong B, Hu Y, Wang S, Yang L, Zheng C. From images to genes: radiogenomics based on artificial intelligence to achieve non-invasive precision medicine in cancer patients. *Adv Sci*. 2025;12(2):2408069.
88. He X, Liu X, Zuo F, Shi H, Jing J. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. *Semin Cancer Biol*. 2023;88:187-200.
89. Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. *Nat Med*. 2024;30(5):1257-68.
90. Stulberg JJ, Huang R, Kreutzer L, Ban K, Champagne BJ, Steele SR, et al. Association between surgeon technical skills and patient outcomes. *JAMA Surg*. 2020;155(10):960-8.
91. Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation*. 1989;79(6):13-12.
92. Hung AJ, Chen J, Gill IS. Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. *JAMA Surg*. 2018;153(8):770-1.
93. Goodman ED, Patel KK, Zhang Y, Locke W, Kennedy CJ, Mehrotra R, et al. Analyzing surgical technique in diverse open surgical videos with multitask machine learning. *JAMA Surg*. 2024;159(2):185-92.
94. Douissard J, Hagen ME, Morel P. The da Vinci surgical system. *Bariatric robotic surgery: a comprehensive guide*: Springer; 2019. p.13-27.
95. Haidegger T, Speidel S, Stoyanov D, Satava RM. Robot-assisted minimally invasive surgery—surgical robotics in the data age. *Proc IEEE*. 2022;110(7):835-46.
96. Marwaha JS, Raza MM, Kvedar JC. The digital transformation of surgery. *NPJ Digit Med*. 2023;6(1):103.
97. Zemmar A, Lozano A, Nelson B. The rise of robots in surgical environments during COVID-19. *Nat Mach Intell*. 2020;2(10):566-72.
98. Hung AJ, Liu Y, Anandkumar A. Deep learning to automate technical skills assessment in robotic surgery. *JAMA Surg*. 2021;156(11):1059-60.
99. Marcus HJ, Ramirez PT, Khan DZ, Layard Horsfall H, Hanrahan JG, Williams SC, et al. The IDEAL framework for surgical robotics: development, comparative evaluation and long-term monitoring. *Nat Med*. 2024;30(1):61-75.
100. Yip M, Salcudean S, Goldberg K, Althoefer K, Mencias A, Opfermann JD, et al. Artificial intelligence meets medical robotics. *Science*. 2023;381(6654):141-6.
101. Khan Y, Ostfeld AE, Lochner CM, Pierre A, Arias AC. Monitoring of vital signs with flexible and wearable medical devices. *Adv Mater*. 2016;28(22):4373-95.
102. Yuan Y, Chen J, Zhang D, Geng R, Gong H, Xu G, et al. Atrial fibrillation detection via contactless radio monitoring and knowledge transfer. *Nat Commun*. 2025;16(1):4317.
103. Kadirvelu B, Gavriel C, Nageshwaran S, Chan JPK, Nethisinghe S, Athanasopoulos S, et al. A wearable motion capture suit and machine learning predict disease progression in Friedreich's ataxia. *Nat Med*. 2023;29(1):86-94.
104. Ricotti V, Kadirvelu B, Selby V, Festenstein R, Mercuri E, Voit T, et al. Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. *Nat Med*. 2023;29(1):95-103.
105. Dentamaro V, Impedovo D, Musti L, Pirlo G, Taurisano P. Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database. *Sci Rep*. 2024;14(1):20941.
106. Schalkamp AK, Peall KJ, Harrison NA, Sandor C. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. *Nat Med*. 2023;29(8):2048-56.
107. Yang Y, Yuan Y, Zhang G, Wang H, Chen YC, Liu Y, et al. Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nat Med*. 2022;28(10):2207-15.
108. Choi Y, Jin P, Lee S, Song Y, Tay RY, Kim G, et al. All-printed chip-less wearable neuromorphic system for multimodal physicochemical health monitoring. *Nat Commun*. 2025;16(1):5689.
109. Dunn J, Kidzinski L, Runge R, Witt D, Hicks JL, Schüssler-Florenz Rose SM, et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. *Nat Med*. 2021;27(6):1105-12.
110. Wong F, de la Fuente-Nunez C, Collins JJ. Leveraging artificial intelligence in the fight against infectious diseases. *Science*. 2023;381(6654):164-70.
111. Oeschger TM, McCloskey DS, Buchmann RM, Choubal AM, Boza JM, Mehta S, et al. Early warning diagnostics for emerging infectious diseases in developing into late-stage pandemics. *Acc Chem Res* 2021;54(19):3656-66.
112. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inf Assoc*. 2008;15(2):150-7.
113. Brownstein JS, Rader B, Astley CM, Tian H. Advances in artificial intelligence for infectious-disease surveillance. *N Engl J Med*. 2023;388(17):1597-607.
114. Schwalbe N, Wahl B. Artificial intelligence and the future of global

- health. *Lancet*. 2020;395(10236):1579-86.
115. Kim H, Ryu S, Jung N, Yang J, Seok C. CSearch: chemical space search via virtual synthesis and global optimization. *J Cheminf*. 2024;16(1):137.
 116. Kraemer MU, Tsui JLH, Chang SY, Lytras S, Khurana MP, Vanderslott S, *et al*. Artificial intelligence for modelling infectious disease epidemics. *Nature*. 2025;638(8051):623-35.
 117. Yang J, Dung NT, Thach PN, Phong NT, Phu VD, Phu KD, *et al*. Generalizability assessment of AI models across hospitals in a low-middle and high income country. *Nat Commun*. 2024;15(1):8270.
 118. Reddy CL, Mitra S, Meara JG, Atun R, Afshar S. Artificial intelligence and its role in surgical care in low-income and middle-income countries. *Lancet Digit Health*. 2019;1(8):e384-6.
 119. Venkateswaran M, Ghanem B, Abbas E, Khader KA, Ward IA, Awwad T, *et al*. A digital health registry with clinical decision support for improving quality of antenatal care in Palestine (eRegQual): a pragmatic, cluster-randomised, controlled, superiority trial. *Lancet Digit Health*. 2022;4(2):e126-36.
 120. Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *NPJ Digit Med*. 2022;5(1):162.
 121. Huang J, Yang DM, Rong R, Nezafati K, Treager C, Chi Z, *et al*. A critical assessment of using ChatGPT for extracting structured data from clinical notes. *NPJ Digit Med*. 2024;7(1):106.
 122. Kather JN, Ferber D, Wiest IC, Gilbert S, Truhn D. Large language models could make natural language again the universal interface of healthcare. *Nat Med*. 2024;30(10):2708-10.
 123. Fahrner LJ, Chen E, Topol E, Rajpurkar P. The generative era of medical AI. *Cell*. 2025;188(14):3648-60.
 124. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-40.
 125. Shea YF, Lee CMY, Ip WCT, Luk DWA, Wong SSW. Use of GPT-4 to analyze medical records of patients with extensive investigations and delayed diagnosis. *JAMA Netw Open*. 2023;6(8):e2325000.
 126. Chua CE, Clara NLY, Furqan MS, Kit JLW, Makmur A, Tham YC, *et al*. Integration of customised LLM for discharge summary generation in real-world clinical settings: a pilot study on RUSSELL GPT. *Lancet Reg Health West Pac*. 2024;51:101211.
 127. Van Veen D, Van Uden C, Blankemeier L, Delbrouck JB, Aali A, Bluethgen C, *et al*. Adapted large language models can outperform medical experts in clinical text summarization. *Nat Med*. 2024;30(4):1134-42.
 128. Croxford E, Gao Y, First E, Pellegrino N, Schnier M, Caskey J, *et al*. Evaluating clinical AI summaries with large language models as judges. *NPJ Digit Med*. 2025;8(1):640.
 129. Menezes MCS, Hoffmann AF, Tan AL, Nalbandyan M, Omenn GS, Mazzotti DR, *et al*. The potential of generative pre-trained transformer 4 (GPT-4) to analyse medical notes in three different languages: a retrospective model-evaluation study. *Lancet Digit Health*. 2025;7(1):e35-43.
 130. Goh E, Gallo RJ, Strong E, Weng Y, Kerman H, Freed JA, *et al*. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. *Nat Med*. 2025;31(4):1233-8.
 131. Tordjman M, Liu Z, Yuze M, Fauveau V, Mei Y, Hadjadj J, *et al*. Comparative benchmarking of the DeepSeek large language model on medical tasks and clinical reasoning. *Nat Med*. 2025;31(8):2550-5.
 132. Sandmann S, Hegselmann S, Fujarski M, Bickmann L, Wild B, Eils R, Varghese J. Benchmark evaluation of DeepSeek large language models in clinical decision-making. *Nat Med*. 2025;31(8):2546-9.
 133. Liu X, Liu H, Yang G, Jiang Z, Cui S, Zhang Z, *et al*. A generalist medical language model for disease diagnosis assistance. *Nat Med*. 2025;31(3):932-42.
 134. Luo Y, Zhang J, Fan S, Yang K, Hong M, Wu Y, *et al*. Biomedgpt: an open multimodal large language model for biomedicine. *IEEE J Biomed Health Inform*. 2024:PP. doi: 10.1109/JBHI.2024.3505955.
 135. Wan P, Huang Z, Tang W, Nie Y, Pei D, Deng S, *et al*. Outpatient reception via collaboration between nurses and a large language model: a randomized controlled trial. *Nat Med*. 2024;30(10):2878-85.
 136. Hou Z, Wu Z, Qu Z, Gong L, Peng H, Jit M, *et al*. A vaccine chatbot intervention for parents to improve HPV vaccination uptake among middle school girls: a cluster randomized trial. *Nat Med*. 2025:1-8.
 137. Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, *et al*. A study of generative large language model for medical research and healthcare. *NPJ Digit Med*. 2023;6(1):210.
 138. Liu F, Zhu T, Wu X, Yang B, You C, Wang C, *et al*. A medical multimodal large language model for future pandemics. *NPJ Digit Med*. 2023;6(1):226.
 139. Li J, Guan Z, Wang J, Cheung CY, Zheng Y, Lim LL, *et al*. Integrated image-based deep learning and language models for primary diabetes care. *Nat Med*. 2024;30(10):2886-96.
 140. Lu MY, Chen B, Williamson DF, Chen RJ, Zhao M, Chow AK, *et al*. A multimodal generative AI copilot for human pathology. *Nature*. 2024;634(8033):466-73.
 141. Pais C, Liu J, Voigt R, Gupta V, Wade E, Bayati M. Large language models for preventing medication direction errors in online pharmacies. *Nat Med*. 2024;30(6):1574-82.
 142. Gupta S, Basu A, Nieves M, Thomas J, Wolfrath N, Ramamurthi A, *et al*. PRISM: patient records interpretation for semantic clinical trial matching system using large language models. *NPJ Digit Med*. 2024;7(1):305.
 143. Li CY, Chang KJ, Yang CF, Wu HY, Chen W, Bansal H, *et al*. Towards a holistic framework for multimodal LLM in 3D brain CT radiology report generation. *Nat Commun*. 2025;16(1):2258.
 144. Singhal K, Tu T, Gottweis J, Sayres R, Wulczyn E, Amin M, *et al*. Toward expert-level medical question answering with large language models. *Nat Med*. 2025;31(3):943-50.
 145. Shashikumar SP, Mohammadi S, Krishnamoorthy R, Patel A, Wardi G, Ahn JC, *et al*. Development and prospective implementation of a large language model based system for early sepsis prediction. *NPJ Digit Med*. 2025;8(1):290.
 146. Zhu M, Lin H, Jiang J, Jinia AJ, Jee J, Pichotta K, *et al*. Large language model trained on clinical oncology data predicts cancer progression. *NPJ Digit Med*. 2025;8(1):397.
 147. Wu C, Qiu P, Liu J, Gu H, Li N, Zhang Y, *et al*. Towards evaluating and building versatile large language models for medicine. *NPJ Digit Med*. 2025;8(1):58.
 148. Wang G, Liu X. Medical large language model for diagnostic reasoning across specialties. *Nat Med*. 2025;31(3):743-4.
 149. Shi D, Zhang W, Yang J, Huang S, Chen X, Xu P, *et al*. A multimodal visual-language foundation model for computational ophthalmology. *NPJ Digit Med*. 2025;8(1):381.
 150. Yao J, Wang Y, Lei Z, Wang K, Feng N, Dong F, *et al*. Multimodal GPT model for assisting thyroid nodule diagnosis and

- management. *NPJ Digit Med*. 2025;8(1):245.
151. Bai L, Wu Y, Li G, Zhang W, Zhang H, Su J. AI-enabled organoids: construction, analysis, and application. *Bioact Mater*. 2024;31:525-48.
 152. Omar M, Soffer S, Agbareia R, Bragazzi NL, Apakama DU, Horowitz CR, et al. Sociodemographic biases in medical decision making by large language models. *Nat Med*. 2025:1-9.
 153. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. *Lancet Digit Health*. 2024;6(1):e12-22.
 154. Johri S, Jeong J, Tran BA, Schlessinger DI, Wongvibulsin S, Barnes LA, et al. An evaluation framework for clinical use of large language models in patient interaction tasks. *Nat Med*. 2025;31(1):77-86.
 155. Alber DA, Yang Z, Alyakin A, Yang E, Rai S, Valliani AA, et al. Medical large language models are vulnerable to data-poisoning attacks. *Nat Med*. 2025;31(2):618-26.
 156. 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):1233-9.
 157. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat Med*. 2025;31(1):60-9.
 158. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput Surv*. 2021;54(6):1-35.
 159. Reddy S. Explainability and artificial intelligence in medicine. *Lancet Digit Health*. 2022;4(4):e214-5.
 160. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digit Med*. 2023;6(1):113.
 161. 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):447-53.
 162. Corfmat M, Martineau JT, Régis C. High-reward, high-risk technologies? An ethical and legal account of AI development in healthcare. *BMC Med Ethics*. 2025;26(1):4.
 163. Laein GD. Global perspectives on governing healthcare AI: prioritising safety, equity and collaboration. *BMJ Leader*. 2024;9(1):e000904.
 164. Mihan A, Pandey A, Van Spall HG. Mitigating the risk of artificial intelligence bias in cardiovascular care. *Lancet Digit Health*. 2024;6(10):e749-54.
 165. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digit Med*. 2020;3(1):81.
 166. Seyyed-Kalantari L, Zhang H, McDermott MB, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med*. 2021;27(12):2176-82.
 167. Bouguettaya A, Stuart EM, Aboujaoude E. Racial bias in AI-mediated psychiatric diagnosis and treatment: a qualitative comparison of four large language models. *NPJ Digit Med*. 2025;8(1):332.
 168. Smith B, Khojandi A, Vasudevan R. Bias in reinforcement learning: a review in healthcare applications. *ACM Comput Surv*. 2023;56(2):1-17.
 169. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv*. 2018;51(5):1-42.
 170. Joyce DW, Kormilitzin A, Smith KA, Cipriani A. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *NPJ Digit Med*. 2023;6(1):6.
 171. Reyes M, Meier R, Pereira S, Silva CA, Dahlweid FM, Tengg-Kobligk Hv, et al. On the interpretability of artificial intelligence in radiology: challenges and opportunities. *Radiol Artif Intell*. 2020;2(3):e190043.
 172. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. *Lancet Digit Health*. 2020;2(4):e179-91.
 173. Chen H, Gomez C, Huang CM, Unberath M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. *NPJ Digit Med*. 2022;5(1):156.
 174. Hossain MI, Zamzmi G, Mouton PR, Salekin MS, Sun Y, Goldgof D. Explainable AI for medical data: current methods, limitations, and future directions. *ACM Comput Surv*. 2025;57(6):1-46.
 175. Holzinger A, Malle B, Saranti A, Pfeifer B. Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI. *Inf Fusion*. 2021;71:28-37.
 176. Alderman JE, Palmer J, Laws E, McCradden MD, Ordish J, Ghassemi M, et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING together consensus recommendations. *Lancet Digit Health*. 2025;7(1):e64-88.
 177. Klauschen F, Dippel J, Keyl P, Jurmeister P, Bockmayr M, Mock A, et al. Toward explainable artificial intelligence for precision pathology. *Annu Rev Pathol: Pathol Mech Dis*. 2024;19(1):541-70.
 178. Wang J, Wang K, Yu Y, Lu Y, Xiao W, Sun Z, et al. Self-improving generative foundation model for synthetic medical image generation and clinical applications. *Nat Med*. 2025;31(2):609-17.
 179. Yang J, Soltan AA, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Digit Med*. 2022;5(1):69.
 180. Price WN, Cohen IG. Privacy in the age of medical big data. *Nat Med*. 2019;25(1):37-43.
 181. Kaissis G, Ziller A, Passerat-Palmbach J, Ryffel T, Usynin D, Trask A, et al. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat Mach Intell*. 2021;3(6):473-84.
 182. Yang Y, Lyu J, Wang R, Wen Q, Zhao L, Chen W, et al. A digital mask to safeguard patient privacy. *Nat Med*. 2022;28(9):1883-92.
 183. Zhang F, Zhai D, Bai G, Jiang J, Ye Q, Ji X, et al. Towards fairness-aware and privacy-preserving enhanced collaborative learning for healthcare. *Nat Commun*. 2025;16(1):2852.
 184. Dou Q, So TY, Jiang M, Liu Q, Vardhanabhuti V, Kaissis G, et al. Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. *NPJ Digit Med*. 2021;4(1):60.
 185. Busch F, Kather JN, Johner C, Moser M, Truhn D, Adams LC, et al. Navigating the European Union artificial intelligence act for healthcare. *NPJ Digit Med*. 2024;7(1):210.
 186. Laux J, Wachter S, Mittelstadt B. Trustworthy artificial intelligence and the European Union AI act: on the conflation of trustworthiness and acceptability of risk. *Regul Gov*. 2024;18(1):3-32.
 187. Forcier MB, Gallois H, Mullan S, Joly Y. Integrating artificial intelligence into health care through data access: can the GDPR act as a beacon for policymakers?. *J Law Biosci*. 2019;6(1):317-35.

188. Lawlor RT. The impact of GDPR on data sharing for European cancer research. *Lancet Oncol.* 2023;24(1):6-8.
189. Yang SR, Chien JT, Lee CY. Advancements in clinical evaluation and regulatory frameworks for AI-driven software as a medical device (SaMD). *IEEE Open J Eng Med Biol.* 2024;6:147-51.
190. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. *JAMA.* 2023;330(4):311-2.
191. Park SH, Han K, Jang HY, Park JE, Lee JG, Kim DW, *et al.* Methods for clinical evaluation of artificial intelligence algorithms for medical diagnosis. *Radiology.* 2023;306(1):20-31.
192. Liu F, Zhou H, Gu B, Zou X, Huang J, Wu J, *et al.* Application of large language models in medicine. *Nat Rev Bioeng.* 2025:1-20.
193. D'Souza M, Papadopoulou A, Girardey C, Kappos L. Standardization and digitization of clinical data in multiple sclerosis. *Nat Rev Neurol.* 2021;17(2):119-25.
194. Prasad D, Khandeshi A, Sartin S, Jain R, Dahdaleh N, Lesniak M, *et al.* Will AI become our Co-PI?. *NPJ Digit Med.* 2025;8(1):440.
195. Reverberi C, Rigon T, Solari A, Hassan C, Cherubini P, Cherubini A. Experimental evidence of effective human-AI collaboration in medical decision-making. *Sci Rep.* 2022;12(1):14952.
196. Park HJ, Kim SH, Choi JY, Cha D. Human-machine cooperation meta-model for clinical diagnosis by adaptation to human expert's diagnostic characteristics. *Sci Rep.* 2023;13(1):16204.
197. Bai L, Su J. Artificial intelligence virtual organoids (AIVOs). *Bioact Mater.* 2026;59:45-68.
198. Yang T, Wang YY, Ma F, Xu BH, Qian HL. Build the virtual cell with artificial intelligence: a perspective for cancer research. *Mil Med Res.* 2025;12(1):4.

<https://doi.org/10.1016/j.jmmr.2026.100012>

Cite this article as: Zhang C, Wang FX, Tang X, Li JL, Ding N, Hong Y, *et al.* Virtual medicine: medical AI in human health and diseases. *Mil Med Res.* 2026;13(1):100012.