CellPress

Cell
Leading Edge

Review

The generative era of medical AI

L. John Fahrner,[1,3] Emma Chen,[1,3] Eric Topol,[2,4,5] and Pranav Rajpurkar[1,4,5,*]
[1]Department of Biomedical Informatics, Harvard Medical School, Cambridge, MA, USA
[2]Scripps Research, La Jolla, CA, USA
[3]These authors contributed equally
[4]These authors contributed equally
[5]Senior author
*Correspondence: pranav_rajpurkar@hms.harvard.edu
https://doi.org/10.1016/j.cell.2025.05.018

SUMMARY

Rapid advancements in artificial intelligence (AI), particularly large language models (LLMs) and multimodal AI, are transforming medicine through enhancements in diagnostics, patient interaction, and medical forecasting. LLMs enable conversational interfaces, simplify medical reports, and assist clinicians with decision making. Multimodal AI integrates diverse data like images and genetic data for superior performance in pathology and medical screening. AI-driven tools promise proactive, personalized healthcare through continuous monitoring and multiscale forecasting. However, challenges like bias, privacy, regulatory hurdles, and integration into healthcare systems must be addressed for widespread clinical adoption.

INTRODUCTION

Technological innovation in biomedicine has directly contributed to improved quality of life and extended healthspan. Historically, advances in drug development, surgical techniques, understanding of biological pathways, imaging techniques, and other areas have propelled this progress. Now we are on the verge of a new phase of growth with the recent progress in artificial intelligence (AI), which we will attempt to summarize here. The weekly Doctor Penguin newsletter has continued to track novel developments in medical and health AI since 2019 and serves as a source of material for this review (https://doctorpenguin.substack.com). From a technical perspective, modern AI advancements have been enabled by several key architectural innovations, including the Transformer architecture, generative adversarial networks, and diffusion models, which together have powered the development of increasingly sophisticated generative AI systems. Research has shown the potential for transformative change because of large language models (LLMs) and multimodal AI, the changing medical practice, and multiscale medical forecasting; this review aims to summarize this seemingly exponential progress over the last 3 years. We will discuss the background, implementation, implications, and some of the persistent challenges associated with these new technologies.

LLMs AND THE PATH TO MULTIMODAL MEDICINE

The promise of AI in healthcare dates to the 1960s, when Joseph Weizenbaum developed ELIZA, one of the first chatbots.[1] ELIZA simulated a Rogerian psychotherapist, engaging in simple dialogue with users. Subsequent efforts to create conversational AI for medicine were hindered by the limited capabilities of early AI systems, which were not reliable enough for real-world clinical applications.

Fast forward to today, LLMs like ChatGPT, Gemini, Claude, and Llama have captured the attention of the world. These models exemplify two important paradigms in modern AI: foundation models—large-scale, general-purpose AI systems trained on vast datasets that can be adapted to numerous downstream tasks—and generative AI, which enables the creation of novel content such as text, images, or molecular designs by modeling complex data distributions. Unlike traditional AI, which predominantly focused on discriminative classification tasks and relied on specialized architectures for specific domains, generative AI learns to generate outputs that statistically resemble the training data. This capability stems primarily from the Transformer architecture, introduced in 2017, which has redefined scalability and performance in AI.[2]

The Transformer's core innovation is self-attention, a mechanism that dynamically weighs the relevance of different input elements, allowing the model to capture long-range dependencies in data, such as relationships across sentences or protein sequences. In generative tasks, the decoder component of the Transformer is critical; it generates outputs sequentially (e.g., one word or token at a time) by attending to both the input context and previously generated elements. This architecture powers LLMs, such as those driving chatbots or text synthesis tools, which are trained on vast datasets to learn intricate patterns in language or other domains. Earlier approaches like recurrent neural networks (RNNs) and convolutional neural networks (CNNs) faced fundamental scaling bottlenecks—RNNs struggled with parallelization and long-range dependencies, while CNNs had locality biases limiting their effectiveness for sequential data.

The key discovery that Transformer models consistently improve as they grow larger accelerated the recent AI development.

Researchers found that simply increasing the model size, training data, and computing power leads to predictable gains in performance—a property not seen with previous approaches where improvements would eventually plateau.[3] This mathematical predictability, coupled with advances in specialized computing hardware and the availability of petabyte-scale datasets, established the precise conditions necessary for the current AI revolution. The convergence of these factors has positioned generative AI as a transformative tool for applications like drug discovery, clinical decision support, and automated analysis of medical literature, offering unprecedented opportunities to accelerate biomedical research.

When applied to direct patient interaction, LLMs are charting a path toward meaningful conversational AI in medicine. In this application, LLMs can provide patients with accessible conversational interfaces to interact directly with their own individual health data in the electronic health record (EHR) and also with general medical information.[4–6] For example, LLM agents have been used to reduce the complexity of pathology reports and for transforming hospital discharge summaries into a patient-friendly format.[7,8] When mental health chatbots are made available to patients, they have shown potential in reducing stigma about mental health care and have demonstrated increased referral rates, most significantly for traditionally underrepresented groups.[9] These results are notable, as the first steps of seeking care and receiving an appropriate referral are common barriers in the mental health pathway. These conversational agents can assist patients in navigating their healthcare course, providing personalized information and support. While many LLM tools await medical approval, early reports suggest patients are already testing their benefits; in one example, a mother was able to diagnose her young son's tethered cord after multiple fruitless physician visits.[10]

In addition to conversational and summarization agents, LLMs can be tools for clinicians.[11] Models in the research setting have demonstrated performance at least comparable to clinicians in history-taking, following diagnostic pathways, communication, and empathy.[11–14] LLMs can also serve as medical knowledge resources for clinicians. Dedicated LLMs have been developed for specialized fields, enabling clinicians to access expert knowledge and to assist with decision-making and guideline adherence, and have already gained certification.[15–17] Models now routinely achieve passing scores on medical licensing exams, showcasing their potential to provide up-to-date and comprehensive medical information.[18–21] By leveraging the vast knowledge encapsulated within LLMs, these diagnostic tools promise to aid clinicians in making accurate and timely diagnoses and guiding management decisions.

Beyond these applications, LLMs could be integrated into healthcare delivery to automate documentation tasks and improve clinician efficiency. AI-powered "scribes" are capable of recording patient histories; creating medical notes; handling pre-authorization requests for medications or tests; scheduling follow-up appointments; and managing lab test results, scans, procedures, billing, and more and are already being used clinically.[11,22–25] Notably, LLMs have shown the ability to summarize medical information as effectively as human experts. Much of the emerging research on LLMs focuses on "agentic" environments, in which AI systems dynamically complete complex tasks by interacting with existing systems, humans, or other agents. Agentic systems promise to automate workflows, validate AI safety and reduce errors, aid in managing disparate AI tools, and to provide outcomes predictions, among other skills.[26] Polaris AI exemplifies this agentic approach through its "constellation architecture," where a primary conversational agent works in concert with specialized LLM agents—including medication specialists that verify dosages, labs specialists that analyze test results, and nutrition specialists that provide tailored dietary guidance—enabling the system to maintain both engaging conversation and medical accuracy while ensuring built-in safety redundancies for healthcare interactions.[27]

Recent advancements in chain-of-thought prompting and reasoning techniques have addressed the challenge of ensuring accurate and clinically relevant outputs from LLMs.[28,29] These approaches have facilitated the development of datasets optimized for reasoning and, more recently, specialized reasoning models.[30–32] As these techniques evolve, large reasoning models are poised to become increasingly prevalent in clinical applications.

Multimodal AI and foundation models

Early medical AI systems were dedicated single-task models predominantly trained on specific medical datasets, which required tedious manual labeling. This burden was slightly reduced by techniques like self-supervised learning (automatic interpretation of training data without explicit human labeling) and few-shot learning (more efficient learning using fewer curated examples).[33,34]

Medicine is inherently a multimodal domain, where clinical insights arise from combining radiology scans, patient records, genomic sequences, and spoken consultations.[35] Traditional AI models, often limited to single modalities, struggled to capture this complexity. Multimodal generative AI overcomes these limitations by learning unified representations across modalities, enabling a deeper understanding of medical data. A pivotal advancement in this field is the Contrastive Language-Image Pretraining (CLIP) model (introduced in 2021 by OpenAI), which uses contrastive learning to align vector representations of images, text, and potentially other modalities (e.g., audio spectrograms) into a shared latent space.[36] CLIP's architecture trains on paired data (e.g., images and captions) by maximizing the similarity between matched pairs while minimizing similarity between unmatched pairs, creating a unified space where related concepts across modalities (like a medical image of a tumor and its textual description) are positioned close together. This alignment enables generative models to process and generate multimodal outputs, such as synthesizing medical reports from imaging data or answering clinical queries by combining visual and textual inputs. Theoretically, this integration mimics human reasoning, where physicians synthesize diverse information to form diagnoses, making multimodal AI a fundamental breakthrough for medicine. By modeling statistical relationships across modalities, multimodal generative AI transforms medical applications; it enhances diagnostic accuracy by correlating imaging and clinical notes, accelerates drug discovery by integrating molecular structures with textual annotations, and personalizes treatment plans by combining patient histories with real-time sensor data. Unlike unimodal models, which risk fragmented insights, multimodal AI captures the holistic nature of medical data, driving progress in precision medicine and clinical decision support. The scalability of models like CLIP, which improve with diverse and large-scale datasets, further amplifies their impact, positioning multimodal generative AI as a cornerstone of the generative AI era in medical research.[37–44] Current multimodality AI research is focused on incorporating more modalities into a single model and incorporating volumetric datasets like magnetic resonance imaging (MRI) and computed tomography (CT) and video.

Multimodal AI is rapidly progressing in the field of pathology. Large pathology training sets can incorporate standardized images of slides and specimens with text reports, genomic data, and EHR data, providing a prime target for AI research.[45] By applying the Transformer architecture, researchers have created “vision-language” models, which incorporate these text and image-analysis components into a unified model. These models are capable of advanced tasks like image captioning and answering questions about an image. An early model used publicly available social media pathology images and captions to produce a model named protein-ligand interaction profiler (PLIP).[46] Later models used larger datasets; PathChat applied an LLM to the large UNI pathology image model to create a vision-language model, while contrastive learning from captions from histopathology (CONCH) trained natively on around 1 million image-text pairs from diverse sources.[34,47] Both approaches resulted in accurate multimodal pathology models. As foundation models become more capable, it is likely that these single larger models will continue to replace dedicated smaller models.

CHANGING MEDICAL PRACTICE

Reviewing the extensive recent research in biomedical AI allows us to anticipate the future direction of medical care. State-of-the-art AI developments point to a new model of health management in which patients have more regular and detailed insight into their health. Patients are empowered to manage their own health with AI tools that provide more timely and personalized feedback. Traditional screening programs can become more tailored and personalized. AI has the potential to enable a transformation in the delivery of healthcare by shifting more care from reactive, hospital-centric treatment to proactive, personalized, and accessible health management. Patients can be triaged to more fine-tuned levels of care, with progressive escalation as needed. Earlier diagnosis and intervention reduce the reliance on acute care resources and can lead to improved outcomes. And finally, these new AI-powered tools promise to revolutionize medical forecasting with multiscale capabilities; predictions can be made at molecular, cellular, individual, and population scales, completely rethinking the standard model of care. Powerful AI models will enable more accurate and dynamic short-term and long-term risk assessment. These new AI-powered tools promise to improve the accessibility of care, improve decision-making, and provide more targeted management.

Continuous monitoring and patient agency

Traditional wearable physical sensors such as fitness trackers have been isolated from the rest of the health ecosystem. Now, newer AI-enabled smartwatches have demonstrated the ability to identify individuals at risk of atrial fibrillation, screen for left ventricular systolic dysfunction, and to monitor cardiac function post-COVID-19 vaccination.[48–51] Devices in research settings include implantable temperature sensors for early detection of acute kidney rejection in transplant patients, sensors for continuous cortisol level detection in sweat, and wearable ultrasound devices for physiologic monitoring that use machine learning to maintain high-quality images during movement.[52–54] Some AI-powered sensors do not provide continuous monitoring but do allow for more accessible diagnosis. For example, smartphone-based diagnostic tools, like dermoscopy lenses coupled with AI models, have demonstrated high accuracy in diagnosing suspicious skin lesions, reducing the need for in-person consultations.[55] Similarly, smartphones with endoscope or otoscope attachments, combined with AI models, have been shown to assist with accurate remote diagnosis of diseases such as acute otitis media.[56] By enabling individuals to take control of their health and well-being, these novel AI-enabled sensors have the potential to regularly monitor health in the home setting to identify developing issues earlier and to reduce the burden on healthcare systems.

Advanced medical screening

AI medical screening tools can detect disease earlier and more efficiently than with traditional methods. Screening programs have traditionally targeted large populations with broad inclusion criteria. AI enables more accurate targeting of high-risk individuals to realize individual and societal benefits. For example, researchers have recalibrated low-dose lung cancer screening recommendations using AI analysis to prioritize workup for higher-risk patients and to decrease screening frequency for those at lower risk.[57]

2D and 3D mammography-based AI interpretation algorithms have matched human abilities in real-world clinical scenarios while clinical rollout continues.[58,59] Research has shown that more advanced models combining lesion detection and texture analysis to determine short-term and long-term breast cancer risks have been shown to improve overall risk assessment.[60] AI applied to traditional mammography has also been shown to help determine which patients would most benefit from supplemental MRI, reducing missed cancers without a large increase in the MRI screening burden.[61]

Photography- and video-based screening tools have been shown to provide affordable and fast analysis of complex neurological disorders, and they have also been able to go a step further and provide clinical predictions about disease progression.[62–64] For example, a retinal image-based system was developed to predict myocardial infarction, providing a less invasive screening option.[65]

Next-generation AI-powered screening promises to more accurately triage patients, improve screening efficiency, and improve predictive analytics with accessible technology (Figure 1). The final large component of this new AI-enabled paradigm in healthcare is multiscale medical forecasting.

Multiscale medical forecasting

Medical forecasting encompasses a broad set of integrated ideas in which the pattern-matching strengths of computers

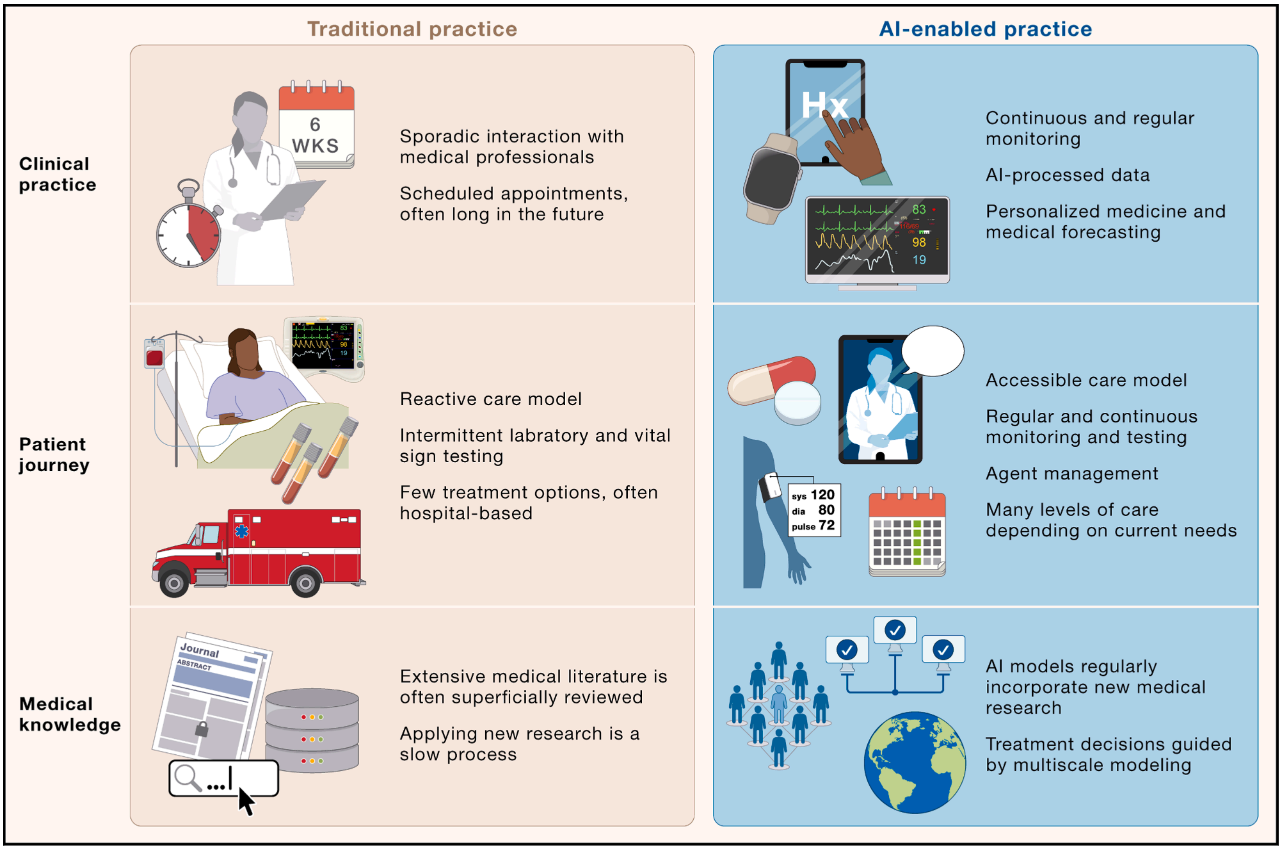

Figure 1. Transformation of medical practice
AI-enabled medical practice transforms clinical care from sporadic interactions to continuous monitoring and regular check-ins. Rather than reactive hospital-based management of more advanced diseases, medical events can be constantly addressed in familiar settings at an earlier stage. New medical knowledge can more easily be integrated into care models, while new medications are created using new AI-enabled techniques.

are paired with traditional and new data inputs to enable earlier and more precise, accurate, personalized, affordable, efficient, equitable, and convenient medical diagnosis.[66] AI algorithms are being used in medical forecasting to predict future events or outcomes based on personalized patient information after training on large datasets. Forecasting applies to the entire context of health, from the molecular level to the cellular level, the organ system level, the individual level, and to the population and global levels (Figure 2).

Molecular-level progress

The field of protein science has been revolutionized by the development of an AI model called AlphaFold2, developed by DeepMind in 2020, which achieved unprecedented accuracy in protein structure prediction.[67] This breakthrough, along with the independently developed RoseTTAFold, used attention mechanisms to predict 3D protein structures from amino acid sequences with near-experimental precision.[68] AlphaFold2's success stemmed from its innovative use of multiple sequence alignments and its ability to learn spatial relationships between amino acids.[67,69] This advancement not only accelerated structural biology research but also catalyzed developments in protein design, function prediction, and drug discovery; the developers of AlphaFold were awarded the 2024 Nobel Prize in Chemistry for this work.[70]

The impact of these advances extends beyond proof of concept. AlphaFold2 quickly expanded to include protein folding predictions for more than 200 million of the most common proteins found in over 1 million species.[71] AlphaFold3 includes more complex biomolecular structures beyond individual proteins, expanding structure prediction capabilities to protein complexes and protein-ligand interactions.[69] The pace of development of new AI tools has been staggering—within just a few weeks near the end of 2024, 10 major molecular-level research projects were released.[72] These projects included a foundation model for DNA, a tool to predict protein-protein interactions, and a Human Cell Atlas, among others.[73–75] These tools are crucial for understanding biological processes and for designing new therapeutics.[70]

However, as these models excel at predicting static structures, a significant challenge remains in capturing protein dynamics and flexibility.[76] Current research is focused on extending these models to predict not just a single structure but also the various conformations a protein might adopt under various conditions.[69] This is particularly important as subtle changes in protein folding can lead to significant physiologic differences

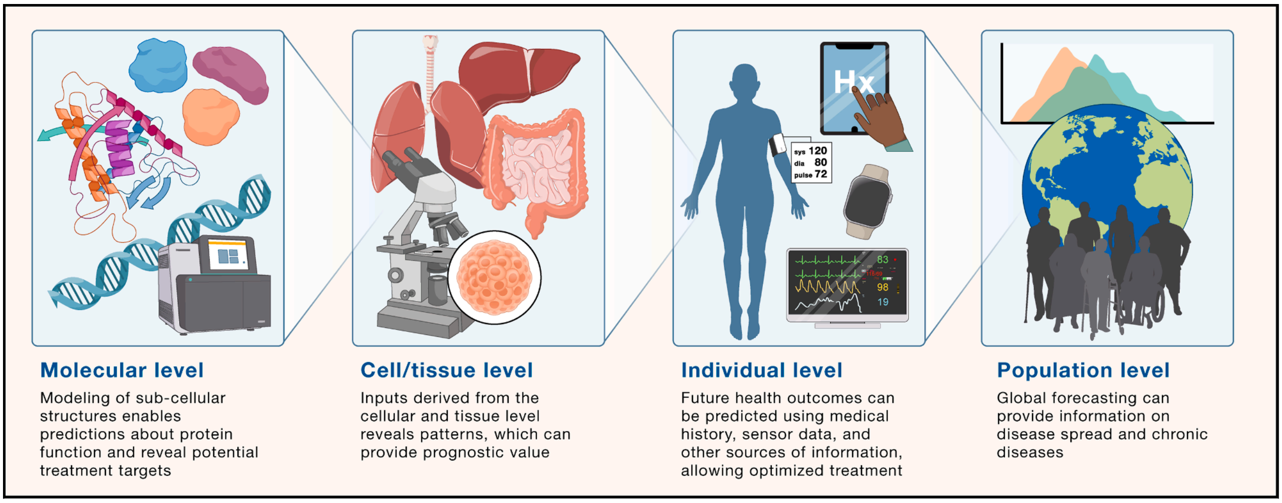

Figure 2. Multiscale medical forecasting
AI algorithms can be used in medical forecasting to predict future medical events, based on various dynamic inputs. These algorithms can be applied at multiple levels, from the molecular level to the population level.

and outcomes, with misfolded proteins implicated in diseases such as cystic fibrosis and Huntington disease.[69] Currently these generative AI tools can be used to predict clinical outcomes of various cystic fibrosis mutations.[77] AlphaFold2 has been helpful for predicting antigen proteins in pathogens like rotavirus and other infections.[78] It has also been used in immunology research to predict antibody structure and assist with vaccine development, predict membrane protein structure and interactions to assist with drug development, characterize enzyme activity for diseases like porphyria, assist with research on drug resistance, and to predict outcomes for certain acute lymphoblastic leukemia (ALL) subtypes, among many other applications.[78] From a clinical perspective, these models promise to enable personalized prediction of the impact of genetic mutations on protein function and disease pathways, as well as help us understand the nature of cancer.

Building on the foundations of protein structure prediction, the field of protein generation and design has seen significant advancements. Tools like RFdiffusion and FrameDiff use generative techniques to generate 3D structural protein backbones.[79,80] Sequence generation tools like ProGen, ProteinMPNN, and Evo can output amino acid sequences based on various inputs and allow researchers to create novel proteins with specific structures or functions. The emergence of Evo 2 also represents a milestone multimodal foundation model incorporating DNA, RNA, and proteins in one large model.[81] This shift from prediction to design represents a new frontier in protein engineering, opening possibilities for creating proteins tailored to specific tasks or environments.[82–86] These models are helping to elucidate the mechanisms of biological action from DNA to RNA to protein, and they enable researchers to efficiently manipulate steps in these processes.

As AI continues to reshape protein science, the field is moving toward more integrated, multiscale approaches. The next frontier will likely involve developing AI systems that can not only design individual proteins but can also engineer entire protein networks or cellular systems, pushing the boundaries of synthetic biology and potentially enabling new approaches to treating complex diseases.

Cellular, organ-system, and individual-level forecasting

Cardiology

A wide range of experimental tools in cardiology can benefit from the pattern-matching capabilities of AI techniques. In the acute setting, AI models have the potential to alert clinicians to developing decompensation. For example, Lin et al. developed an electrocardiogram (ECG) model that monitored 12-lead ECGs of hospitalized patients and was able to alert providers of impending decompensation and to improve clinical outcomes.[87] Other models have been able to identify patients at risk of hypotension, tachycardia, or hypoxia, based on standard vital sign monitors. Researchers were able to use ECG data to detect a pattern of occlusion myocardial infarction even without ST elevation, surpassing human abilities and allowing for earlier intervention.[88] Sundrani et al. developed a bimodal model to predict tachycardia, hypotension, or hypoxia in the emergency department (ED), based on triage data and ECG/pulse plethysmograph (PPG) waveforms.[89]

AI-powered models have been developed to forecast the risk of future cardiovascular disease events such as heart attacks or strokes, based on various factors such as age, gender, and medical history. By identifying a unique set of patient variables from a potential list of thousands of variables, models have been shown to more accurately predict coronary artery disease (CAD) risk than was previously possible.[90] In another model, researchers were able to identify 27 specific proteins in blood samples that can be used to create a personalized survival model that is more accurate than previous methods.[91] ECG analysis can be used to predict the risk of future atrial fibrillation or LV dysfunction after percutaneous coronary intervention (PCI), which in turn predicts which patients would most benefit from medical intervention.[92,93]

Cardiovascular disease risks can be estimated using imaging techniques. An AI tool called EchoCLIP was able to characterize subtle clinically significant changes over time on echocardiograms, which would be difficult for a human interpreter.[94] The timing of future arrhythmic sudden death can be predicted based on myocardial scarring seen on MRI.[95] Coronary artery CTA studies are time consuming studies when interpreted manually, but Lin et al. were able to automate the process and show prognostic value for predicting future myocardial infarction.[96] Coronary CTA can also show perivascular fat inflammation, allowing researchers to create an AI algorithm to estimate the risk of future cardiac events even when there is no obstructive coronary disease.[97]

Radiology, oncology, and other fields

In radiology, AI algorithms have been applied to standard MRI or CT studies to identify subtle image texture patterns that are not detectable by human clinicians. For example, researchers were able to use MRI data to reliably classify pediatric medulloblastoma into four subtypes based on image characteristics alone, facilitating the development of treatment regimens when there is no access to molecular testing.[98] AI-based tools have similarly been developed for classifying lung cancer, breast cancer, neuroendocrine tumor, gastrointestinal stromal tumor, colorectal cancer, and other tumors, and they can be used to predict histopathology, grading, metastatic potential, and other clinically useful characteristics.[99–103]

Research in the field of oncology has made significant strides in leveraging large multimodal AI models for automated analysis of whole-slide pathology images. AI models have been shown to help determine susceptibility to chemotherapy agents in pancreatic adenocarcinoma by analyzing subtle morphological features in the tumor microenvironment, ultimately informing clinical outcomes.[104,105] AI has facilitated the development of tools like tumor origin differentiation using cytological histology (TORCH), which can more reliably identify the origin of cancers with unknown primary sites, using cytological samples from pleural and peritoneal fluid.[106] Models have been developed that can predict the risk of a patient developing pancreatic adenocarcinoma based on the patient's historical diagnoses and trajectory of diseases.[107] An AI algorithm trained on pancreatic cancer patients was able to predict future complications, following pancreatic resection, and was able to show reduced mortality by approximately 50% at 90 days when compared with the usual care in which the clinician did not have access to the algorithm.[108]

Other AI-based tools have been able to extract additional clinically relevant information from traditional sources. For example, a tool called RETFound used fundus photography and retinal optical coherence tomography to predict the presence of systemic conditions like heart failure and myocardial infarction in addition to more predictably identifying sight-threatening diseases of the retina.[109]

Harnessing contactless sensor data through the application of AI allows "ambient intelligence," in which the patterns of sensors are interpreted by AI algorithms to learn about the surrounding environment and patient movement; this has the potential to improve patient safety and clinical efficiency.[110] For example, Liu et al. developed a low-power radio-based sensor to monitor gait.[111] This device was able to identify statistically significant declines in gait speed in Parkinson disease patients, providing clinically helpful information about progression of the disease. Other contactless sensors such as cameras have been used to track neurodegenerative disease progression and even to identify the likely molecular etiology in patients with Friedreich's ataxia[62] and Duchenne muscular dystrophy,[112] allowing for earlier diagnosis, intervention, and personalized treatment plans.

AI models based on EHR data are capable of predicting readmission, mortality, and length-of-stay.[89,113] An AI model trained on EHR data was able to predict the International Classification of Diseases (ICD) codes of a patient's next visit, increasing the ability to predict uncommon outcomes like pancreatic cancer and self-harm.[114] Models have been able to predict seizure recurrence risk in pediatric patients, based on routine clinical notes, chart messages, and diagnostic studies.[115]

Population-level forecasting

Medical resources are fundamentally limited in our current medical system. Global forecasting allows for the optimal distribution of resources in order to provide the most benefit. For example, by modeling brain aging across populations, researchers are able to identify and address geographic, socioeconomic, and health factors, which are associated with increased risks of dementia.[116] Modeling the global spread of infections can provide information about where a disease may next present. And by modeling weather and population data, governments can anticipate impending heatstroke events.[117]

CHALLENGES AND LIMITATIONS

AI has the potential to transform healthcare delivery, but multiple challenges and limitations need to be addressed before its potential can be realized.

LLM development and use presents challenges. Much of what has been shown is not based on real-world, prospective studies, but it is instead simulated with patient actors and theoretical cases. As the infrastructure supporting LLM deployment evolves, with the development of foundation models and standardized benchmarking,[19] challenges related to accuracy, bias, privacy, and ethics persist.[24] Earlier LLMs were prone to "hallucinating" information, although recent efforts have shown promise in mitigating this issue.[118,119] Bias in training data and the tendency of LLMs to accept input text as truthful can also limit output accuracy.[24,120] According to Han et al., medical LLM models currently do not meet standards of general or medical safety, although efforts to improve safety have been promising.[121] Combining human skills with AI tools has the potential to improve care.[122,123] However, more research is needed to understand how to effectively integrate these tools into medical workflows.[120,124]

Another fundamental challenge lies in the development and validation of AI models. Older models were relatively small and used a smaller set of curated training data. Determining how to regulate these tools is and was a challenge; these models do not necessarily generalize well across different populations, again raising concerns about bias. Significant research and resources will be required to ensure that AI models are appropriate for any given situation or population. Larger and more diverse datasets may be required for training to ensure accurate

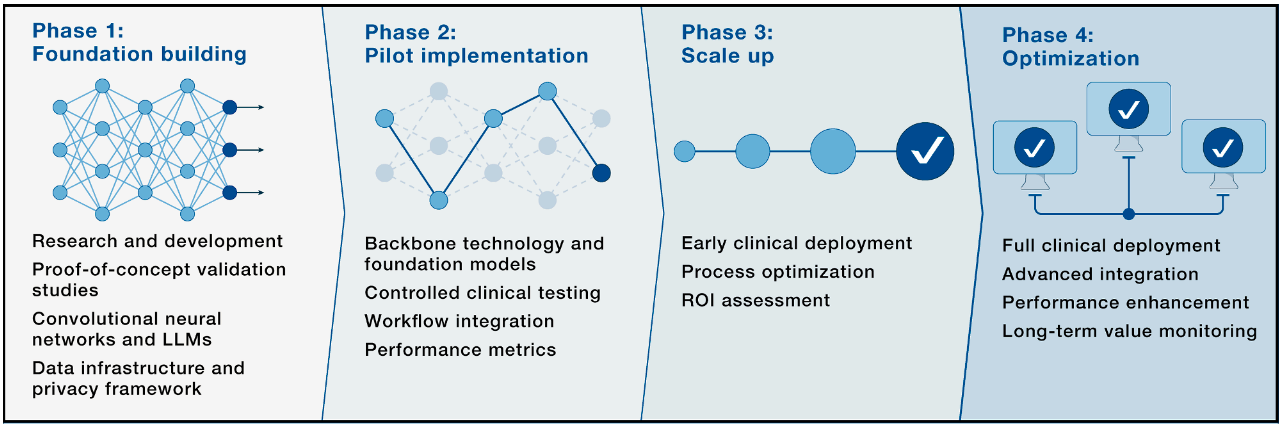

Figure 3. Medical AI implementation roadmap
Basic science research slowly led to proof-of-concept models. Larger models and early clinical deployment can open the door to eventual clinical deployment and optimization.

performance. To address the complexity of regulating regularly updated AI tools, the Food and Drug Administration (FDA) has developed the Predetermined Change Control Plan (PCCP) framework, in which a vendor undergoes initial certification for a product but then is able to update the product within certain boundaries.[125] This acknowledges the benefits of product updates while it maintains safety standards and avoids overwhelming FDA resources. Multiple other frameworks and entities can be used for regulation of AI products.

Generative AI models present significantly more challenges from a regulatory perspective.[125] It becomes more unclear about what data were used for training, how the model performs on any one task or population, and how the output varies based on the non-deterministic nature of generative AI tools. Additionally, any updates to a model or training data can significantly alter the model's output. Medical device companies have begun to integrate generative AI features into their products, although more integral use and approval of generative AI remain in question.[126] As AI tools become more generally capable and more unknowable, evaluating them may begin to look more like the evaluation of physicians, incorporating licensing exams and monitoring.[125,127]

Most AI models today are not transparent enough about their design and datasets to allow for recreations of the model. The black-box nature of these tools reduces understanding of the mechanisms of the model and uses and limitations of the outputs. Additionally, current AI models may not be truly reasoning but rather repeating trained information. This limitation raises questions about their reliability in novel clinical scenarios.[128] Many prediction and forecasting studies are retrospective and may not be rigorous enough for clinical implementation. It may be years before widespread clinical benefit can be realized or proven for some of these tools, as the scientific and healthcare communities adapt to the new capabilities. Furthermore, when implementing a new model, questions arise about how to validate, test, and continuously update it with new information or advancements.

The pathway to integrating AI tools into existing healthcare systems presents challenges. Healthcare systems worldwide are complex, often outdated, and heterogeneous. Data integration between devices and into preexisting IT systems will require significant effort. One solution is unlikely to work everywhere. The infrastructure must be in place to seamlessly incorporate this information into EHRs and decision-making processes. Incorporating AI models into practice remains a long and arduous process (Figure 3).

The role of physicians is likely to evolve. In the future, physicians may become orchestrators or directors, managing the most complex cases and overseeing other providers who review the more routine cases. Physicians could manage departmental operations, quality control, tumor boards, and procedures, while also bearing legal liability. Research has shown that different physicians can respond to AI tools variably, highlighting the need for additional research into understanding the use of AI tools.[124,129] Healthcare providers must be trained to interpret and act upon the data generated by these tools. Skepticism or hesitancy about new AI tools is inevitable and will need to be addressed, particularly when considering the perception of potential job loss or loss of autonomy. Resources such as the AI for Medicine Specialization courses in Coursera (https://www.coursera.org/specializations/ai-for-medicine), Udacity courses, and books like Co-Intelligence: Living and Working with AI by Mollick and The AI Revolution in Medicine by Lee can serve as starting points.[130,131]

The costs of implementing new AI tools must be addressed to ensure accessibility and prevent exacerbation of healthcare disparities. It is unclear where the expenses for AI tools will fall—on patients, populations, or third parties. In the US, the dominant fee-for-service model leaves open the question of reimbursement by the government or private insurers. Most reimbursement for medical care currently is initiated by current procedural terminology (CPT) or diagnosis-related groups (DRG) codes; these codes generally do not cover AI tools in their current form. A few early AI tools received temporary coverage in part because of their novelty, but a more robust and predictable system of reimbursement must be developed, tested, and implemented if fee-for-service coverage continues. Generalist medical AI (GMAI) tools may not fit well into the traditional fee-for-service model, however.

Possible alternative reimbursement models for GMAI include assigning an overarching care management coordinated activity code for assisting with existing clinical services or value-based reimbursement. Proving clinical and economic value for AI tools is more complicated and multifaceted than would be initially expected, however. Advanced statistical analysis techniques may be required to assess the value or return on investment of a given AI project or tool. Survival, quality of life, and other variables can be weighted to form a proxy for the "value" of a high-investment technology. Finally, the realized value of a new tool often lags the implementation, as users learn how to incorporate the new information and optimize patient selection. Health systems and physicians are unlikely to implement AI tools on a larger scale if there is no quantifiable benefit. The overall impact on the healthcare system remains uncertain.

Privacy and data security are paramount as the medical system relies on increasingly connected technology. AI tools trained on real-world data carry the additional risk of exposing patient information from the training set. As these technologies become more powerful, the scientific and clinical communities are actively developing guidelines for responsible use.

A significant challenge lies in the disparity between the pace of AI development and traditional medical progress. While medical advancements often occur slowly and methodically, AI research has progressed at a whirlwind pace. This disconnect poses challenges for integration and regulation. Regulatory agencies, often under-resourced, may struggle to adapt quickly enough to this rapidly evolving field.

CONCLUSION

Over the past few years substantial progress has been made in the use of AI in health and medicine. The future of medicine incorporates tools that can process vast amounts of information on every scale and has the potential to meaningfully improve diagnostic accuracy and patient outcomes. AI advancements like advanced screenings, innovative imaging technologies, predictive analytics in medical forecasting, and personalized management plans promise to transform patient care from a reactive hospital-based model to a proactive health-optimizing system with fine-tuned levels of intervention.

Despite these promises, full clinical acceptance and regular sanctioned use of AI tools are not imminent. Serious challenges remain and will need to be addressed before there is widespread adoption of AI in clinical practice. Most AI tools are still in developmental phases. While some show clinical benefit in a controlled setting, few can claim to unequivocally improve health for all users. Also, few can claim to clearly reduce costs in all settings, and few can claim to have a clear path to implementation in the current medical systems. Clinical implementation remains the main hurdle to more widespread use of AI tools by health professionals.

ACKNOWLEDGMENTS

We thank Sid Dogra, Vish Rao, Rohit Reddy, and Hong-Yu Zhou for their feedback. E.T. was funded by the National Institutes of Health National grant UL1 TR001114.

AUTHOR CONTRIBUTIONS

L.J.F. wrote the original draft and edited the work. E.C. performed conceptualization and reviewed and edited the work. E.T. performed conceptualization, investigation, data curation, review and editing, and project administration and provided supervision. P.R. performed conceptualization and review and editing and provided supervision.

DECLARATION OF INTERESTS

P.R. is a co-founder, part-time employee, and equity holder of a2z Radiology AI.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) used ChatGPT o4-mini and Grok 3 in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

REFERENCES

1. Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM *9*, 36–45. https://doi.org/10.1145/365153.365168.
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2023). Attention is all you need. Preprint at arXiv. https://doi.org/10.48550/arXiv.1706.03762.
3. Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. (2020). Scaling laws for neural language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2001.08361.
4. Liu, S., McCoy, A.B., Wright, A.P., Carew, B., Genkins, J.Z., Huang, S.S., Peterson, J.F., Steitz, B., and Wright, A. (2024). Leveraging large language models for generating responses to patient messages—a subjective analysis. J. Am. Med. Inform. Assoc. *31*, 1367–1379. https://doi.org/10.1093/jamia/ocae052.
5. Bernstein, I.A., Zhang, Y.V., Govil, D., Majid, I., Chang, R.T., Sun, Y., Shue, A., Chou, J.C., Schehlein, E., Christopher, K.L., et al. (2023). Comparison of Ophthalmologist and Large Language Model Chatbot Responses to Online Patient Eye Care Questions. JAMA Netw. Open *6*, e2330320. https://doi.org/10.1001/jamanetworkopen.2023.30320.
6. Mika, A.P., Martin, J.R., Engstrom, S.M., Polkowski, G.G., and Wilson, J. M. (2023). Assessing ChatGPT Responses to Common Patient Questions Regarding Total Hip Arthroplasty. J. Bone Joint Surg. Am. *105*, 1519–1526. https://doi.org/10.2106/JBJS.23.00209.
7. Steimetz, E., Minkowitz, J., Gabutan, E.C., Ngichabe, J., Attia, H., Hershkop, M., Ozay, F., Hanna, M.G., and Gupta, R. (2024). Use of Artificial Intelligence Chatbots in Interpretation of Pathology Reports. JAMA Netw. Open *7*, e2412767. https://doi.org/10.1001/jamanetworkopen.2024.12767.
8. Zaretsky, J., Kim, J.M., Baskharoun, S., Zhao, Y., Austrian, J., Aphinyanaphongs, Y., Gupta, R., Blecker, S.B., and Feldman, J. (2024). Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format. JAMA Netw. Open *7*, e240357. https://doi.org/10.1001/jamanetworkopen.2024.0357.
9. Habicht, J., Viswanathan, S., Carrington, B., Hauser, T.U., Harper, R., and Rollwage, M. (2024). Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. Nat. Med. *30*, 595–602. https://doi.org/10.1038/s41591-023-02766-x.
10. Holohan, M. (2023). A boy saw 17 doctors over 3 years for chronic pain. ChatGPT found the diagnosis. https://www.today.com/health/mom-chatgpt-diagnosis-pain-rcna101843.

11. Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.-B., Aali, A., Bluethgen, C., Pareek, A., Polacin, M., Reis, E.P., Seehofnerová, A., et al. (2024). Adapted large language models can outperform medical experts in clinical text summarization. Nat. Med. *30*, 1134–1142. https://doi.org/10.1038/s41591-024-02855-5.

12. Tu, T., Schaekermann, M., Palepu, A., Saab, K., Freyberg, J., Tanno, R., Wang, A., Li, B., Amin, M., Cheng, Y., et al. (2025). Towards conversational diagnostic artificial intelligence. Nature. https://doi.org/10.1038/s41586-025-08866-7.

13. Johri, S., Jeong, J., Tran, B.A., Schlessinger, D.I., Wongvibulsin, S., Cai, Z.R., Daneshjou, R., and Rajpurkar, P. (2023). Guidelines For Rigorous Evaluation of Clinical LLMs For Conversational Reasoning. Preprint at medRxiv. https://doi.org/10.1101/2023.09.12.23295399.

14. Li, J., Guan, Z., Wang, J., Cheung, C.Y., Zheng, Y., Lim, L.-L., Lim, C.C., Ruamviboonsuk, P., Raman, R., Corsino, L., et al. (2024). Integrated image-based deep learning and language models for primary diabetes care. Nat. Med. *30*, 2886–2896. https://doi.org/10.1038/s41591-024-03139-8.

15. Huang, A.S., Hirabayashi, K., Barna, L., Parikh, D., and Pasquale, L.R. (2024). Assessment of a Large Language Model's Responses to Questions and Cases About Glaucoma and Retina Management. JAMA Ophthalmol. *142*, 371–375. https://doi.org/10.1001/jamaophthalmol.2023.6917.

16. Ferber, D., Wiest, I.C., Wölflein, G., Ebert, M.P., Beutel, G., Eckardt, J.-N., Truhn, D., Springfeld, C., Jäger, D., and Kather, J.N. (2024). GPT-4 for Information Retrieval and Comparison of Medical Oncology Guidelines. NEJM AI *1*, AIcs2300235. https://doi.org/10.1056/AIcs2300235.

17. Koch, M.-C. (2025). Clinical co-pilot receives first approval for Class IIb medical device. Heise. https://www.heise.de/en/news/Clinical-co-pilot-receives-first-approval-for-Class-IIb-medical-device-10348301.html.

18. Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., et al. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit. Health *2*, e0000198. https://doi.org/10.1371/journal.pdig.0000198.

19. Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., et al. (2023). Large language models encode clinical knowledge. Nature *620*, 172–180. https://doi.org/10.1038/s41586-023-06291-2.

20. Lee, P., Bubeck, S., and Petro, J. (2023). Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. N. Engl. J. Med. *388*, 1233–1239. https://doi.org/10.1056/NEJMsr2214184.

21. Saab, K., Tu, T., Weng, W.-H., Tanno, R., Stutz, D., Wulczyn, E., Zhang, F., Strother, T., Park, C., Vedadi, E., et al. (2024). Capabilities of Gemini models in medicine. Preprint at arXiv. https://doi.org/10.48550/arXiv.2404.18416.

22. Liu, S., McCoy, A.B., Wright, A.P., Carew, B., Genkins, J.Z., Huang, S.S., Peterson, J.F., Steitz, B., and Wright, A. (2023). Leveraging Large Language Models for Generating Responses to Patient Messages. Preprint at medRxiv, 2023.07.14.23292669. https://doi.org/10.1101/2023.07.14.23292669.

23. Tierney, A.A., Gayre, G., Hoberman, B., Mattern, B., Ballesca, M., Kipnis, P., Liu, V., and Lee, K. (2024). Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation. NEJM Catal. *5*. https://doi.org/10.1056/CAT.23.0404.

24. Omiye, J.A., Gui, H., Rezaei, S.J., Zou, J., and Daneshjou, R. (2024). Large Language Models in Medicine: The Potentials and Pitfalls: A Narrative Review. Ann. Intern. Med. *177*, 210–220. https://doi.org/10.7326/M23-2772.

25. Grewal, H., Dhillon, G., Monga, V., Sharma, P., Buddhavarapu, V.S., Sidhu, G., and Kashyap, R. (2023). Radiology Gets Chatty: The ChatGPT Saga Unfolds. Cureus *15*, e40135. https://doi.org/10.7759/cureus.40135.

26. Qiu, J., Lam, K., Li, G., Acharya, A., Wong, T.Y., Darzi, A., Yuan, W., and Topol, E.J. (2024). LLM-based agentic systems in medicine and healthcare. Nat. Mach. Intell. *6*, 1418–1420. https://doi.org/10.1038/s42256-024-00944-1.

27. Mukherjee, S., Gamble, P., Ausin, M.S., Kant, N., Aggarwal, K., Manjunath, N., Datta, D., Liu, Z., Ding, J., Busacca, S., et al. (2024). Polaris: A safety-focused LLM constellation architecture for healthcare. Preprint at arXiv. https://doi.org/10.48550/arXiv.2403.13313.

28. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2201.11903.

29. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. (2023). ReAct: synergizing reasoning and acting in language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2210.03629.

30. Xu, F., Hao, Q., Zong, Z., Wang, J., Zhang, Y., Wang, J., Lan, X., Gong, J., Ouyang, T., Meng, F., et al. (2025). Towards large reasoning models: A survey of reinforced reasoning with large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2501.09686.

31. Savage, T., Nayak, A., Gallo, R., Rangan, E., and Chen, J.H. (2024). Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. npj Digit. Med. *7*, 20. https://doi.org/10.1038/s41746-024-01010-1.

32. Wu, J., Deng, W., Li, X., Liu, S., Mi, T., Peng, Y., Xu, Z., Liu, Y., Cho, H., Choi, C.-I., et al. (2025). MedReason: eliciting factual medical reasoning steps in LLMs via knowledge graphs. Preprint at arXiv. https://doi.org/10.48550/arXiv.2504.00993.

33. Moor, M., Huang, Q., Wu, S., Yasunaga, M., Zakka, C., Dalmia, Y., Reis, E.P., Rajpurkar, P., and Leskovec, J. (2023). Med-flamingo: a multimodal medical few-shot learner. Preprint at arXiv. https://doi.org/10.48550/arXiv.2307.15189.

34. Lu, M.Y., Chen, B., Williamson, D.F.K., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al. (2024). A visual-language foundation model for computational pathology. Nat. Med. *30*, 863–874. https://doi.org/10.1038/s41591-024-02856-4.

35. Acosta, J.N., Falcone, G.J., Rajpurkar, P., and Topol, E.J. (2022). Multimodal biomedical AI. Nat. Med. *28*, 1773–1784. https://doi.org/10.1038/s41591-022-01981-2.

36. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models From natural language supervision. Preprint at arXiv. https://doi.org/10.48550/arXiv.2103.00020.

37. Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., Arx, S. von, Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2022). On the opportunities and risks of foundation models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2108.07258.

38. Zhang, K., Zhou, R., Adhikarla, E., Yan, Z., Liu, Y., Yu, J., Liu, Z., Chen, X., Davison, B.D., Ren, H., et al. (2024). A generalist vision–language foundation model for diverse biomedical tasks. Nat. Med. *30*, 3129–3141. https://doi.org/10.1038/s41591-024-03185-2.

39. Zhou, H.-Y., Yu, Y., Wang, C., Zhang, S., Gao, Y., Pan, J., Shao, J., Lu, G., Zhang, K., and Li, W. (2023). A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. Nat. Biomed. Eng. *7*, 743–755. https://doi.org/10.1038/s41551-023-01045-x.

40. Khader, F., Müller-Franzes, G., Wang, T., Han, T., Tayebi Arasteh, S., Haarburger, C., Stegmaier, J., Bressem, K., Kuhl, C., Nebelung, S., et al. (2023). Multimodal Deep Learning for Integrating Chest Radiographs and Clinical Parameters: A Case for Transformers. Radiology *309*, e230806. https://doi.org/10.1148/radiol.230806.

41. Chen, R.J., Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B., et al. (2022). Pan-cancer

integrative histology-genomic analysis via multimodal deep learning. Cancer Cell *40*, 865–878.e6. https://doi.org/10.1016/j.ccell.2022.07.004.

42. Minoura, K., Abe, K., Nam, H., Nishikawa, H., and Shimamura, T. (2021). A mixture-of-experts deep generative model for integrated analysis of single-cell multiomics data. Cell Rep. Methods *1*, 100071. https://doi.org/10.1016/j.crmeth.2021.100071.

43. Vanguri, R.S., Luo, J., Aukerman, A.T., Egger, J.V., Fong, C.J., Horvat, N., Pagano, A., Araujo-Filho, J.A.B., Geneslaw, L., Rizvi, H., et al. (2022). Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. Nat. Cancer *3*, 1151–1164. https://doi.org/10.1038/s43018-022-00416-8.

44. Rao, V.M., Hla, M., Moor, M., Adithan, S., Kwak, S., Topol, E.J., and Rajpurkar, P. (2025). Multimodal generative AI for medical image interpretation. Nature *639*, 888–896. https://doi.org/10.1038/s41586-025-08675-y.

45. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F.K., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al. (2024). Towards a general-purpose foundation model for computational pathology. Nat. Med. *30*, 850–862. https://doi.org/10.1038/s41591-024-02857-3.

46. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., and Zou, J. (2023). A visual–language foundation model for pathology image analysis using medical Twitter. Nat. Med. *29*, 2307–2316. https://doi.org/10.1038/s41591-023-02504-3.

47. Lu, M.Y., Chen, B., Williamson, D.F.K., Chen, R.J., Zhao, M., Chow, A.K., Ikemura, K., Kim, A., Pouli, D., Patel, A., et al. (2024). A multimodal generative AI copilot for human pathology. Nature *634*, 466–473. https://doi.org/10.1038/s41586-024-07618-3.

48. Gadaleta, M., Harrington, P., Barnhill, E., Hytopoulos, E., Turakhia, M.P., Steinhubl, S.R., and Quer, G. (2023). Prediction of atrial fibrillation from at-home single-lead ECG signals without arrhythmias. npj Digit. Med. *6*, 229. https://doi.org/10.1038/s41746-023-00966-w.

49. Gavidia, M., Zhu, H., Montanari, A.N., Fuentes, J., Cheng, C., Dubner, S., Chames, M., Maison-Blanche, P., Rahman, M.M., Sassi, R., et al. (2024). Early warning of atrial fibrillation using deep learning. Patterns (N Y) *5*, 100970. https://doi.org/10.1016/j.patter.2024.100970.

50. Attia, Z.I., Harmon, D.M., Dugan, J., Manka, L., Lopez-Jimenez, F., Lerman, A., Siontis, K.C., Noseworthy, P.A., Yao, X., Klavetter, E.W., et al. (2022). Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. Nat. Med. *28*, 2497–2503. https://doi.org/10.1038/s41591-022-02053-1.

51. Guan, G., Mofaz, M., Qian, G., Patalon, T., Shmueli, E., Yamin, D., and Brandeau, M.L. (2022). Higher sensitivity monitoring of reactions to COVID-19 vaccination using smartwatches. npj Digit. Med. *5*, 140. https://doi.org/10.1038/s41746-022-00683-w.

52. Madhvapathy, S.R., Wang, J.-J., Wang, H., Patel, M., Chang, A., Zheng, X., Huang, Y., Zhang, Z.J., Gallon, L., and Rogers, J.A. (2023). Implantable bioelectronic systems for early detection of kidney transplant rejection. Science *381*, 1105–1112. https://doi.org/10.1126/science.adh7726.

53. Torrente-Rodríguez, R.M., Tu, J., Yang, Y., Min, J., Wang, M., Song, Y., Yu, Y., Xu, C., Ye, C., IsHak, W.W., et al. (2020). Investigation of Cortisol Dynamics in Human Sweat Using a Graphene-Based Wireless mHealth System. Matter *2*, 921–937. https://doi.org/10.1016/j.matt.2020.01.021.

54. Lin, M., Zhang, Z., Gao, X., Bian, Y., Wu, R.S., Park, G., Lou, Z., Zhang, Z., Xu, X., Chen, X., et al. (2024). A fully integrated wearable ultrasound system to monitor deep tissues in moving subjects. Nat. Biotechnol. *42*, 448–457. https://doi.org/10.1038/s41587-023-01800-0.

55. Menzies, S.W., Sinz, C., Menzies, M., Lo, S.N., Yolland, W., Lingohr, J., Razmara, M., Tschandl, P., Guitera, P., Scolyer, R.A., et al. (2023). Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. Lancet Digit. Health *5*, e679–e691. https://doi.org/10.1016/S2589-7500(23)00130-9.

56. Shaikh, N., Conway, S.J., Kovačević, J., Condessa, F., Shope, T.R., Haralam, M.A., Campese, C., Lee, M.C., Larsson, T., Cavdar, Z., et al. (2024). Development and Validation of an Automated Classifier to Diagnose Acute Otitis Media in Children. JAMA Pediatr. *178*, 401–407. https://doi.org/10.1001/jamapediatrics.2024.0011.

57. Landy, R., Wang, V.L., Baldwin, D.R., Pinsky, P.F., Cheung, L.C., Castle, P.E., Skarzynski, M., Robbins, H.A., and Katki, H.A. (2023). Recalibration of a Deep Learning Model for Low-Dose Computed Tomographic Images to Inform Lung Cancer Screening Intervals. JAMA Netw. Open *6*, e233273. https://doi.org/10.1001/jamanetworkopen.2023.3273.

58. Lång, K., Josefsson, V., Larsson, A.-M., Larsson, S., Högberg, C., Sartor, H., Hofvind, S., Andersson, I., and Rosso, A. (2023). Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol. *24*, 936–944. https://doi.org/10.1016/S1470-2045(23)00298-X.

59. Ng, A.Y., Oberije, C.J.G., Ambrózay, É., Szabó, E., Serfőző, O., Karpati, E., Fox, G., Glocker, B., Morris, E.A., Forrai, G., et al. (2023). Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. Nat. Med. *29*, 3044–3049. https://doi.org/10.1038/s41591-023-02625-9.

60. Lauritzen, A.D., Von Euler-Chelpin, M.C., Lynge, E., Vejborg, I., Nielsen, M., Karssemeijer, N., and Lillholm, M. (2023). Assessing Breast Cancer Risk by Combining AI for Lesion Detection and Mammographic Texture. Radiology *308*, e230227. https://doi.org/10.1148/radiol.230227.

61. Salim, M., Liu, Y., Sorkhei, M., Ntoula, D., Foukakis, T., Fredriksson, I., Wang, Y., Eklund, M., Azizpour, H., Smith, K., et al. (2024). AI-based selection of individuals for supplemental MRI in population-based breast cancer screening: the randomized ScreenTrustMRI trial. Nat. Med. *30*, 2623–2630. https://doi.org/10.1038/s41591-024-03093-5.

62. Kadirvelu, B., Gavriel, C., Nageshwaran, S., Chan, J.P.K., Nethisinghe, S., Athanasopoulos, S., Ricotti, V., Voit, T., Giunti, P., Festenstein, R., et al. (2023). A wearable motion capture suit and machine learning predict disease progression in Friedreich's ataxia. Nat. Med. *29*, 86–94. https://doi.org/10.1038/s41591-022-02159-6.

63. Mekkes, N.J., Groot, M., Hoekstra, E., De Boer, A., Dagkesamanskaia, E., Bouwman, S., Wehrens, S.M.T., Herbert, M.K., Wever, D.D., Rozemuller, A., et al. (2024). Identification of clinical disease trajectories in neurodegenerative disorders with natural language processing. Nat. Med. *30*, 1143–1153. https://doi.org/10.1038/s41591-024-02843-9.

64. Dingemans, A.J.M., Hinne, M., Truijen, K.M.G., Goltstein, L., Van Reeuwijk, J., De Leeuw, N., Schuurs-Hoeijmakers, J., Pfundt, R., Diets, I.J., Den Hoed, J., et al. (2023). PhenoScore quantifies phenotypic variation for rare genetic diseases by combining facial analysis with other clinical features using a machine-learning framework. Nat. Genet. *55*, 1598–1607. https://doi.org/10.1038/s41588-023-01469-w.

65. Diaz-Pinto, A., Ravikumar, N., Attar, R., Suinesiaputra, A., Zhao, Y., Levelt, E., Dall'armellina, E., Lorenzi, M., Chen, Q., Keenan, T.D.L., et al. (2022). Predicting myocardial infarction through retinal scans and minimal personal information. Nat. Mach. Intell. *4*, 55–61. https://doi.org/10.1038/s42256-021-00427-7.

66. Topol, E.J. (2024). Medical forecasting. Science *384*, eadp7977. https://doi.org/10.1126/science.adp7977.

67. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

68. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876. https://doi.org/10.1126/science.abj8754.

69. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

70. Yang, Z., Zeng, X., Zhao, Y., and Chen, R. (2023). AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduct. Target. Ther. *8*, 115. https://doi.org/10.1038/s41392-023-01381-z.

71. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061.

72. Topol, E. (2024). Learning the Language of Life with A.I.: A Hyper-Accelerated Phase of New Foundation Models. Ground Truths. https://erictopol.substack.com/p/learning-the-language-of-life-with.

73. Nguyen, E., Poli, M., Durrant, M.G., Kang, B., Katrekar, D., Li, D.B., Bartie, L.J., Thomas, A.W., King, S.H., Brixi, G., et al. (2024). Sequence modeling and design from molecular to genome scale with Evo. Science *386*, eado9336. https://doi.org/10.1126/science.ado9336.

74. Rood, J.E., Wynne, S., Robson, L., Hupalowska, A., Randell, J., Teichmann, S.A., and Regev, A. (2025). The Human Cell Atlas from a cell census to a unified foundation model. Nature *637*, 1065–1071. https://doi.org/10.1038/s41586-024-08338-4.

75. Xiong, D., Qiu, Y., Zhao, J., Zhou, Y., Lee, D., Gupta, S., Torres, M., Lu, W., Liang, S., Kang, J.J., et al. (2024). A structurally informed human protein–protein interactome reveals proteome-wide perturbations caused by disease mutations. Nat. Biotechnol., 1–15. https://doi.org/10.1038/s41587-024-02428-4.

76. Karelina, M., Noh, J.J., and Dror, R.O. (2023). How accurately can one predict drug binding modes using AlphaFold models? eLife *12*, RP89386. https://doi.org/10.7554/eLife.89386.

77. Drysdale, E. (2023). A multitask neural network trained on embeddings from ESMFold can accurately rank order clinical outcomes for different cystic fibrosis mutations. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.26.564274.

78. Zhang, H., Lan, J., Wang, H., Lu, R., Zhang, N., He, X., Yang, J., and Chen, L. (2024). AlphaFold2 in biomedical research: facilitating the development of diagnostic strategies for disease. Front. Mol. Biosci. *11*, 1414916. https://doi.org/10.3389/fmolb.2024.1414916.

79. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., et al. (2023). De novo design of protein structure and function with RFdiffusion. Nature *620*, 1089–1100. https://doi.org/10.1038/s41586-023-06415-8.

80. Yim, J., Trippe, B.L., Bortoli, V.D., M, E., Doucet, A., Barzilay, R., and Jaakkola, T. (2023). SE(3) diffusion model with application to protein backbone generation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2302.02277.

81. Brixi, G., Durrant, M.G., Ku, J., Poli, M., Brockman, G., Chang, D., Gonzalez, G.A., King, S.H., Li, D.B., Merchant, A.T., et al. (2025). Genome modeling and design across all domains of life with Evo 2. Preprint at bioRxiv. https://doi.org/10.1101/2025.02.18.638918.

82. Ni, B., Kaplan, D.L., and Buehler, M.J. (2023). Generative design of de novo proteins based on secondary structure constraints using an attention-based diffusion model. Chem *9*, 1828–1849. https://doi.org/10.1016/j.chempr.2023.03.020.

83. Lutz, I.D., Wang, S., Norn, C., Courbet, A., Borst, A.J., Zhao, Y.T., Dosey, A., Cao, L., Xu, J., Leaf, E.M., et al. (2023). Top-down design of protein architectures with reinforcement learning. Science *380*, 266–273. https://doi.org/10.1126/science.adf6591.

84. Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M., Olmos, J.L., Xiong, C., Sun, Z.Z., Socher, R., et al. (2023). Large language models generate functional protein sequences across diverse families. Nat. Biotechnol. *41*, 1099–1106. https://doi.org/10.1038/s41587-022-01618-2.

85. Altman, R.B. (2023). A Holy Grail — The Prediction of Protein Structure. N. Engl. J. Med. *389*, 1431–1434. https://doi.org/10.1056/NEJMcibr2307735.

86. Anishchenko, I., Pellock, S.J., Chidyausiku, T.M., Ramelot, T.A., Ovchinnikov, S., Hao, J., Bafna, K., Norn, C., Kang, A., Bera, A.K., et al. (2021). De novo protein design by deep network hallucination. Nature *600*, 547–552. https://doi.org/10.1038/s41586-021-04184-w.

87. Lin, C.-S., Liu, W.-T., Tsai, D.-J., Lou, Y.-S., Chang, C.-H., Lee, C.-C., Fang, W.-H., Wang, C.-C., Chen, Y.-Y., Lin, W.-S., et al. (2024). AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat. Med. *30*, 1461–1470. https://doi.org/10.1038/s41591-024-02961-4.

88. Al-Zaiti, S.S., Martin-Gill, C., Zègre-Hemsey, J.K., Bouzid, Z., Faramand, Z., Alrawashdeh, M.O., Gregg, R.E., Helman, S., Riek, N.T., Kraevsky-Phillips, K., et al. (2023). Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. Nat. Med. *29*, 1804–1813. https://doi.org/10.1038/s41591-023-02396-3.

89. Sundrani, S., Chen, J., Jin, B.T., Abad, Z.S.H., Rajpurkar, P., and Kim, D. (2023). Predicting patient decompensation from continuous physiologic monitoring in the emergency department. npj Digit. Med. *6*, 60. https://doi.org/10.1038/s41746-023-00803-0.

90. Agrawal, S., Klarqvist, M.D.R., Emdin, C., Patel, A.P., Paranjpe, M.D., Ellinor, P.T., Philippakis, A., Ng, K., Batra, P., and Khera, A.V. (2021). Selection of 51 predictors from 13,782 candidate multimodal features using machine learning improves coronary artery disease prediction. Patterns (N Y) *2*, 100364. https://doi.org/10.1016/j.patter.2021.100364.

91. Williams, S.A., Ostroff, R., Hinterberg, M.A., Coresh, J., Ballantyne, C.M., Matsushita, K., Mueller, C.E., Walter, J., Jonasson, C., Holman, R.R., et al. (2022). A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk. Sci. Transl. Med. *14*, eabj9625. https://doi.org/10.1126/scitranslmed.abj9625.

92. Khurshid, S., Friedman, S., Reeder, C., Di Achille, P., Diamant, N., Singh, P., Harrington, L.X., Wang, X., Al-Alusi, M.A., Sarma, G., et al. (2022). ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. Circulation *145*, 122–133. https://doi.org/10.1161/CIRCULATIONAHA.121.057480.

93. Jeon, K.-H., Lee, H.S., Kang, S., Jang, J.-H., Jo, Y.-Y., Son, J.M., Lee, M.S., Kwon, J.-M., Kwun, J.-S., Cho, H.-W., et al. (2024). AI-enabled ECG index for predicting left ventricular dysfunction in patients with ST-segment elevation myocardial infarction. Sci. Rep. *14*, 16575. https://doi.org/10.1038/s41598-024-67532-6.

94. Christensen, M., Vukadinovic, M., Yuan, N., and Ouyang, D. (2024). Vision–language foundation model for echocardiogram interpretation. Nat. Med. *30*, 1481–1488. https://doi.org/10.1038/s41591-024-02959-y.

95. Popescu, D.M., Shade, J.K., Lai, C., Aronis, K.N., Ouyang, D., Moorthy, M.V., Cook, N.R., Lee, D.C., Kadish, A., Albert, C.M., et al. (2022). Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart. Nat CardioVasc Res *1*, 334–343. https://doi.org/10.1038/s44161-022-00041-9.

96. Lin, A., Manral, N., McElhinney, P., Killekar, A., Matsumoto, H., Kwiecinski, J., Pieszko, K., Razipour, A., Grodecki, K., Park, C., et al. (2022). Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. Lancet Digit. Health *4*, e256–e265. https://doi.org/10.1016/S2589-7500(22)00022-X.

97. Chan, K., Wahome, E., Tsiachristas, A., Antonopoulos, A.S., Patel, P., Lyasheva, M., Kingham, L., West, H., Oikonomou, E.K., Volpe, L., et al. (2024). Inflammatory risk and cardiovascular events in patients without obstructive coronary artery disease: the ORFAN multicentre, longitudinal cohort study. Lancet *403*, 2606–2618. https://doi.org/10.1016/S0140-6736(24)00596-8.

98. Zhang, M., Wong, S.W., Wright, J.N., Wagner, M.W., Toescu, S., Han, M., Tam, L.T., Zhou, Q., Ahmadian, S.S., Shpanskaya, K., et al. (2022). MRI Radiogenomics of Pediatric Medulloblastoma: A Multicenter Study. Radiology *304*, 406–416. https://doi.org/10.1148/radiol.212137.

99. Rengo, M., Onori, A., Caruso, D., Bellini, D., Carbonetti, F., De Santis, D., Vicini, S., Zerunian, M., Iannicelli, E., Carbone, I., et al. (2023). Development and Validation of Artificial-Intelligence-Based Radiomics Model Using Computed Tomography Features for Preoperative Risk Stratification of Gastrointestinal Stromal Tumors. J. Pers. Med. *13*, 717. https://doi.org/10.3390/jpm13050717.

100. Li, L., Zhou, X., Cui, W., Li, Y., Liu, T., Yuan, G., Peng, Y., and Zheng, J. (2023). Combining radiomics and deep learning features of intra-tumoral and peri-tumoral regions for the classification of breast cancer lung metastasis and primary lung cancer with low-dose CT. J. Cancer Res. Clin. Oncol. *149*, 15469–15478. https://doi.org/10.1007/s00432-023-05329-2.

101. Granata, V., Fusco, R., Setola, S.V., Galdiero, R., Maggialetti, N., Silvestro, L., De Bellis, M., Di Girolamo, E., Grazzini, G., Chiti, G., et al. (2023). Risk Assessment and Pancreatic Cancer: Diagnostic Management and Artificial Intelligence. Cancers *15*, 351. https://doi.org/10.3390/cancers15020351.

102. Chiti, G., Grazzini, G., Flammia, F., Matteuzzi, B., Tortoli, P., Bettarini, S., Pasqualini, E., Granata, V., Busoni, S., Messserini, L., et al. (2022). Gastroenteropancreatic neuroendocrine neoplasms (GEP-NENs): a radiomic model to predict tumor grade. Radiol. Med. *127*, 928–938. https://doi.org/10.1007/s11547-022-01529-x.

103. Swanson, K., Wu, E., Zhang, A., Alizadeh, A.A., and Zou, J. (2023). From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell *186*, 1772–1791. https://doi.org/10.1016/j.cell.2023.01.035.

104. Nimgaonkar, V., Krishna, V., Krishna, V., Tiu, E., Joshi, A., Vrabac, D., Bhambhvani, H., Smith, K., Johansen, J.S., Makawita, S., et al. (2023). Development of an artificial intelligence-derived histologic signature associated with adjuvant gemcitabine treatment outcomes in pancreatic cancer. Cell Rep. Med. *4*, 101013. https://doi.org/10.1016/j.xcrm.2023.101013.

105. Sorin, M., Rezanejad, M., Karimi, E., Fiset, B., Desharnais, L., Perus, L.J. M., Milette, S., Yu, M.W., Maritan, S.M., Doré, S., et al. (2023). Single-cell spatial landscapes of the lung tumour immune microenvironment. Nature *614*, 548–554. https://doi.org/10.1038/s41586-022-05672-3.

106. Tian, F., Liu, D., Wei, N., Fu, Q., Sun, L., Liu, W., Sui, X., Tian, K., Nemeth, G., Feng, J., et al. (2024). Prediction of tumor origin in cancers of unknown primary origin with cytology-based deep learning. Nat. Med. *30*, 1309–1319. https://doi.org/10.1038/s41591-024-02915-w.

107. Placido, D., Yuan, B., Hjaltelin, J.X., Zheng, C., Haue, A.D., Chmura, P.J., Yuan, C., Kim, J., Umeton, R., Antell, G., et al. (2023). A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. Nat. Med. *29*, 1113–1122. https://doi.org/10.1038/s41591-023-02332-5.

108. Smits, F.J., Henry, A.C., Besselink, M.G., Busch, O.R., Van Eijck, C.H., Arntz, M., Bollen, T.L., Van Delden, O.M., Van Den Heuvel, D., Van Der Leij, C., et al. (2022). Algorithm-based care versus usual care for the early recognition and management of complications after pancreatic resection in the Netherlands: an open-label, nationwide, stepped-wedge cluster-randomised trial. Lancet *399*, 1867–1875. https://doi.org/10.1016/S0140-6736(22)00182-9.

109. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., et al. (2023). A foundation model for generalizable disease detection from retinal images. Nature *622*, 156–163. https://doi.org/10.1038/s41586-023-06555-x.

110. Haque, A., Milstein, A., and Fei-Fei, L. (2020). Illuminating the dark spaces of healthcare with ambient intelligence. Nature *585*, 193–202. https://doi.org/10.1038/s41586-020-2669-y.

111. Liu, Y., Zhang, G., Tarolli, C.G., Hristov, R., Jensen-Roberts, S., Waddell, E.M., Myers, T.L., Pawlik, M.E., Soto, J.M., Wilson, R.M., et al. (2022). Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response. Sci. Transl. Med. *14*, eadc9669. https://doi.org/10.1126/scitranslmed.adc9669.

112. Ricotti, V., Kadirvelu, B., Selby, V., Festenstein, R., Mercuri, E., Voit, T., and Faisal, A.A. (2023). Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. Nat. Med. *29*, 95–103. https://doi.org/10.1038/s41591-022-02045-1.

113. Jiang, L.Y., Liu, X.C., Nejatian, N.P., Nasir-Moin, M., Wang, D., Abidin, A., Eaton, K., Riina, H.A., Laufer, I., Punjabi, P., et al. (2023). Health system-scale language models are all-purpose prediction engines. Nature *619*, 357–362. https://doi.org/10.1038/s41586-023-06160-y.

114. Yang, Z., Mitra, A., Liu, W., Berlowitz, D., and Yu, H. (2023). TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records. Nat. Commun. *14*, 7857. https://doi.org/10.1038/s41467-023-43715-z.

115. Beaulieu-Jones, B.K., Villamar, M.F., Scordis, P., Bartmann, A.P., Ali, W., Wissel, B.D., Alsentzer, E., De Jong, J., Patra, A., and Kohane, I. (2023). Predicting seizure recurrence after an initial seizure-like episode from routine clinical notes using large language models: a retrospective cohort study. Lancet Digit. Health *5*, e882–e894. https://doi.org/10.1016/S2589-7500(23)00179-6.

116. Moguilner, S., Baez, S., Hernandez, H., Migeot, J., Legaz, A., Gonzalez-Gomez, R., Farina, F.R., Prado, P., Cuadros, J., Tagliazucchi, E., et al. (2024). Brain clocks capture diversity and disparities in aging and dementia across geographically diverse populations. Nat. Med. *30*, 3646–3657. https://doi.org/10.1038/s41591-024-03209-x.

117. Ogata, S., Takegami, M., Ozaki, T., Nakashima, T., Onozuka, D., Murata, S., Nakaoku, Y., Suzuki, K., Hagihara, A., Noguchi, T., et al. (2021). Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. Nat. Commun. *12*, 4575. https://doi.org/10.1038/s41467-021-24823-0.

118. Zakka, C., Shad, R., Chaurasia, A., Dalal, A.R., Kim, J.L., Moor, M., Fong, R., Phillips, C., Alexander, K., Ashley, E., et al. (2024). Almanac — Retrieval-Augmented Language Models for Clinical Medicine. NEJM AI *1*, AIoa2300068. https://doi.org/10.1056/AIoa2300068.

119. Menz, B.D., Kuderer, N.M., Bacchi, S., Modi, N.D., Chin-Yee, B., Hu, T., Rickard, C., Haseloff, M., Vitry, A., McKinnon, R.A., et al. (2024). Current safeguards, risk mitigation, and transparency measures of large language models against the generation of health disinformation: repeated cross sectional analysis. BMJ *384*, e078538. https://doi.org/10.1136/bmj-2023-078538.

120. (2024). How to support the transition to AI-powered healthcare. Nat. Med. *30*, 609–610. https://doi.org/10.1038/s41591-024-02897-9.

121. Han, T., Kumar, A., Agarwal, C., and Lakkaraju, H. (2024). MedSafetyBench: evaluating and improving the medical safety of large language models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2403.03744.

122. Shah, N.H., Entwistle, D., and Pfeffer, M.A. (2023). Creation and Adoption of Large Language Models in Medicine. JAMA *330*, 866–869. https://doi.org/10.1001/jama.2023.14217.

123. Sharma, A., Lin, I.W., Miner, A.S., Atkins, D.C., and Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat. Mach. Intell. *5*, 46–57. https://doi.org/10.1038/s42256-022-00593-2.

124. Yu, F., Moehring, A., Banerjee, O., Salz, T., Agarwal, N., and Rajpurkar, P. (2024). Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat. Med. *30*, 837–849. https://doi.org/10.1038/s41591-024-02850-w.

125. Blumenthal, D., and Patel, B. (2024). The Regulation of Clinical Artificial Intelligence. NEJM AI *1*, AIpc2400545. https://doi.org/10.1056/AIpc2400545.

126. Capoot, A. (2024). Dexcom's over-the-counter glucose monitor now offers users an AI summary of how sleep, meals and more impact sugar levels. CNBC. https://www.cnbc.com/2024/12/17/dexcom-launches-generative-ai-platform-for-stelo-users.html.

127. Rajpurkar, P., and Topol, E.J. (2025). A clinical certification pathway for generalist medical AI systems. Lancet *405*, 20. https://doi.org/10.1016/S0140-6736(24)02797-1.

128. Kim, W. (2024). Seeing the Unseen: Advancing Generative AI Research in Radiology. Radiology *311*, e240935. https://doi.org/10.1148/radiol.240935.

129. Togher, D., Dean, G., Moon, J., Mayola, R., Medina, A., Repec, J., Meheux, M., Mather, S., Storey, M., Rickaby, S., et al. (2025). Evolution of radiology staff perspectives during artificial intelligence (AI) implementation for expedited lung cancer triage. Clin. Rad. *81*, 106704. https://doi.org/10.1016/j.crad.2024.09.010.

130. Mollick, E. (2024). Co-Intelligence: Living and Working with AI (Portfolio).

131. Lee, P., Goldberg, C., and Kohane, I. (2023). The AI Revolution in Medicine: GPT-4 and Beyond (Pearson).