nature medicine

REVIEW ARTICLE

https://doi.org/10.1038/s41591-021-01614-0

AI in health and medicine

Pranav Rajpurkar[1,4], Emma Chen[2,4], Oishi Banerjee[2,4] and Eric J. Topol[3]✉

Artificial intelligence (AI) is poised to broadly reshape medicine, potentially improving the experiences of both clinicians and patients. We discuss key findings from a 2-year weekly effort to track and share key developments in medical AI. We cover prospective studies and advances in medical image analysis, which have reduced the gap between research and deployment. We also address several promising avenues for novel medical AI research, including non-image data sources, unconventional problem formulations and human–AI collaboration. Finally, we consider serious technical and ethical challenges in issues spanning from data scarcity to racial bias. As these challenges are addressed, AI's potential may be realized, making healthcare more accurate, efficient and accessible for patients worldwide.

In the years ahead, AI is poised to broadly reshape medicine. Just a few years since the first landmark demonstrations of medical AI algorithms that are able to detect disease from medical images at the level of experts[1–4], the landscape of medical AI has matured considerably. Today, the deployment of medical AI systems in routine clinical care presents an important yet largely unfulfilled opportunity, as the medical AI community navigates the complex ethical, technical and human-centered challenges required for safe and effective translation.

In this review, we summarize major advances and highlight overarching trends, providing a concise overview of the state of medical AI. Our review is informed by our efforts over the past 2 years, during which we tracked and shared recent developments in medical AI on a weekly basis (https://doctorpenguin.com). First, we summarize recent progress, highlighting studies that have rigorously demonstrated the utility of medical AI systems. Second, we examine promising avenues for medical AI research in the form of novel data sources and discuss collaboration setups between AI and humans, which are more likely to reflect real medical practice than typical study designs that pit AI against humans. Finally, we discuss major challenges facing the field, including the technological limitations of AI as it stands and ethical concerns about regulating AI systems, holding people accountable when AI error occurs, respecting patient privacy and consent in data collection and safeguarding against the reinforcement of inequities (Fig. 1).

Recent progress in deployment of AI algorithms in medicine

Although AI systems have repeatedly been shown to be successful in a wide variety of retrospective medical studies, relatively few AI tools have been translated into medical practice[5]. Critics point out that AI systems may in practice be less helpful than retrospective data would suggest[6]; systems may be too slow or complicated to be useful in real medical settings[7], or unforeseen complications may arise from the way in which humans and AIs interact[8]. Moreover, retrospective in silico datasets undergo extensive filtering and cleaning, which may make them less representative of real-world medical practice. Randomized controlled trials (RCTs) and prospective studies can bridge this gap between theory and practice, more rigorously demonstrating that AI models can have a quantifiable, positive impact when deployed in real healthcare settings. Recently, RCTs have tested the usefulness of AI systems in healthcare. In addition to accuracy, a variety of other metrics have been used to assess the utility of AI, providing a holistic view of its impact on medical systems[9–13]. For example, an RCT evaluating an AI system for managing insulin doses measured the amount of time patients spent within the target glucose range[14]; a study that evaluated a monitoring system for intraoperative hypotension tracked the average duration of hypotension episodes[15], while a system that flagged cases of intracranial hemorrhage for human review was judged by its reduction of turnaround time[16]. Recent guidelines, such as AI-specific extensions to the SPIRIT and CONSORT guidelines and upcoming guidelines such as STARD-AI, may help standardize medical AI reporting, including clinical trials protocols and results, making it easier for the community to share findings and rigorously investigate the usefulness of medical AI[17,18].

In recent years, some AI tools have moved past testing to deployment, winning administrative support and clearing regulatory hurdles. The Center for Medicare and Medicaid Services, which approves public insurance reimbursement costs, has facilitated the adoption of AI in clinical settings by allowing reimbursement for the use of two specific AI systems for medical image diagnosis[19]. Furthermore, a 2020 study found that the US Food and Drug Administration (FDA) is approving AI, particularly machine learning (ML; a type of AI), products at an accelerating rate[20]. These advances largely take the form of FDA clearances, which require products to meet a lower regulatory bar than full-fledged approvals do, but they are nonetheless clearing a path for AI/ML systems to be used in real clinical settings. It is important to point out that the datasets used for these regulatory clearances are often made up of retrospective, single-institution data that are mostly unpublished and considered proprietary. To build trust in medical AI systems, stronger standards for reporting transparency and validation will be required, including demonstrations of impact on clinical outcomes.

Deep learning for interpretation of medical images. In recent years, deep learning, in which neural networks learn patterns directly from raw data, has achieved remarkable success in image classification. Medical AI research has consequently blossomed in specialties that rely heavily on the interpretation of images, such as radiology, pathology, gastroenterology and ophthalmology.

[1]Department of Biomedical Informatics, Harvard University, Cambridge, MA, USA. [2]Department of Computer Science, Stanford University, Stanford, CA, USA. [3]Scripps Translational Science Institute, San Diego, CA, USA. [4]These authors contributed equally: Pranav Rajpurkar, Emma Chen, Oishi Banerjee. ✉e-mail: etopol@scripps.edu

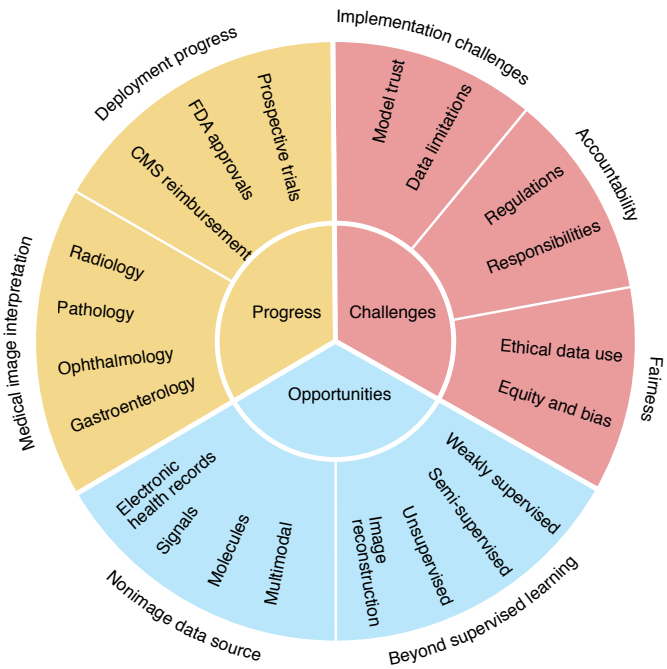

Fig. 1 | Overview of the progress, challenges and opportunities for AI in health. CMS, Centers for Medicare & Medicaid Services.

AI systems have achieved considerable improvements in accuracy for radiology tasks, including mammography interpretation[21,22], cardiac function assessment[23,24] and lung cancer screening[25], tackling not only diagnosis but also risk prediction and treatment[26]. For instance, one AI system was trained to estimate 3-year lung cancer risk from radiologists' computed tomography (CT) readings and other clinical information[27]. These predictions could then be used to schedule follow-up CT scans for patients with cancer, augmenting current screening guidelines. Validation of such systems on multiple clinical sites and an increasing number of prospective evaluations have brought AI closer to being deployed and making a practical impact in the field of radiology.

In the field of pathology, AI has made major strides in diagnosing cancers and providing new disease insights[28–33], largely through the use of whole-slide imaging. Models have been able to efficiently identify areas of interest within slides, potentially speeding up workflows for diagnosis. Beyond this practical impact, deep neural networks have been trained to discern the primary tumor origin and detect structural variants or driver mutations, providing benefits beyond even expert pathologist reviews. Furthermore, AI has been shown to make more accurate survival predictions for a wide range of cancer types compared to conventional grading and histopathological subtyping[31]. Such studies have demonstrated how AI can make pathology interpretations more efficient, accurate and useful.

Deep learning has also made progress in gastroenterology, especially in terms of improving colonoscopy, a key procedure used to detect colorectal cancer. Deep learning has been used to automatically predict whether colonic lesions are malignant, with performance comparable to skilled endoscopists[34]. Additionally, because polyps and other possible signs of disease are frequently missed during the exam[35], AI systems have been developed to assist endoscopists. Such systems have been shown to improve endoscopists' ability to detect irregularities, potentially improving sensitivity and making colonoscopy a more reliable tool for diagnosis[10,11,36].

Deep learning models have been applied widely in the area of ophthalmology, making important advances toward deployment[7,37–41]. Besides quantifying model performance, studies have investigated the human impact of such models on health systems. For example, one study examined how an AI system for eye disease screening affected patient experience and medical workflows, using human observation and interviews[7]. Other studies have looked at the financial impact of AI in the ophthalmology setting, finding that semi-automated[40] or fully automated AI screening[39] might provide cost savings in specific contexts, such as the detection of diabetic retinopathy.

Opportunities for development of AI algorithms

Medical AI studies often follow a familiar pattern, tackling an image classification problem, using supervised learning on labeled data to train an AI system and then evaluating the system by comparing it against human experts. Although such studies have achieved noteworthy advances, we present three other promising avenues of research that break from this mold (Fig. 2). First, we address non-image data sources such as text, chemical and genomic sequences, which can provide rich medical insights. Second, we discuss problem formulations that go beyond supervised learning, obtaining insights from unlabeled or otherwise imperfect data through paradigms such as unsupervised or semi-supervised learning. Finally, we look at AI systems that collaborate with humans instead of competing against them, which is a path toward achieving better performance than either AI or humans alone.

Medical data beyond images. Moving beyond image classification, deep learning models can learn from many kinds of input data, including numbers, text or even combinations of input types. Recent work has drawn on a variety of rich data sources involving molecular information, natural language, medical signals such as electroencephalogram (EEG) data and multimodal data. The following is a summary of applications using these data sources.

AI has enabled recent advances in the area of biochemistry, improving understanding of the structure and behavior of biomolecules[42–45]. The work of Senior et al. on AlphaFold represented a breakthrough in the key task of protein folding, which involves predicting the 3D structure of a protein from its chemical sequence[42]. Improvements in protein structure prediction can provide mechanistic insight into a range of phenomena, such as drug–protein interactions or the effects of mutations. Alley et al. also made strides in the area of protein analysis, creating statistical summaries that capture key properties of proteins and help neural networks learn with less data[43]. By using such summaries rather than raw chemical sequences, models for downstream tasks like predicting molecular function may obtain high performance with much less labeled data.

AI has also made strides in the field of genomics, despite the complexity of modeling 3D genomic interactions. When applied to data on circulating cell-free DNA, AI has enabled noninvasive cancer detection, prognosis and tumor origin identification[46–48]. Deep learning has enhanced CRISPR-based gene editing efforts, helping to predict guide-RNA activity and identify anti-CRISPR protein families[49,50]. Additionally, AI-based analysis of microbial transcriptomic and genomic data has been used to quickly detect antibiotic resistance in pathogens. This advance allows doctors to rapidly select the most effective treatments, potentially reducing mortality and preventing the unnecessary use of broad-spectrum antibiotics[51].

Furthermore, AI is now beginning to accelerate the process of drug discovery. Deep learning models for molecular analysis have been shown to accelerate the discovery of novel drugs by reducing the need for slower, more costly physical experiments. Such models have proven useful for predicting relevant physical properties such as the bioactivity or toxicity of potential drugs. One study used AI to

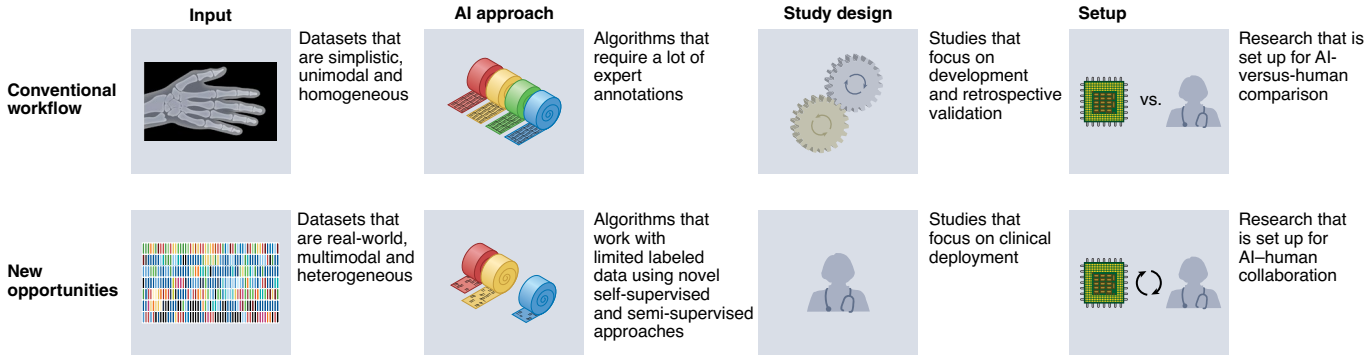

Fig. 2 | Opportunities for the development of AI algorithms.

identify a drug that was subsequently proven to be effective at fighting antibiotic-resistant bacteria in experimental models[52]. Another drug designed by AI was shown to inhibit DDR1 (a receptor implicated in several diseases, including fibrosis) in experimental models; remarkably, it was discovered in only 21 days and experimentally tested in 46 days, dramatically accelerating a process that usually takes several years[53]. Importantly, deep learning models can select effective molecules that differ from existing drugs in clinically meaningful ways, thereby opening novel pathways for treatment and providing new tools in the fight against drug-resistant pathogens.

Recent research has exploited the availability of large medical text datasets for healthcare-related natural language processing tasks, taking advantage of technical advances like transformers and contextual word embeddings (two technologies that help models consider surrounding context when interpreting each part of a text). One study presented BioBERT, a model trained on a large corpus of medical texts that surpassed prior state-of-the-art performance on natural language tasks like answering biomedical questions[54]. Such models have been used to improve performance on tasks such as learning from biomedical literature which drugs are known to interact with each other[55] or automatically labeling radiology reports[56]. Sizable text datasets have also been mined from social media and used to track large-scale mental health trends[57]. Thus, advances in natural language processing have opened up a wealth of new datasets and AI opportunities, although major limitations still exist due to the difficulty of extracting information from long text sequences.

Additionally, ML methods have been used to predict outcomes from medical signal data, such as EEG[58], electrocardiogram[59,60] and audio data[61]. For example, ML applied to EEG signals from clinically unresponsive patients with brain injuries allowed the detection of brain activity, a predictor of eventual recovery[58]. Moreover, AI's ability to directly transform brain waves to speech or text has remarkable potential value for patients with aphasia or locked-in syndrome who have had strokes[62]. Medical signal data can also be passively collected outside a clinical setting in the real world by using wearable sensors such as smartwatches that enable remote health monitoring[59,63].

Some deep learning models integrate multiple sources of medical data for a multimodal approach[64–68]. For instance, one model for diagnosing respiratory disorders took audio recordings of patients' coughs as well as reports of their symptoms as input[65]. Multimodal models have also taken advantage of far more complex inputs, such as electronic health records, which encompass a wide variety of data such as medical diagnoses, vital signs, prescriptions and laboratory results[66,67]. Such models can make predictions based on diverse types of data, much as human clinicians rely on multiple types of information when making decisions in practice. Despite its potential, this area of research seems relatively underdeveloped, in part because of the challenges of gathering multiple types of data consistently across departments or institutions. We nonetheless expect to see the use of multimodal models increase over time.

AI setups beyond supervised learning. In addition to using novel data sources, recent studies have tried unconventional problem formulations. Conventionally, datasets derive inputs and labels from real data, and models like neural networks are used to learn functions mapping from inputs to labels. However, because labeling can be expensive and time-consuming, datasets containing both accurate inputs and labels are often difficult to obtain and are frequently reused across many studies. Other paradigms, including unsupervised learning (specifically self-supervised learning), semi-supervised learning, causal inference and reinforcement learning (Box 1), have been used to tackle problems in which data are unlabeled or otherwise noisy. These advances have pushed the boundary of medical AI, enhancing existing technologies and deepening the understanding of diseases.

Unsupervised learning, which involves learning from data without any labels, has provided actionable insights, allowing models to find novel patterns and categories rather than being limited to existing labels, as in the supervised paradigm[69–73,74]. For example, clustering algorithms, which organize unlabeled data points by grouping similar data points together, have been applied to conditions such as sepsis, breast cancer and endometriosis, identifying clinically meaningful patient subgroups[29,74,75,76]. These categories can reveal novel patterns in disease manifestation that may eventually help to determine diagnosis, prognosis and treatment.

Other formulations rely on extracting information out of noisy or otherwise imperfect data, dramatically reducing the cost of data collection[30,77]. As an example, Campanella et al. trained a weakly supervised model to diagnose several types of cancer from whole-slide images, using only the final diagnoses as labels and skipping the pixel-wise annotation expected in a supervised learning setup. With this approach, they were able to achieve excellent classification results, even with annotation costs lowered[30]. Unconventional problem formulations have also been used to enhance and reconstruct images[78–81]. For instance, when creating a model to enhance spatial detail in low-quality magnetic resonance imaging (MRI) images, Masutani et al. synthetically generated input data; they took high-quality MRI images, randomly added noise and then trained a convolutional neural network (a type of neural network commonly used for image data) to recover the original high-quality MRI images from their simulated 'low-quality' inputs[80]. Such formulations allow researchers to leverage large datasets, despite their imperfections, to train high-performing models.

Setups beyond human versus AI. Although the majority of studies have focused on a head-to-head comparison of AI with humans[82], real-life medical practice is more likely to involve human-in-the-loop setups, where in humans actively collaborate with AI systems and

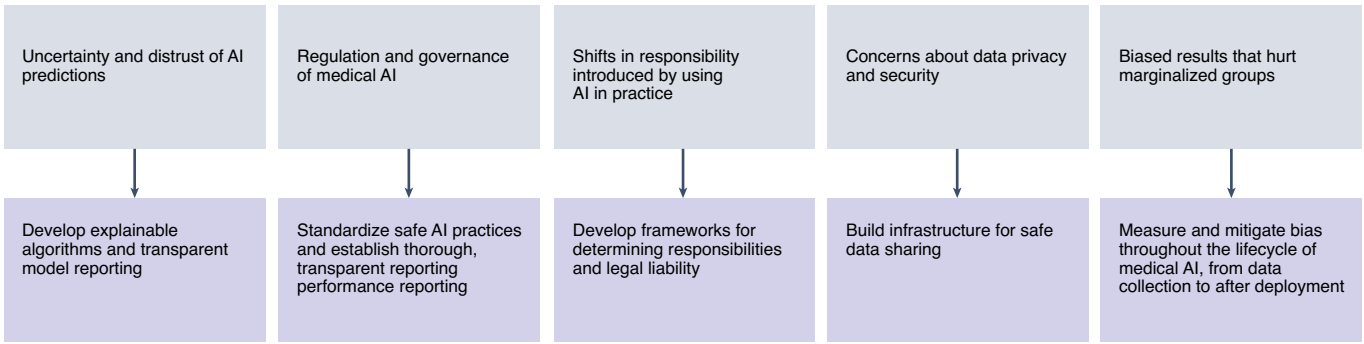

Fig. 3 | Ethical challenges for AI in medicine.

provide oversight[83,84]. Thus, recent studies have begun to explore such collaborative setups between AI and humans. These setups typically feature humans receiving assistance from AI, although occasionally AI and humans work separately and have their predictions averaged or otherwise combined afterward. Multiple studies on a variety of tasks have shown that clinical experts and AI in combination achieve better performance than experts alone[21,85–89]. For example, Sim et al. found that AI-assisted clinical experts surpassed both humans and AI alone when detecting malignant nodules on chest radiographs[85]. The usefulness of human–AI collaboration will likely depend on the specifics of the task and the clinical context.

There are still open questions about exactly how AI assistance affects human performance. For instance, AI assistance has sometimes been shown to improve clinical experts' sensitivity while lowering their specificity[8,86], and some studies, both prospective and retrospective, have found that combined AI–human performance could not surpass the performance of AI alone[90,91]. Furthermore, some clinicians may benefit more from AI assistance than others; studies suggest that less experienced clinicians, such as trainees, benefit more from AI input than their more experienced peers[8,92].

Technical considerations also play a major role in determining the effectiveness of AI assistance. Predictably, the accuracy of AI advice can affect its usefulness, so incorrect predictions have been found to hinder clinician performance even if correct predictions prove helpful[8]. Additionally, AI predictions can be communicated in multiple ways, appearing, for example, as probabilities, text recommendations or images edited to highlight areas of interest. The presentation format of AI assistance has been shown to affect its helpfulness to human users[90,91], so future work on optimizing medical AI assistance may draw on existing research on human–computer interactions.

Challenges for the future of the field

Despite striking advances, the field of medical AI faces major technical challenges, particularly in terms of building user trust in AI systems and composing training datasets. Questions also remain about the regulation of AI in medicine and the ways in which AI may shift and create responsibilities throughout the healthcare system, affecting researchers, physicians and patients alike. Finally, there exist important ethical concerns about data use and equity in medical AI (Fig. 3).

Implementation challenges. *Dataset limitations.* Medical AI data often raise specific, practical challenges. Although it is hoped that AI will reduce medical costs, the devices required to obtain the inputs for AI systems can be prohibitively expensive. Specifically, the equipment needed to capture images of whole slides is costly and is therefore unavailable in many health systems, impeding both data collection for and deployment of AI systems for pathology.

Additional concerns arise from large image sizes, because the amount of memory required by a neural network can increase with both the complexity of the model and the number of pixels in the input. As a result, many medical images, especially whole-slide images, which can easily contain billions of pixels each, are too large to fit into the average neural network. There exist many ways to address this issue. Pictures may be resized at the expense of fine details, or they may be split into multiple small patches, although this will hinder the system's ability to draw connections between different areas of the image. In other cases, humans may identify a smaller region of interest, such as part of a slide image that contains a tumor, and crop the image before feeding it into an AI system, though this intervention adds a manual step into what might otherwise be a fully automated workflow[32,93]. Some studies use large custom models that can accept whole medical images, but running these models can require expensive hardware with more memory. Thus, systems for medical image classification often involve trade-offs to make inputs compatible with neural networks.

Another issue affecting images as well as many other types of medical data is a shortage of the labels required for supervised learning[94]. Labels are often hand-assigned by medical experts, but this approach can prove difficult due to dataset size, time constraints or shortage of expertise. In other cases, labels can be provided by non-expert humans, for example, via crowdsourcing. However, such labels may be less accurate, and crowdsourced labeling projects face complications associated with privacy, as the data must be shared with many labelers. Labels can also be applied by other AI models, as in some weak-supervision setups[95], but these labels again carry the risk of noise. Currently, the difficulty of obtaining quality labels is a major blockade for supervised learning projects, driving interest in platforms that make labeling more efficient and weakly supervised and unsupervised setups that require less labeling effort.

Problems also arise when technological factors lead to bias in datasets. For example, single-source bias occurs when a single system generates an entire dataset, as when all the images in a collection come from a single camera with fixed settings. Models that exhibit single-source bias may underperform on inputs collected from other sources. To improve generalization, models can undergo site-specific training to adapt to the specific quirks of each place they are deployed, and they can also be trained and validated on datasets collected from different sources[94,96]. However, the latter approach must be undertaken with care, especially when the distribution of labels differs dramatically across datasets. For instance, if a model is trained on datasets from two institutions, one containing only positive cases and one containing only negative cases, then it can achieve high performance through spurious 'shortcuts' without learning about the relevant pathology. An image classification model might thus base its predictions entirely on the differences between the two institutions' cameras; such a model would likely learn nothing about the underlying disease and fail to generalize

Box 1 | AI setups beyond supervised learning

Self-supervised learning	Learning from unlabeled data by leveraging information extracted from the data itself
Semi-supervised learning	Learning from a small amount of labeled data combined with a large amount of unlabeled data
Causal inference	Finding the effect of a component or treatment on a system using data
Reinforcement learning	Learning in an interactive environment using feedback from actions and past experiences

elsewhere. We therefore encourage researchers to be wary of technological bias, even when using data from diverse sources[97].

Building model trust. A variety of qualities are desired for an AI system to garner user trust. For example, it is useful for AI systems to be reliable, convenient to use and easy to integrate into clinical workflows[98]. AI systems can be packaged with easy-to-read instructions, explaining how and when they should be used; it may be helpful for such user manuals to be standardized across systems[99].

Explainability is another key aspect of earning trust, as it is easier to buy into an AI system's predictions when the system can explain how it reached its conclusions. Because many AI systems currently function as uninterpretable 'black boxes', explaining their predictions poses a serious technical challenge. Some methods for explaining AI predictions exist, such as saliency methods that highlight regions of an image that most contribute to a prediction of a disease by a model. However, these methods may not be reliable[100], and further research is necessary to interpret AI decision-making processes, quantify their reliability and convey those interpretations clearly to human audiences[101]. In addition to building trust among users, enhanced explainability will allow developers to check models more thoroughly for errors and verify to what degree AI decision-making mirrors expert human approaches[102]. Moreover, when medical AI models achieve novel insights that go beyond current human knowledge, improved explainability may help researchers grasp those new insights and thus better understand the biological mechanisms behind disease.

Perhaps the most obvious component of trustworthiness is accuracy, because users are unlikely to trust a model that has not been rigorously shown to give correct predictions. Additionally, trustworthy AI studies should be reproducible, so that repeatedly training a model with a given dataset and protocol produces consistent results. Studies should also be replicable, so that models perform consistently even when trained with different samples of data. Unfortunately, proving the reproducibility and replicability of AI studies raises unique challenges. Datasets, code and trained models are often not released publicly, making it difficult for the wider AI community to independently verify and build on previous results[103,104].

Accountability. *Regulatory challenges.* Recent work highlights regulatory issues regarding the deployment of AI models for healthcare. Beyond accuracy, regulators can look at a variety of criteria to evaluate models. For example, they may require validation studies showing that AI systems are robust and generalizable across clinical settings and patient populations and ensure that systems protect patient privacy. Additionally, because the usefulness of AI systems can depend heavily on how humans provide input and interpret output, regulators may require testing of human factors and adequate training for the human users of medical AI systems[105].

Specific regulatory challenges arise from continual learning, where models learn from new data over time and adjust to shifts in patient populations, as this may come at the risk of overwriting previously learned patterns or otherwise causing new mistakes[106]. Traditionally, regulators of AI systems approve only one locked set of parameters, yet this approach does not account for the necessity to update models, as data evolve due to changes in patient populations, data collection tools and care management. Regulators must therefore develop novel certification processes to handle such systems. Importantly, the FDA has recently proposed a framework for adaptive AI systems in which they would approve not only an initial model but also a process for updating it over time[107].

Shifts in responsibility. Although AI systems have the potential to empower humans in medical decision-making, they also run the risk of limiting personal autonomy and creating new obligations. As AI systems take on more responsibilities in the healthcare setting, a concern facing the system is that clinicians may become overly reliant on AI, perhaps seeing a gradual decline in their own skills or personal connections with patients. In turn, medical AI developers may gain outsized influence on healthcare and should thus be obliged to create safe, useful AI systems and responsibly influence public views on health. As medical decision-making becomes more reliant on potentially unexplained AI judgments, individual patients might lose some understanding of, or control over, their own care. Patients might at the same time gain new responsibilities as AI makes healthcare more pervasive in daily life. As an example, if smart devices provide patients with constant advice, then those patients may be expected to follow those recommendations or else be responsible for negative health outcomes[108].

The proliferation of AI also raises concerns around accountability, as it is currently unclear whether developers, regulators, sellers or healthcare providers should be held accountable if a model makes mistakes even after being thoroughly clinically validated. Currently, doctors are held liable when they deviate from the standard of care and patient injury occurs. If doctors are generally skeptical of medical AI, then individual doctors may be adversely influenced to ignore AI recommendations that conflict with standard practice, even if those recommendations may be personalized and beneficial for a specific patient. However, if the standard of care shifts so that doctors routinely use AI tools, then there will be a strong medicolegal incentive for doctors to follow AI recommendations[109].

Fairness. *Ethical data use.* There are concerns that bad actors interested in identity theft and other misconduct might take advantage of medical datasets, which often contain large amounts of sensitive information about real patients. Decentralizing data storage is one way to reduce the potential damage of any individual hack or data leak. The process of federated learning facilitates such decentralization while also making it easier to collaborate across institutions without complicated data-sharing agreements (Fig. 4). When using federated learning, developers send AI models to different institutions that have private datasets; the institutions train the models on their data and send back model updates without ever sharing the data[110]. However, even after models are trained, there remains the risk that AI systems will face privacy attacks, which can sometimes reconstruct original data points used in training just by examining the resulting model. Patient data can be better protected from such attacks if inputs are encrypted before training, but this approach comes at the cost of model interpretability[111].

Looking beyond such bad-faith attacks, there are other questions about how to respect patients' privacy. Sensitive data should typically be collected and used in research with patient consent and, where practical, anonymization and aggregation strategies should be used to obscure personal details. It is necessary to ensure that any institutions working with patient data handle them responsibly, for

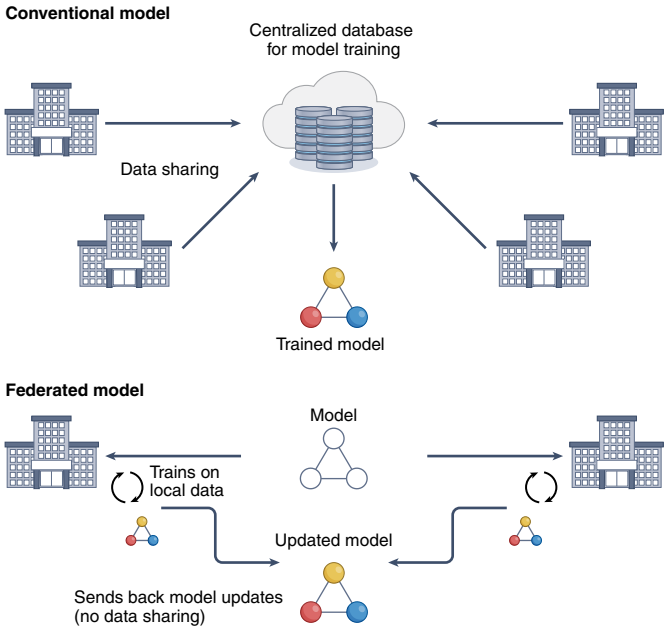

Fig. 4 | Evolving procedures for data sharing. An advantage of federated learning is that it is decentralized, representing a major potential advance in data security.

example, by using appropriate security protocols. At the same time, it is also important that patient data be used for the good of patients. Out of respect for the patients who have agreed to share their personal information, patient data ideally would be used for research that promotes future patient well-being. Unfortunately, these goals can sometimes conflict with each other; implementing security measures like federated learning can require considerable resources and effort, and institutions that cannot make those investments may be unable to access certain datasets, even when their research would benefit the patients in question. Additionally, reusing data across projects may make it more difficult to obtain informed consent, because patients allured by one study may be hesitant to join others. We hope and expect the AI community will continue exploring these trade-offs and find new ways to balance a variety of patient interests[110].

Equity and bias. AI can make healthcare more accessible to underserved groups, but it also risks reinforcing existing inequities, because AI models can perpetuate biases lurking in the data[112]. Medical AI systems can fail to generalize to new kinds of data they were not trained on; thus, training on datasets that underrepresent marginalized groups is well known to result in biased systems that underperform on those groups. Systems that explicitly factor race into their predictions are also at risk of perpetuating prejudice, because racial categories are difficult to define and obscure the diversity within racial groups[113]. Bias can creep in due to other design choices, such as the choice of target label. For example, a risk-assessment algorithm used to guide clinical decision-making for 200 million patients was found to give racially biased predictions, such that white patients assigned a certain predicted risk score tended to be healthier than Black patients with the same score. This bias was due in large part to the original labels used in training. The system was trained to predict future healthcare costs, but because Black patients had historically received less expensive care than white patients due to existing systematic biases, the system reproduced those racial biases in its predictions[114]. Extensive research is necessary to detect and correct bias in medical AI models, because bias can cause widespread harm to marginalized groups if left unchecked. In the future, AI tools may systematically undergo special testing before deployment to verify that neural networks serve the well-being of marginalized populations equitably. Additionally, it may become easier to identify dangerous bias if model explainability improves, because human monitors will be able to double check the reasoning of AI systems and identify problematic elements[115].

Conclusion

The field of medical AI has made considerable progress toward large-scale deployment, especially through prospective studies such as RCTs and through medical image analysis, yet medical AI remains in an early phase of validation and implementation. To date, a limited number of studies have used external validation, prospective evaluation and diverse metrics to explore the full impact of AI in real clinical settings, and the range of assessed use cases has been relatively narrow. Although the field requires more testing and practical solutions, there is also a need for bold imagination. AI has proven capable of extracting insights from unexpected sources and drawing connections that humans would not normally anticipate, so we hope to see even more creative, out-of-the-box approaches to medical AI. There are rich opportunities for novel AI research involving non-image data types and unconventional problem formulations, which open a broader array of possible datasets. Opportunities also exist in AI–human collaboration, an alternative to the AI-versus-human competitions common in research; we would like to see collaborative setups receive more study, as they may provide better results than either AI or humans alone and are more likely to reflect real medical practice. Despite the potential of the field, major technical and ethical questions remain for medical AI. As these pivotal issues are systematically addressed, the potential of AI to markedly improve the future of medicine may be realized.

Received: 23 July 2021; Accepted: 5 November 2021;
Published online: 20 January 2022

References

1. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *J. Am. Med. Assoc.* **316**, 2402–2410 (2016).
2. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
3. Rajpurkar, P. et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med.* **15**, e1002686 (2018).
4. Hannun, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **25**, 65–69 (2019).
5. Wiens, J. et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat. Med.* **25**, 1337–1340 (2019).
6. Kanagasingam, Y. et al. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. *JAMA Netw. Open* **1**, e182665 (2018).
7. Beede, E. et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* 1–12 (Association for Computing Machinery, 2020); https://dl.acm.org/doi/abs/10.1145/3313831.3376718
8. Kiani, A. et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *NPJ Digit. Med.* **3**, 23 (2020).
9. Lin, H. et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine* **9**, 52–59 (2019).
10. Gong, D. et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol. Hepatol.* **5**, 352–361 (2020).
11. Wang, P. et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. *Lancet Gastroenterol. Hepatol.* **5**, 343–351 (2020).
12. Hollon, T. C. et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat. Med.* **26**, 52–58 (2020).

13. Phillips, M. et al. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Netw. Open* **2**, e1913436 (2019).
14. Nimri, R. et al. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. *Nat. Med.* **26**, 1380–1384 (2020).
15. Wijnberge, M. et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs. standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery. *J. Am. Med. Assoc.* **323**, 1052–1060 (2020).
16. Wismüller, A. & Stockmaster, L. A prospective randomized clinical trial for measuring radiology study reporting time on Artificial Intelligence-based detection of intracranial hemorrhage in emergent care head CT. in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging* vol. 11317, 113170M (International Society for Optics and Photonics, 2020).
17. Liu, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Br. Med. J.* **370**, m3164 (2020).
18. Rivera, S. C. et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat. Med.* **26**, 1351–1363 (2020).
19. Centers for Medicare & Medicaid Services. Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Final Policy Changes and Fiscal Year 2021 Rates; Quality Reporting and Medicare and Medicaid Promoting Interoperability Programs Requirements for Eligible Hospitals and Critical Access Hospitals. *Fed. Regist.* **85**, 58432–59107 (2020).
20. Benjamens, S., Dhunnoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit. Med.* **3**, 118 (2020).
21. Wu, N. et al. Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Trans. Med. Imaging* **39**, 1184–1194 (2020).
22. McKinney, S. M. et al. International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020).
23. Ghorbani, A. et al. Deep learning interpretation of echocardiograms. *NPJ Digit. Med.* **3**, 10 (2020).
24. Ouyang, D. et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).
25. Ardila, D. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **25**, 954–961 (2019).
26. Huynh, E. et al. Artificial intelligence in radiation oncology. *Nat. Rev. Clin. Oncol.* **17**, 771–781 (2020).
27. Huang, P. et al. Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method. *Lancet Digit. Health* **1**, e353–e362 (2019).
28. Kather, J. N. et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
29. Jackson, H. W. et al. The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).
30. Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).
31. Fu, Y. et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* **1**, 800–810 (2020).
32. Courtiol, P. et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat. Med.* **25**, 1519–1525 (2019).
33. Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology: new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019).
34. Zhou, D. et al. Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. *Nat. Commun.* **11**, 2961 (2020).
35. Zhao, S. et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. *Gastroenterology* **156**, 1661–1674 (2019).
36. Freedman, D. et al. Detecting deficient coverage in colonoscopies. *IEEE Trans. Med. Imaging* **39**, 3451–3462 (2020).
37. Liu, H. et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol.* **137**, 1353–1360 (2019).
38. Milea, D. et al. Artificial intelligence to detect papilledema from ocular fundus photographs. *N. Engl. J. Med.* **382**, 1687–1695 (2020).
39. Wolf, R. M., Channa, R., Abramoff, M. D. & Lehmann, H. P. Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmol.* **138**, 1063–1069 (2020).
40. Xie, Y. et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit. Health* **2**, e240–e249 (2020).
41. Arcadu, F. et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit. Med.* **2**, 92 (2019).
42. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
43. Alley, E. C., Khimulya, G., Biswas, S., AlQuraishi, M. & Church, G. M. Unified rational protein engineering with sequence-based deep representation learning. *Nat. Methods* **16**, 1315–1322 (2019).
44. Gainza, P. et al. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **17**, 184–192 (2020).
45. Greener, J.G. et al. Deep learning extends de novo protein modelling coverage of genomes using iteratively predicted structural constraints. *Nat. Commun.* **10**, 3977 (2019).
46. Chabon, J. J. et al. Integrating genomic features for non-invasive early lung cancer detection. *Nature* **580**, 245–251 (2020).
47. Luo, H. et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. *Sci. Transl. Med.* **12**, eaax7533 (2020).
48. Cristiano, S. et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389 (2019).
49. Gussow, A. B. et al. Machine-learning approach expands the repertoire of anti-CRISPR protein families. *Nat. Commun.* **11**, 3784 (2020).
50. Wang, D. et al. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat. Commun.* **10**, 4284 (2019).
51. Bhattacharyya, R. P. et al. Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. *Nat. Med.* **25**, 1858–1864 (2019).
52. Stokes, J. M. et al. A deep learning approach to antibiotic discovery. *Cell* **181**, 475–483 (2020).
53. Zhavoronkov, A. et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040 (2019).
54. Lee, J. et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **36**, 1234–1240 (2020).
55. Zhu, Y., Li, L., Lu, H., Zhou, A. & Qin, X. Extracting drug-drug interactions from texts with BioBERT and multiple entity-aware attentions. *J. Biomed. Inform.* **106**, 103451 (2020).
56. Smit, A. et al. CheXbert: Combining automatic labelers and expert annotations for accurate radiology report labeling using BERT. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* 1500–1519 (2020).
57. Sarker, A., Gonzalez-Hernandez, G., Ruan, Y. & Perrone, J. Machine learning and natural language processing for geolocation-centric monitoring and characterization of opioid-related social media chatter. *JAMA Netw. Open* **2**, e1914672 (2019).
58. Claassen, J. et al. Detection of brain activation in unresponsive patients with acute brain injury. *N. Engl. J. Med.* **380**, 2497–2505 (2019).
59. Porumb, M., Stranges, S., Pescapè, A. & Pecchia, L. Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG. *Sci. Rep.* **10**, 170 (2020).
60. Attia, Z. I. et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* **394**, 861–867 (2019).
61. Chan, J., Raju, S., Nandakumar, R., Bly, R. & Gollakota, S. Detecting middle ear fluid using smartphones. *Sci. Transl. Med.* **11**, eaav1102 (2019).
62. Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M. & Shenoy, K. V. High-performance brain-to-text communication via handwriting. *Nature* **593**, 249–254 (2021).
63. Green, E. M. et al. Machine learning detection of obstructive hypertrophic cardiomyopathy using a wearable biosensor. *NPJ Digit. Med.* **2**, 57 (2019).
64. Thorsen-Meyer, H.-C. 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* **2**, e179–e191 (2020).
65. Porter, P. et al. A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children. *Respir. Res.* **20**, 81 (2019).
66. Tomašev, N. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* **572**, 116–119 (2019).
67. Kehl, K. L. et al. Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. *JAMA Oncol.* **5**, 1421–1429 (2019).
68. Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit. Med.* **3**, 136 (2020).

69. Wang, C. et al. Quantitating the epigenetic transformation contributing to cholesterol homeostasis using Gaussian process. *Nat. Commun.* **10**, 5052 (2019).
70. Li, Y. et al. Inferring multimodal latent topics from electronic health records. *Nat. Commun.* **11**, 2536 (2020).
71. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
72. Li, X. et al. Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nat. Commun.* **11**, 2338 (2020).
73. Amodio, M. et al. Exploring single-cell data with deep multitasking neural networks. *Nat. Methods* **16**, 1139–1145 (2019).
74. Urteaga, I., McKillop, M. & Elhadad, N. Learning endometriosis phenotypes from patient-generated data. *NPJ Digit. Med.* **3**, 88 (2020).
75. Brbić, M. et al. MARS: discovering novel cell types across heterogeneous single-cell experiments. *Nat. Methods* **17**, 1200–1206 (2020).
76. Seymour, C. W. et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *J. Am. Med. Assoc.* **321**, 2003–2017 (2019).
77. Fries, J. A. et al. Weakly supervised classification of aortic valve malformations using unlabeled cardiac MRI sequences. *Nat. Commun.* **10**, 3111 (2019).
78. Jin, L. et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nat. Commun.* **11**, 1934 (2020).
79. Vishnevskiy, V. et al. Deep variational network for rapid 4D flow MRI reconstruction. *Nat. Mach. Intell.* **2**, 228–235 (2020).
80. Masutani, E. M., Bahrami, N. & Hsiao, A. Deep learning single-frame and multiframe super-resolution for cardiac MRI. *Radiology* **295**, 552–561 (2020).
81. Rana, A. et al. Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. *JAMA Netw. Open* **3**, e205111 (2020).
82. Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit. Health* **1**, e271–e297 (2019).
83. Chen, P.-H. C. et al. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat. Med.* **25**, 1453–1457 (2019).
84. Patel, B. N. et al. Human–machine partnership with artificial intelligence for chest radiograph diagnosis. *NPJ Digit. Med.* **2**, 111 (2019).
85. Sim, Y. et al. Deep convolutional neural network–based software improves radiologist detection of malignant lung nodules on chest radiographs. *Radiology* **294**, 199–209 (2020).
86. Park, A. et al. Deep learning–assisted diagnosis of cerebral aneurysms using the HeadXNet model. *JAMA Netw. Open* **2**, e195600 (2019).
87. Steiner, D. F. et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am. J. Surg. Pathol.* **42**, 1636–1646 (2018).
88. Jain, A. et al. Development and assessment of an artificial intelligence-based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. *JAMA Netw. Open* **4**, e217249 (2021).
89. Seah, J. C. Y. et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit. Health* **3**, e496–e506 (2021).
90. Rajpurkar, P. et al. CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV. *NPJ Digit. Med.* **3**, 115 (2020).
91. Kim, H.-E. et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit. Health* **2**, e138–e148 (2020).
92. Tschandl, P. et al. Human–computer collaboration for skin cancer recognition. *Nat. Med.* **26**, 1229–1234 (2020).
93. van der Laak, J., Litjens, G. & Ciompi, F. Deep learning in histopathology: the path to the clinic. *Nat. Med.* **27**, 775–784 (2021).
94. Willemink, M. J. et al. Preparing medical imaging data for machine learning. *Radiology* **295**, 4–15 (2020).
95. Irvin, J. et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. in *Proceedings of the AAAI Conference on Artificial Intelligence* vol. 33, 590–597 (2019).
96. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G. & King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* **17**, 195 (2019).
97. DeGrave, A. J., Janizek, J. D. & Lee, S.-I. AI for radiographic COVID-19 detection selects shortcuts over signal. *Nat. Mach. Intell.* **3**, 610–619 (2021).
98. Cutillo, C. M. et al. Machine intelligence in healthcare: perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit. Med.* **3**, 47 (2020).
99. Sendak, M. P., Gao, M., Brajer, N. & Balu, S. Presenting machine learning model information to clinical end users with model facts labels. *NPJ Digit. Med.* **3**, 41 (2020).
100. Saporta, A. et al. Deep learning saliency maps do not accurately highlight diagnostically relevant regions for medical image interpretation. Preprint at *medRxiv* https://doi.org/10.1101/2021.02.28.21252634 (2021).
101. Ehsan, U. *et al*. The who in explainable AI: how AI background shapes perceptions of AI explanations. Preprint at https://arxiv.org/abs/2107.13509 (2021).
102. Reyes, M. et al. On the interpretability of artificial intelligence in radiology: Challenges and opportunities. *Radio. Artif. Intell.* **2**, e190043 (2020).
103. Liu, C. *et al*. On the replicability and reproducibility of deep learning in software engineering. Preprint at https://arxiv.org/abs/2006.14244 (2020).
104. Beam, A. L., Manrai, A. K. & Ghassemi, M. Challenges to the reproducibility of machine learning models in health care. *J. Am. Med. Assoc.* **323**, 305–306 (2020).
105. Gerke, S., Babic, B., Evgeniou, T. & Cohen, I. G. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ Digit. Med.* **3**, 53 (2020).
106. Lee, C. S. & Lee, A. Y. Clinical applications of continual learning machine learning. *Lancet Digit. Health* **2**, e279–e281 (2020).
107. Food and Drug Administration. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD): Discussion Paper and Request for Feedback* (FDA, 2019).
108. Morley, J. et al. The debate on the ethics of AI in health care: a reconstruction and critical review. *SSRN* http://dx.doi.org/10.2139/ssrn.3486518 (2019.
109. Price, W. N., Gerke, S. & Cohen, I. G. Potential liability for physicians using artificial intelligence. *J. Am. Med. Assoc.* **322**, 1765–1766 (2019).
110. Larson, D. B., Magnus, D. C., Lungren, M. P., Shah, N. H. & Langlotz, C. P. Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology* **295**, 675–682 (2020).
111. Kaissis, G. A., Makowski, M. R., Rückert, D. & Braren, R. F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat. Mach. Intell.* **2**, 305–311 (2020).
112. Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H. & Ferrante, E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc. Natl Acad. Sci. USA* **117**, 12592–12594 (2020).
113. Vyas, D. A., Eisenstein, L. G. & Jones, D. S. Hidden in plain sight: reconsidering the use of race correction in clinical algorithms. *N. Engl. J. Med.* **383**, 874–882 (2020).
114. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
115. Cirillo, D. et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digit. Med.* **3**, 81 (2020).

Acknowledgements

We thank A. Tamkin and N. Phillips for their feedback. E.J.T. receives funding support from US National Institutes of Health grant UL1TR002550.

Author contributions

P.R. and E.J.T. conceptualized this Review. E.C., O.B. and P.R. were responsible for the design and synthesis of this Review. All authors contributed to writing and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Eric J. Topol.

Peer review information *Nature Medicine* thanks Despina Kontos and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Karen O'Leary was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.