Current Status and Future Directions: The Application of Artificial Intelligence/Machine Learning for Precision Medicine

Kunal Naik[1], Rahul K. Goyal[2,†], Luca Foschini[3], Choi Wai Chak[4], Christian Thielscher[5], Hao Zhu[1], James Lu[6], Joseph Lehár[4], Michael A. Pacanoswki[1], Nadia Terranova[7], Neha Mehta[1], Niklas Korsbo[8], Tala Fakhouri[9], Qi Liu[1] and Jogarao Gobburu[2,*]

Technological innovations, such as artificial intelligence (AI) and machine learning (ML), have the potential to expedite the goal of precision medicine, especially when combined with increased capacity for voluminous data from multiple sources and expanded therapeutic modalities; however, they also present several challenges. In this communication, we first discuss the goals of precision medicine, and contextualize the use of AI in precision medicine by showcasing innovative applications (e.g., prediction of tumor growth and overall survival, biomarker identification using biomedical images, and identification of patient population for clinical practice) which were presented during the February 2023 virtual public workshop entitled "Application of Artificial Intelligence and Machine Learning for Precision Medicine," hosted by the US Food and Drug Administration (FDA) and University of Maryland Center of Excellence in Regulatory Science and Innovation (M-CERSI). Next, we put forward challenges brought about by the multidisciplinary nature of AI, particularly highlighting the need for AI to be trustworthy. To address such challenges, we subsequently note practical approaches, viz., differential privacy, synthetic data generation, and federated learning. The proposed strategies – some of which are highlighted presentations from the workshop – are for the protection of personal information and intellectual property. In addition, methods such as the risk-based management approach and the need for an agile regulatory ecosystem are discussed. Finally, we lay out a call for action that includes sharing of data and algorithms, development of regulatory guidance documents, and pooling of expertise from a broad-spectrum of stakeholders to enhance the application of AI in precision medicine.

BACKGROUND

The pace of medical innovation over the past couple of decades is remarkable. The menu of therapeutic modalities has grown more diverse with proliferation of biologics, oligonucleotides, cell, and gene therapies. Diagnostics have also evolved as high throughput approaches to sequence the genome in days (or less) are now widely available. Activity trackers and smartphones, though consumer lifestyle focused, capture and store more data than ever imagined when the first smartphone was unveiled in the late 2000s.[1,2] All these novel tools and technologies have enabled greater precision in medicine, that is, the ability to select a drug, dosing, and monitoring strategy that is most likely to result in the greatest benefit to the patient while minimizing the potential for iatrogenic adverse outcomes. However, our ability to match treatments to any given individual's disease is likely still in its infancy. In looking forward, increasing amounts of data and computational power will give rise to tools that will accelerate drug development and enable greater individualization of medical care.

In the current environment, precision medicine is commonly thought of as tailoring treatments based on individual characteristics.[3] This is often accomplished using a biomarker test or some other type of tool that is an "indicator of normal biological processes, pathogenic processes, or biological responses to an exposure or intervention, including therapeutic interventions."[4] Susceptibility, prognostic, and predictive biomarkers are measured at a single point of time to forecast future events, whereas monitoring, response, and safety biomarkers are assessed serially to provide information on improvement or worsening of a clinical outcome

[1]Office of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [2]Center for Translational Medicine, University of Maryland School of Pharmacy, Baltimore, Maryland, USA; [3]Sage Bionetworks, Seattle, Washington, USA; [4]Business Strategy, Owkin, Paris, France; [5]Competence Center for Medical Economics, FOM University, Essen, Germany; [6]Modeling & Simulation/Clinical Pharmacology, Genentech Inc., South San Francisco, California, USA; [7]Quantitative Pharmacology, Ares Trading S.A. (an affiliate of Merck KGaA, Darmstadt, Germany), Lausanne, Switzerland; [8]Pumas AI, Annapolis, Maryland, USA; [9]Office of Medical Policy, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA. *Correspondence: Jogarao Gobburu (jgobburu@rx.umaryland.edu)

† These authors contributed equally to this work.

Received October 16, 2023; accepted November 28, 2023. doi:10.1002/cpt.3152

relative to a previous timepoint(s).[4] In clinical practice, the results of a biomarker assessment may inform the risk for safety events, confirm, or refine a diagnosis, identify potential responders to a particular intervention, or guide the dosage of a drug. In drug development, biomarkers can support various clinical trial maneuvers, such as enriched trial designs to allow treatment effects to be detected if they exist, or to provide evidence that a drug is effective and safe.

Most disease therapies only show profound benefit in a minority of patients because diagnoses encompass many distinct biologies.[5,6] Precision medicine aims to break cancer diagnoses into biologically distinct subtypes, then pursue personalized therapies for each group of patients.[7] This strategy is enabled by recent advances in technologies for medical imaging, molecular profiling, and artificial intelligence (AI). The precision medicine approach is neatly summarized by MD Anderson, one of the earliest adopters of this approach for oncology (see **Figure 1**). Using a variety of biomarkers (in oncology molecular profiling and pathology imaging predominate), patients can be categorized into subtypes with distinct biologies for which different treatments can be applied or developed. Using this approach, response rates can be increased dramatically within each biologic subtype, and the overall outcomes for all patients in the therapeutic area can be dramatically improved.

Most current examples of precision medicines have focused on matching a drug that has a targeted mechanism of action to a relatively specific and well-understood dimension of disease pathology. However, the biological complexity of human disease is not so straightforward. For example, molecular alterations in tumors are commonly tested to identify oncogenic drivers. The results can be used to determine whether the tumor is potentially amenable to a drug that is pharmacologically targeted to a specific driver. Indeed, numerous drugs have been approved for molecular subsets of lung tumors (e.g., cemiplimab, nivolumab, ipilimumab, and atezolizumab for tumors with high PD-L1 expression; mobocertinib, amivantamab, osimertinib, ramucirumab, and erlotinib for tumors with EGFR alterations, pralsetinib and selpercatinib for tumors with RET fusions). In almost every case, the drug was targeted for patients whose tumors have a pathogenic driver relevant to the mechanism of the drug. However, other factors, such as methylation patterns, copy number variations, and other aberrations, may be present, and even coexist.[8] If integrated, response prediction could potentially be improved. Furthermore, changes in the tumor's molecular composition over time are not currently assessed, and more dynamic approaches based on repeated measures could facilitate designing more effective treatment strategies, such as different combinations or sequences of therapies. Few examples of multifactorial and dynamic models of drug response have been implemented in practice.

Advanced data analytics are a key facet of the next generation of precision health care, which will be enabled by omic technologies, the internet of things,[9] complex diagnostics, innovative therapies, and improved infrastructure for data aggregation. It is now possible to generate unprecedented amounts of data on any given patient. What remains is an ability to harness those data, find a mechanism for a differential response or identify a subset in whom the benefit–risk profile may differ, and produce a simplistic tool that can be applied at the patient-doctor interface. Precision medicine approaches rely on technologies that improve mechanistic understanding of disease and drug response. For example, AI and complex data analytics can augment use of existing tools to further optimize drug development. AI tools in practice can also help resolve multiple factors to support individualized therapeutic decision making.

AI refers to the theory and development of computer systems able to perform tasks normally requiring human intelligence,

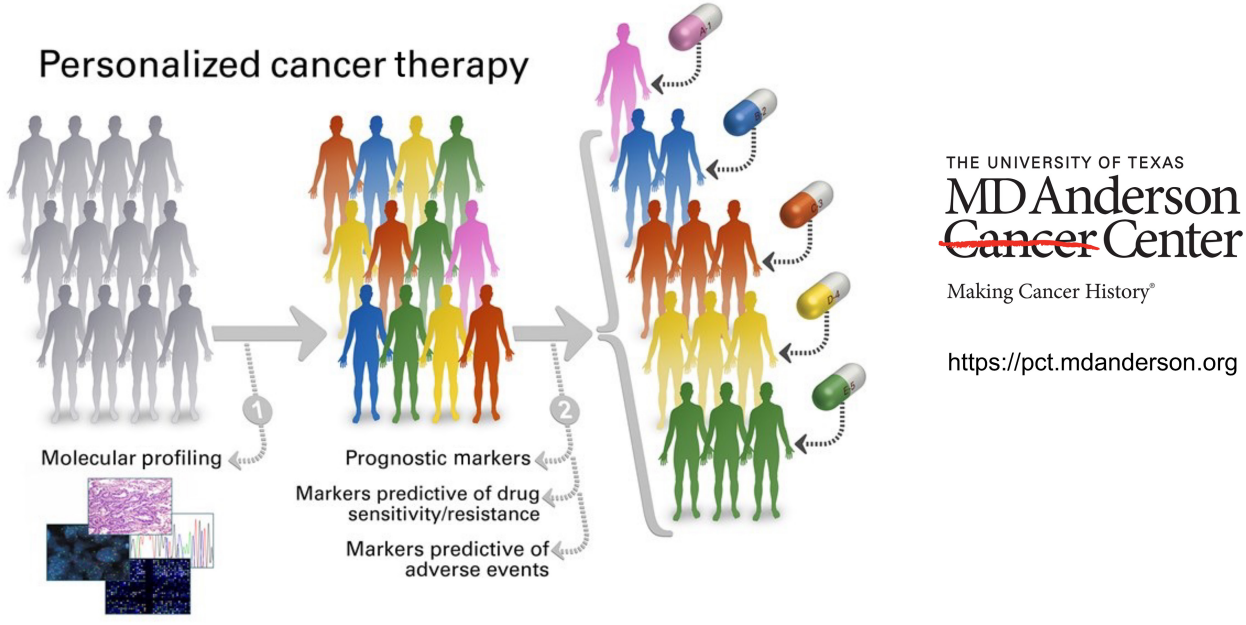

Figure 1 Precision medicine – finding the right drug for each patient (from MD Anderson Cancer Center).

in order to deliver solutions that can automate routine tasks, draw data-based insights, or augment human activities.[10] AI is classed into two categories based on the degree of autonomy – narrow AI and general AI. Narrow AI systems can perform predetermined singular tasks, such as cancer diagnosis, disease risk stratification, monitoring patient adherence to treatment, and prediction of outcomes from medical signal data (e.g., electrocardiograms (ECGs), computed tomography (CT) scans, magnetic resonance imaging (MRI) data, and genome sequences). On the other hand, general AI involves systems that exhibit human-like intelligence and tasks that can even surpass human capabilities. Although there have been great strides in narrow AI systems, general AI remains an emerging goal for current research and development.

Machine learning (ML) is a subset of AI that focuses on models and algorithms which can learn automatically from data without being explicitly programmed. Most of the recent groundbreaking advancement in the AI field are examples of ML, in particular, deep learning (DL; e.g., convolutional neural networks for computer vision, recurrent neural networks, or transformer neural networks for natural language processing). DL is a subfield of ML that uses multiple processing layers to learn representations of data with multiple levels of abstraction.[11] Insights generated from AI/ML analyses can be used for personalized diagnosis, disease prevention, and personalized treatment.

Generally speaking, ML can be divided into three major categories: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is the most widely used AI/ML methodology. At its core, supervised learning trains an ML algorithm on a labeled dataset of inputs (e.g., tabular data, text, and image) and outputs (e.g., a continuous or categorical outcome). Through the training process, the algorithm learns the mapping between inputs and outputs for performing predictions on new and unseen data. On the other hand, unsupervised learning is a methodology where an ML algorithm is trained on unlabeled data. The algorithm extracts hidden intrinsic structures in the data without prior knowledge of output labels. Unsupervised learning is generally used for tasks such as clustering (or segmentation), dimensionality reduction, and anomaly detection.

In reinforcement learning, an agent learns to make decisions by interacting with its environment and receives rewards or punishments based on the actions. The goal of the agent is to learn a policy that can maximize long-term rewards that culminates in an optimal behavior in an environment. An innovative application of reinforcement learning in precision medicine involved using a multi-cytokine therapy for the treatment of sepsis. The simulation results suggest a potential to significantly reduce mortality rate with this "adaptive" therapy as compared with antibiotics.[12]

Building upon the foundational understanding and methodologies, we now move to examining the current landscape of AI in precision medicine. In the subsequent section, we will explore the breadth of literature, regulatory framework, and practical applications of AI in precision medicine, thus bridging the gap between theoretical knowledge and real-world applications.

CURRENT LANDSCAPE

The literature landscape analyses

Over the past decade, the number of AI health applications has exponentially grown.[13] This rapid expansion in the use of AI in medical research can be attributed to several factors. First, the abundance of diverse biomedical data, including data from administrative and claims records, electronic health records (EHRs), data from registries, and more, has played a pivotal role.[2] Second, the utilization of multimodal datasets, such as data originating from digital health technologies, genomics and microbiome data, clinical laboratory and biomarker data, and data from medical imaging, has further fueled this growth.[2] Another contributing factor is related to the continuous improvements in data standards, interoperability, and healthcare data exchange, which has enabled efficient data sharing.[14] Advancements in data privacy preservation approaches have instilled confidence and trust in the utilization of AI by both clinicians and patients and are alleviating privacy concerns that have traditionally prevented data owners from providing access to their data.[15–17] However, none of these advancements would have been sufficient with the massive increase in computing power that has been the driving force behind this acceleration.[18] Notably, the application of AI methods using causal inference approaches has been instrumental in advancing the use of AI in clinical research, specifically, and particularly in areas where understanding the causal relationship between a drug and health outcomes is imperative.[19]

This rapid growth in AI health applications is reflected by the exponential increase in the number of studies found on PubMed.com using the search terms "Artificial intelligence [Mesh] OR generative artificial intelligence OR large language models," for example. Between 2016 and 2022, the count of publication increased from 6,904 to 32,429, an increase over 4.5-fold (**Figure 2**). This growth occurred over a large number of medical specialties,[13] with the largest number of applications seen in pathology, radiology, surgery, psychiatry, and oncology.

Moreover, applications of AI are within every step of the Drug Development Lifecycle leading to disease therapies (**Figure 3**). The classical drug research and development (R&D) pathway proceeds from target identification through molecular lead discovery and culminates with clinical trials. AI solutions are brought to bear to accelerate every step of the process during R&D and post-approval. In drug R&D, AI is used to interpret complex disease phenotypes, discover new targets, identify/optimize new compounds, and mine experimental and clinical data sources.[20,21] Post-approval, AI is being used to optimize drug manufacture and distribution, as well as to ensure market access. For the current pharmacopeia, AI is being increasingly used to repurpose existing drugs into new disease indications. Regulatory organizations, such as the US Food and Drug Administration (FDA), play a pivotal role in ensuring meaningful and valid applications of AI and ML on the path leading to disease therapies. Therefore, a landscape analysis was performed to summarize AI- and ML-related submissions to the Center for Drug Evaluation and Research at the FDA from 2016 to 2022.

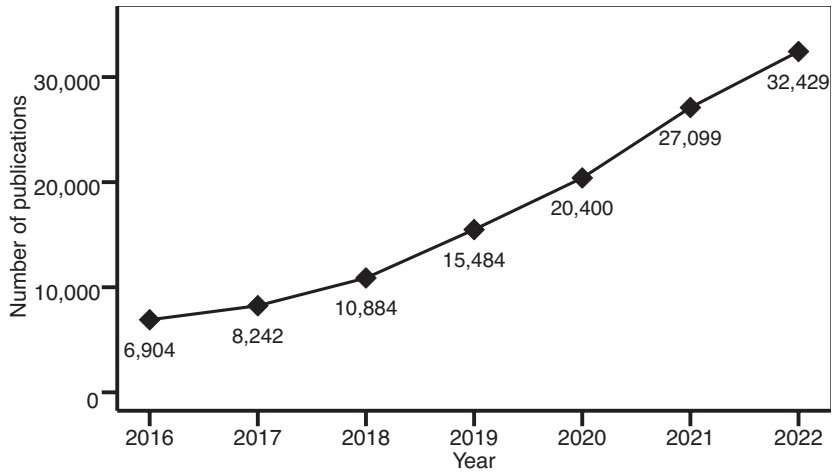

Figure 2 PubMed Search query results for: Artificial intelligence [Mesh] OR generative artificial intelligence OR large language models from 2016 to 2022.

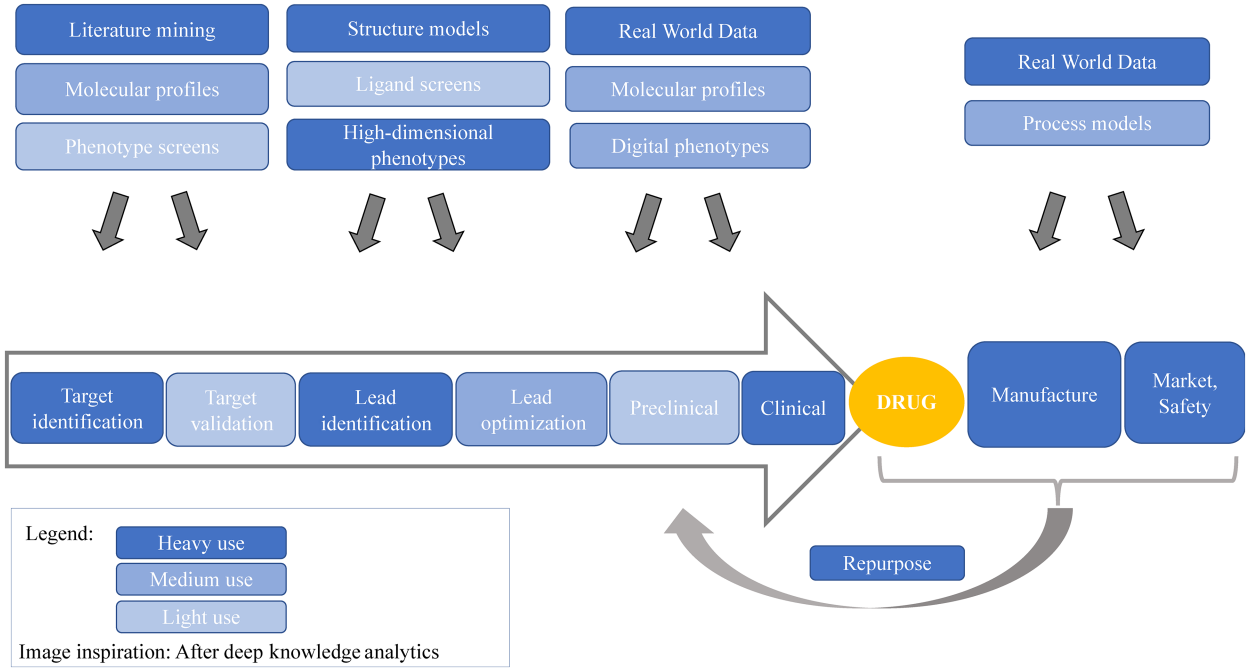

Figure 3 AI has a role in every aspect of drug discovery and development. Both during R&D (enclosed in arrow) and post-approval, AI approaches (examples above the path) are being developed to greatly accelerate each step of the process. Shading shows the authors' heuristic estimate of AI intensity at the time of publication, where less used areas are due to data paucity, undeveloped methods, or slow adoption of AI. AI, artificial intelligence; R&D, research and development.

The analysis by Liu *et al.*,[22] has provided an overview of the yearly AI and ML-related submissions in years 2016 to 2021 in the literature. The current analysis provides an update. A remarkable increase in the number of submissions has been observed in the recent 2 years. There are < 4 submissions per year prior to year 2019. Around 2019 to 2020, there are ~ 10 submissions per year. In the years of 2021 and 2022, the numbers of submissions are more than 130 and 170, respectively. Most submissions are at the Investigational New Drug stage, where drug developers are exploring the potential usage of various AI and ML tools in drug development. Only a few submissions are under New Drug Applications or Biological License Applications (**Figure 4a**).

From 2016 to the end of 2022, more than 300 meeting packages containing AI and ML components were submitted to the FDA. Among them, only a limited number of AI and ML submissions are related to drug discovery. In contrast, using AI and ML tools to guide drug design and discovery appears to be very active among drug developers. The discrepancy is anticipated because the FDA is not actively involved in the drug discovery stage. At the preclinical stage, AI and ML approaches have been applied to understand potential toxicity of new compounds. As expected, most submissions are at the clinical stage with a few submissions centered around unresolved issues after the compound is approved (**Figure 4b**). There appears to be an imbalance on the usage of AI and ML tools to

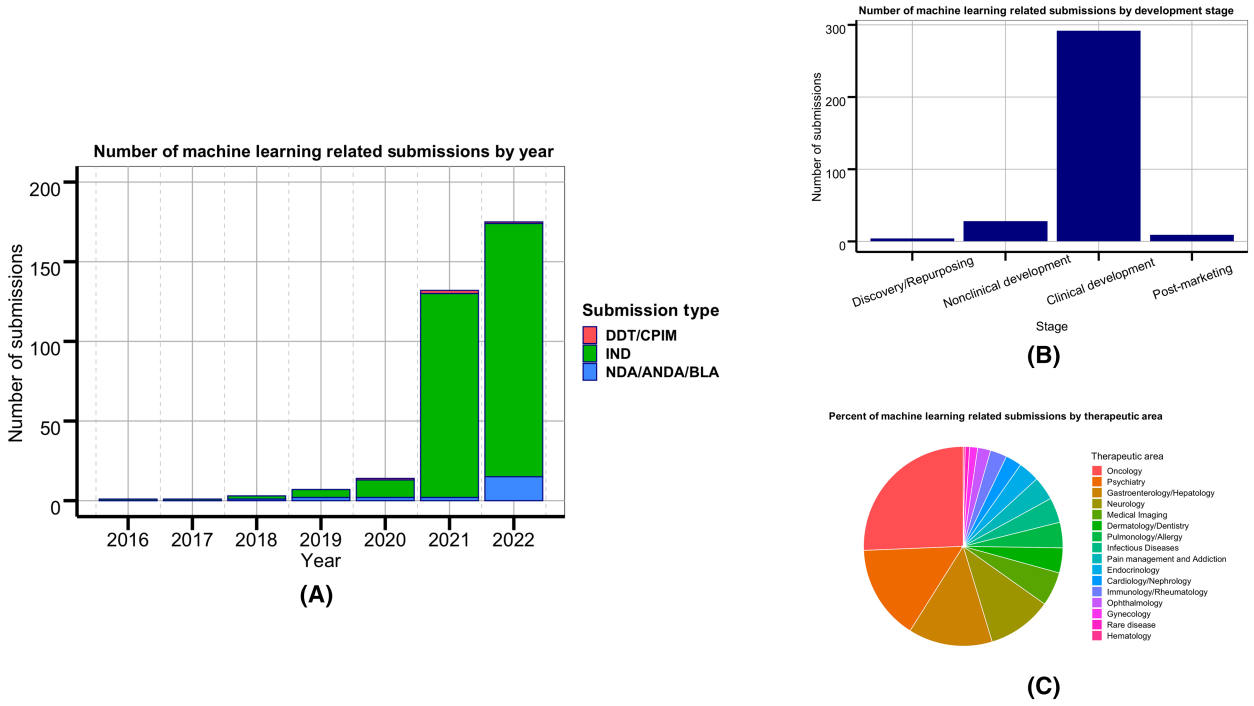

Figure 4 Overview of AI/ML Related Submissions for New Drug Development to the FDA. (**a**) The Number of Submissions by Year. (**b**) AI/ML Submissions by Development Stage. (**c**) AI/ML Submissions by Therapeutic Areas. AI, artificial intelligence; ANDA, abbreviated new drug application; BLA, biologics license application; CPIM, Certified in Production and Inventory Management; DDT, drug development tool; FDA, US Food and Drug Administration; IND, investigational new drug; ML, machine learning; NDA, new drug application.

assist drug development across different therapeutic areas. About a quarter of submissions is in oncology-related fields. Psychiatry, neurology, gastroenterology, and hepatology, and medical imaging are the other therapeutic areas where submissions are commonly seen (**Figure 4c**).

There is a broad usage of AI and ML approaches in clinical drug development, as reflected in the submissions. For example, some drug developers propose AI and ML modeling to enrich patients, to stratify patients with different safety risk factors, to select or optimize dosing, or to ensure adherence to the assigned dosing regimen. In recent years, some ML algorithms have been used to select end points, identify biomarkers, or synthesize control groups. To meet these objectives, various modeling approaches have been utilized for covariate selection, anomaly detection, image/video/audio assessment, and real-world data phenotyping. With new tools and methodologies available, AI and ML are expected to improve the efficiency of drug development and enhance patient care. By peering into transformative and specific applications of AI/ML in precision medicine using compelling case studies (**Table 1**), we can gain insights into how precision medicine is adapting to accommodate the surge of technological transformation.

Examples

AI-partnered dynamical model discovery for precision medicine. Traditional approaches to modeling patients' treatment data rely upon the development of empirical or mechanistic models that leverage existing data and/or capture known interactions between disease pathophysiology and drug effects. However, the construction of accurate dynamic models to facilitate precision medicine have proven to be a formidable challenge, primarily due to the complexity of biological systems and the paucity of data. The recent emergence of collections of large scale, high content data have opened new possibilities. This growth of data has occurred along the following three dimensions: (1) the number of patients (e.g., from real-world data sources[23]; (2) the dimensionality of data (e.g., from -omics[8]); and (3) as the number of measurements (e.g., from liquid biopsies[24] and wearable devices[1]). The daunting task of making sense of this deluge of data can easily overwhelm the capabilities of human intellect alone.

One effective approach to overcome the analytical bottleneck is to leverage the AI-partnered dynamical model discovery approach, whereby the computational strength of AI/ML algorithms is combined with the pharmacological domain knowledge of the human modelers.[25] As an example, a neural-pharmacokinetic/pharmacodynamic (PK/PD) model that expresses the causality among dose, PK, and PD has been developed.[25] In the setting of individualized predictions of patients' platelet count from early data, it has been demonstrated that the proposed DL approach can outperform the existing population-PK/PD model on certain metrics of predictive performance. Furthermore, the importance of using a pharmacology-informed neural network architecture for PK predictions to unseen dosing regimens has been demonstrated in the setting of trastuzumab emtansine (T-DM1).[26] The enhanced accuracy provided by AI models can unlock new opportunities for precision medicine.

Oncology is a particularly promising domain for partnering with AI in data-driven dynamic model discovery. Due to the

Table 1 Case studies of AI/ML application in precision medicine

Precision medicine application	Application description	AI/ML method(s) used
Longitudinal patient response predictions to dose-optimize trastuzumab emtansine[25]	Neural-PK and PD model which expresses causality between dose, PK and PD was implemented. In the setting of individualized predictions of patients' platelet count from early data, it has been demonstrated that the proposed approach can outperform the existing population-PK/PD model on certain metrics of predictive performance.	Pharmacology-informed recurrent neural networks
Prediction of OS using longitudinal tumor data[27]	Explainable DL was developed for tumor growth prediction, discover novel TGD metrics, perform unbiased individual-level predictions of OS using tumor data with short follow-up. GitHub link – https://github.com/jameslu01/TDNODE	Pharmacology-informed, encoder-decoder neural network, XGBoost and SHapley additive explanations
Identification of prognostic factors of long-term OS in gastric cancer[30]	Prognostic factors of long-term OS and TGD were identified using ML with potential to inform future trial design in gastric cancer	Random survival forests, SIDEScreen method
Early prediction of disease activity in multiple sclerosis patients[33]	An ML model was developed to predict future disease activity in patients with multiple sclerosis to facilitate early detection of disease onset, improved treatment planning, and augment understanding of the illness	XGBoost with SHapley additive explanations
Prognostic and predictive factors identification of TGD in advanced melanoma patients[31]	Potential prognostic and predictive biomarkers of TGD in advanced melanoma patients were identified using high-dimensional AI/ML covariate screening approaches applied on tabular and imaging data (radiomics features obtained from positron emission tomography/computed tomography images)	Random regression forest with several variable importance metrics for robust feature selection
Patient population identification in clinical practice of COVID-19 pneumonia[37]	A scoring system was developed to identify patients with plasma SuPAR level no less than 6 ng/mL (no commercially available assays) for administration of Anakinra. The developed scoring system is part of the final fact sheet for Anakinra administration.	Elastic-net regression and neural networks

Abbreviations: AI, artificial intelligence; COVID-19, coronavirus disease 2019; DL, deep learning; ML, machine learning; OS, overall survival; PD, pharmacodynamic; PK, pharmacokinetic; SuPAR, soluble urokinase plasminogen activator receptor; TDNODE, Tumor Dynamics Neural-ODE; TGD, tumor growth dynamics.

complexity and heterogeneity of tumors, identifying the right anticancer therapies and adjusting treatments based on a patient's response are highly challenging tasks. As highlighted by Acosta *et al.*,[2] there is a need for AI in integrating data to capture dynamic and real-time information to support precision medicine approaches. To this end, a pharmacology-informed, encoder-decoder neural network architecture of Tumor Dynamics Neural-ODE (TDNODE) has been proposed.[27] This formulation generates patient embeddings from longitudinal tumor data at the individual level, which can then be used to predict the patients' overall survival (OS). The model has demonstrated the ability to overcome a significant limitation found in existing tumor dynamic models, specifically the issue of prediction bias when utilizing tumor data with short follow-up periods.[28] Furthermore, the methodology has been shown to significantly increase the accuracy of predicting OS at the individual patient level, hence with the potential for enabling precision medicine applications.[29]

As technology further advances and more data become available, AI-partnered dynamic model discovery is likely to emerge as a crucial way to surmount the challenge of data deluge and help us continue to push the boundaries of precision medicine and lead to improved outcomes.

AI/ML-enabled drug-disease modeling. A striking demonstration of the value of AI/ML-enabled drug-disease modeling lies in its applications to advance understanding of disease and drug mechanism of action in a totality of evidence mindset. For instance, ML methods were used to evaluate prognostic and predictive factors governing long-term survival and tumor growth dynamics in the Avelumab JAVELIN Gastric trial.[30] Through a systematic analysis of a vast array of covariates, AI/ML techniques provided invaluable insights into response outcomes and enabled an extensive investigation of patient subpopulations that could potentially derive substantial benefits from the treatment.

Furthermore, the synergy between ML and the integration of high-dimensional datasets presents an innovative avenue for leveraging novel biomarkers such as liquid biopsy circulating tumor DNA and image-based radiomics. For the latter, it first attempts to integrate radiomic features into models of tumor growth dynamics report the use of ML-based radiomics feature selection in the real-world setting.[31] Leveraging tumor phenotype features extracted from images into longitudinal disease models has a great potential to deepen the comprehension of tumor evolution and progression dynamics, thereby contributing to informed clinical decision making and to advance precision medicine. In fact, tumor heterogeneity is critical in cancer treatment due to drug resistance and failure in advanced tumors. Intra-tumor variation and inter-tumoral genetic differences shape disease progression. Standard assessment criteria often miss nuances in patients with different lesion dynamics. To address this, CICIL – short for ***Class****Ification* ***C****lustering of* ***I****ndividual* ***L****esions* – a novel methodology integrating signal processing and ML, was introduced to offer an ML-based metric of lesion dynamics similarity.[32] Applying CICIL to over 1,700 patients with metastatic colorectal cancer receiving cetuximab, yielded uncovered insights. Lesion dynamics differ across tissues, especially in KRAS-mutated patients. Notably, a multivariate Cox model highlighted the CICIL-based metric of tumor heterogeneity as a significant predictor of OS, revealing its potential. This is an example of how ML-derived metrics may not only advance current approaches but also incorporate a holistic view into predictive models.

AI/ML approaches have been further leveraged for precision medicine to enhance existing models and provide predictions of response dynamics and clinical outcomes in mechanism-agnostic manners. This predictive prowess was demonstrated through early disease activity prediction in patients with multiple sclerosis, achieved by integrating and assessing diverse data dimensions, such as demographics, response data, MRI scans, and neurological assessments from pivotal clinical trials of cladribine.[33] Integration of explainable AI like SHAPley measures enabled a thorough understanding of the model results and impact of factors towards predictions of disease activity.

The ability of AI and ML to navigate high-dimensional multimodal heterogenous datasets, recover nonlinear effects, and unveil complex variable interactions holds profound potential in reshaping drug discovery and therapeutic development. Moreover, the integration of AI/ML tools into drug-disease modeling processes opens new avenues for understanding novel biomarkers, digital biomarkers, real-world data, and translational safety, augmenting our understanding of the intricate interplay between drugs and diseases.

Accelerated drug development and precision pharmacotherapy using DeepNLME. DeepNLME is a seamless mix of dynamic, statistical, and ML modeling, the primary benefit of which is the ability to utilize disparate sources of information in the same model simultaneously. Known interactions in disease biology and medical interventions constitute a valuable source of information that can be encoded in differential equation-based dynamic models.[34,35] Mixed effects enable the specification of what biological properties are expected to be shared across a population of patients and what properties have between-patient variability. ML is a flexible way of identifying complex relationships from data, be it how a medical image might inform patient-specific model parameters or how one dynamic model variable affects the rate of change of another. DeepNLME enables explicit encoding of incomplete scientific knowledge, letting ML identify the missing pieces in a data-driven and individualizable way.

A recent study tested the effectiveness of DeepNLME for enabling individualized predictions of survival probabilities based on covariates and early tumor size biomarker data. In the model, the general structure of the problem was defined such that tumor size dynamics were split into one drug responsive and one drug-unresponsive compartment. The total tumor size then affected the hazard of death, enabling the modeling of tumor-size-dependent survival probability. Although the structure of the model was explicitly defined based on biological principles, the exact functional forms of how drug concentration and current tumor size affected tumor growth/shrinkage, how tumor size affected hazard, and how patient covariates could inform patient outcomes were left for neural networks to automatically discover based on a synthetic dataset that was used to train the model (200 patients). The trained DeepNLME model could accurately reproduce the longitudinal tumor sizes predicted by the data-generating model (best possible prediction) on unseen data under different scenarios of data availability (with or without covariates, with or without access to early tumor size measurements). Furthermore, the DeepNLME model could reproduce the probability of survival for individual patients that the data-generating model would predict under the same conditions of data availability. The resulting DeepNLME model could thus utilize not only baseline covariates but also all tumor size measurements until a given time to predict the survival probability over time and how that probability is affected by different chemotherapeutic dosing schemes.

The ability of a model to train its behavior based on data from previous patients yet individualize predictions of how a patient outcome would be affected by different treatment options based on all the covariate and longitudinal biomarker information that is available for the specific patient at the very day that a treatment decision must be made enables tailored treatment optimization. The ability to automatically identify complex relationships within the model and complex predictive factors not only facilitates more rapid modeling but also the discovery of more complex and predictive relationships than a manual trial-and-error approach to modeling might have identified. Although this study has yet to be published in full detail, ongoing work is already utilizing the same approach with real oncology data, as well as in different domains such as ophthalmology and epidemiology.

Application of artificial intelligence and machine learning in drug development and precision medicine. Anakinra is an interleukin-1 antagonist indicated for the treatment of rheumatoid arthritis, cryopyrin-associated periodic syndromes, and deficiency of interleukin-1 receptor antagonist in the United States.[36] A clinical trial was conducted to evaluate the efficacy and safety in patients with coronavirus disease 2019 (COVID-19) pneumonia who were at risk of developing severe respiratory failure. The COVID-19

pneumonia was radiologically confirmed by chest X-ray or CT scan. As part of the enrollment criteria, all enrolled patients were required to have a plasma soluble urokinase plasminogen activator receptor (SuPAR) level no less than 6 ng/mL. Based on the clinical trial results, the sponsor sought emergency use authorization.[37] It has been noted that the SuPAR assay is not commercially available for use in the United States. To align with the clinically tested patient population, an ML analysis was conducted with the objective to search for alternative method based on commonly measured patient characteristics to identify patients with SuPAR no less than 6 ng/mL.[38] The final identified patient identification method is a scoring system which comprises eight items, including age, severe pneumonia defined by World Health Organization (WHO) criteria, smoking status, sequential organ failure score, neutrophil-to-lymphocyte ratio, hemoglobin, medical history of ischemic stroke, blood urea, or medical history of renal disease. Patients with at least 3 out of the 8 items meeting established threshold are likely to be the patients with SuPAR no less than 6 ng/mL. The established scoring system demonstrated high positive predictive value, high specificity, and reasonable sensitivity in both the training and external validation dataset. The selected items are also biologically plausible. This score is recommended in the fact sheet of anakinra to guide patient population identification in clinical practice.[37]

From literature summary to regulatory landscape to concrete case studies, we have encapsulated the potential that AI/ML currently holds in transforming the field of precision medicine. The transformative promise, however, presents a gamut of challenges that merit careful consideration and require addressing of nuanced issues related to bias, generalizability, reliability, interpretability, and privacy. With the foundation firmly established, in the next section, we will navigate current challenges, and suggestions to address these challenges, and elucidate how AI's integration in precision medicine can be harnessed responsibly and ethically to revolutionize patient care.

CHALLENGES AND FUTURE DIRECTION

Challenges in AI/ML for health

The capabilities of AI systems are increasing by the day. Large language models (LLMs) are now able to confidently engage in any topic, create rhymes, debug code, and even pass the US medical licensing examination.[39] These tools are being rapidly adopted by the community. Take the example of ChatGPT, an AI Chatbot developed by OpenAI (https://chat.openai.com/). It reached one million users in just 5 days,[40] highlighting the transformative potential – and perils – of AI technologies. Nevertheless, these achievements of AI have also sparked critical debates around its accuracy, safety, privacy, reliability, and, most importantly, its trustworthiness. The various challenges brought about by AI in precision medicine and potential solutions are outlined in **Table 2**.

Trust is a critical yet elusive concept, particularly in the AI domain. As it happens among humans, trust requires time to build but can be lost instantly. For instance, ChatGPT is known for confidently providing inaccurate answers.[41] These occurrences, referred to as "hallucinations," are instances where the system confidently and erroneously produces outputs. These seemingly trivial yet crucial mistakes are an example of how the trustworthiness of AI systems can be easily undermined, casting a shadow over their potential benefits. Therefore, it is clear that establishing trust in AI necessitates the absence of such untrustworthy behaviors.

To address these challenges, it is essential to identify a comprehensive set of conditions deemed necessary for trustworthiness in AI systems. The National Institute of Standards and Technology (NIST) has made strides in this direction, outlining key dimensions of trustworthiness in their recently-released guideline.[42] These include validity, reliability, safety, security, explainability, interpretability, privacy, and fairness (**Figure 5**). These dimensions bridge the gap between scientific and engineering perspectives, highlighting the multidisciplinary nature of AI and the unique considerations each discipline brings to the table.

Table 2 Challenges and solutions for AI/ML systems in precision medicine

Challenges recognized	Proposed solutions
AI systems may exhibit untrustworthy behavior. Trustworthiness requires validity and reliability – reproducibility, generalizability, and adversarial resilience – accountability and transparency.	• Promote the use of AI Risk Management Framework[42] • Creation of societal level agreed upon criteria for trustworthiness • Implement model public auditability • Development of proofs and certificates for AI systems' design • Outline context of use of AI systems in depth • Standardization of protocols for scoring datasets • Incentivize model builders to demonstrate trustworthiness in AI systems
ML algorithms may manifest racial bias, economic bias, and biases related to cultures, identities, and values of humans labeling the datasets. This can result in biased and unfair behavior of AI systems towards certain patient populations.	• Encourage usage of the "ethical algorithm" to reduce bias in model predictions[51] • Use federated learning to ensure fairness
AI systems are susceptible to data extraction attacks or might inadvertently divulge confidential information. Ensuring privacy and safety of personal information and intellectual property is crucial.	• Use homomorphic encryption and multi-party computation to ensure algorithm privacy • Use federated learning and differential privacy to safeguard personal information[57] • Use synthetic data generation as a solution to protect patient privacy

Abbreviations: AI, artificial intelligence; ML, machine learning.

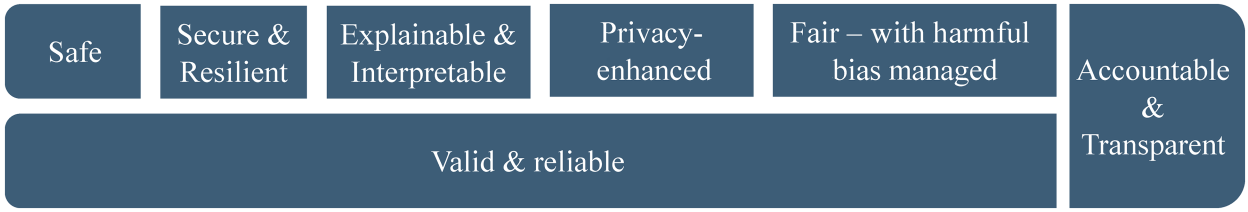

Figure 5 Characteristics of trustworthy AI systems. Valid & Reliable is a necessary condition of trustworthiness and is shown as the base for other trustworthiness characteristics. Accountable & Transparent is shown as a vertical box because it relates to all other characteristics.[42] AI, artificial intelligence.

Validity and reliability form the cornerstone of any empirical research. The findings should not only be reproducible but also capable of generalizing across similar application domains and remaining stable over time. However, an examination of over 300 papers in ML for health (ML4H) health applications revealed significant reproducibility issues.[43] These range from issues related to code and data availability, variances in model performance, and the utilization of multiple datasets for testing. These factors contribute significantly to the challenges of achieving reproducibility in AI research.

Reliability can be many things, one example for AI systems extends to its resilience to adversarial examples. There have been instances where ML systems could be fooled by strategically constructed inputs. For instance, stickers on stop signs can render them invisible to self-driving cars,[44] or for the medical domain specifically, the fact that addition of noise to ECG readings can alter the prediction of an AI classifier.[45]

Bias is another critical issue that can lead to irreproducibility and unfair outcomes across different subgroups of people. For example, a recent review of many ML models trained on Fitbit data to detect COVID showed widely different performances when tested on different datasets.[46] This discrepancy goes above and beyond model drifts phenomena that are already well studied in other fields of medical machine learning, such as EHR,[47] and suggest that different ML models have been trained with significantly differing definitions of "COVID cases," some more and some less severe stemming from the fact that each real-world dataset has its unique characteristics, introducing potential for bias, and consequently, irreproducibility.

Bias in AI for health applications is not a new phenomenon. There have been documented cases of algorithmic bias in health care, such as classifiers exhibiting unequal error rates across races in mortality prediction tasks,[48] and algorithms misidentifying historical health costs as a proxy for health needs, leading to unfair triaging of different populations.[49] Even algorithms measuring oxygen saturation have exhibited racial bias, leading to unfair treatment, demonstrating that bias can affect any algorithm, ML-based or not.[50] Therefore, ensuring trust in AI systems necessitates careful navigation and balance of these complex factors to guarantee reliability, fairness, and validity.

The last 5 years of research have indicated that it is possible to develop fair algorithms despite biased data[51]; however, it becomes increasingly difficult when the discriminated-against group is heavily under-represented or completely absent from the training set. This absence of diversity not only affects the generalizability of the research findings but also hampers efforts toward the creation of fair AI algorithms.

Bias can subtly infiltrate AI models and persistently affect their outputs. For instance, ChatGPT, based on GPT-3.5, incorporates a supervised and reinforcement learning fine-tuning step, which has improved the accuracy of the model.[52] However, this step introduces a layer of bias through the human labelers, or annotators, whose identities, cultures, beliefs, and values influence the responses provided by ChatGPT. This additional source of bias complicates efforts to mitigate the overall bias of the model, emphasizing the need for diversity among annotators, not only for text, but for medical applications alike.

Privacy is another major concern, particularly in the context of large language models that have been proven susceptible to extraction attacks. GPT-2, the predecessor of ChatGPT, has been shown to be amenable to being tricked into divulging confidential information from its training data.[53] The same risk applies to health settings, especially with the increasing use of wearable devices that collect sensitive health data. Such data are high-dimensional and even seemingly innocuous data, such as a series of step counts from a Fitbit device, could potentially be used as a key for extraction attacks if the AI systems are not trained in a privacy-protected manner.

Fortunately, there are ways to mitigate these issues. Models can be trained to generalize across datasets and time, maintain a certain level of privacy, and ensure fairness.[51] However, these mitigations often involve trade-offs among different factors, such as validity, fairness, and privacy. The NIST in their recently released framework[42] recommends a risk-based approach, which involves identifying the context of use of the AI system and striking an appropriate balance among the different factors (**Figure 6**).

Ultimately, establishing trust in AI cannot be proven exhaustively but can be described using a set of criteria agreed upon at the societal level. Full openness of the models may not be possible due to privacy, cost, or Intellectual Property (IP) reasons, but that does not rule out trade-offs such as model public auditability that it is crucial to ensure the trustworthiness of AI systems (https://www.medperf.org/). This process may involve creating proofs and certificates that attest to the involvement of end-users during the AI system's design. It also includes the development of standardized protocols for scoring datasets[54,55] for representativeness and describing the context of use and the suggested trade-offs around different trustworthiness criteria. These efforts require creating a system of incentives at the societal level

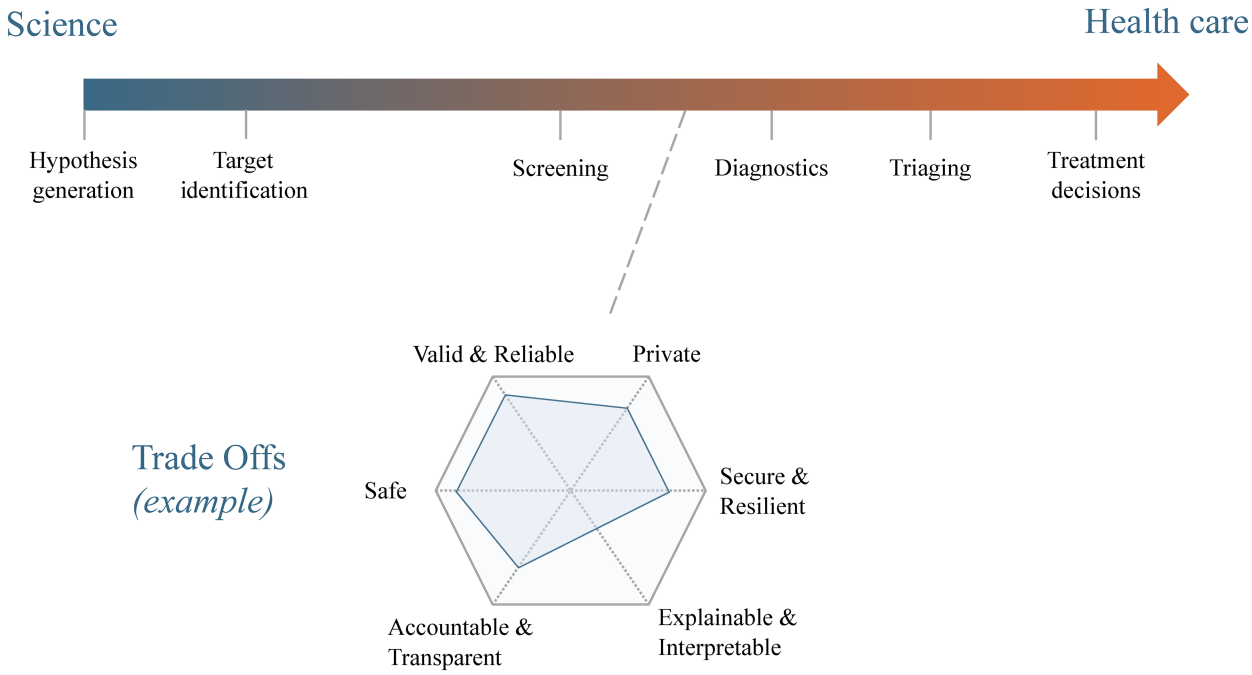

Figure 6 Depending on the chosen Context of Use (CoU), as represented as a point on the dimension from blue to red (science to health care applications) different tradeoffs between different dimensions of trustworthiness are necessary.

to encourage model builders to demonstrate the trustworthiness of their AI systems, even when the underlying data or algorithms cannot be openly shared.

Data sharing, distributed learning, and synthetic data generation

Despite a global surge in health data, access to health data remains a major challenge in AI/ML development for health. There are three reasons. First, health data are often decentralized and generated by many different actors, such as hospitals, companies, and devices.[56] Second, health data contain personal and sensitive information and are therefore highly regulated.[17] Consequently, managing and using medical data requires strict adherence to data protection guidelines and regulations, namely the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the United States. Third, because it is costly to generate health data, they are valuable and require high market prices to acquire. In addition, data providers (e.g., academics, startups, and pharmaceutical companies) must generate and share data in an ecosystem of competing interests.

To address these challenges, institutions seeking collaboration in the health sector need to identify an efficient, scalable, and secure approach to share their data. Conventionally, health data are shared through a *data pooling* approach. The data generated by multiple sources are aggregated and stored in a central location before analyses and modeling. However, this method requires high levels of computation power and storage space from a centralized provider to handle large data and is prone to data security risks.

Federated data access offers a novel way to break these data silos and enable collaborative ML efforts among different data sources (i.e., pharmaceutical companies, clinical researchers, manufacturers, and hospitals).[57] In contrast to the conventional data pooling approach, the federated approach trains models locally within each institution.[58] For instance, in forming a federated model, only the models trained in the institutions are shared and aggregated by a central organizing company. Over time, the federated model is then optimized by integrating the insights from all different sources without directly accessing individual patient data from the institutions. There are three advantages: first, it provides a secure, private path for diverse institutions to share data. Because the aggregating provider does not have direct access to individual patient data, it protects the patients' privacy and maintains the security of proprietary source information. Second, the data providers retain full ownership and control of their data. Finally, it enhances the efficiency and scalability in accessing large-volume data as no copies of private data are downloaded and centrally stored.

Three aspects must be considered to realize a secured and efficient collaborative ML: data privacy, algorithm privacy, and performance (or compute power). Several approaches can be taken to safeguard these three aspects. For example, "differential privacy" can be deployed to protect data privacy. This method allows masking and de-identification of the original data sources by imposing additional, structured noises to the data. "Homomorphic encryption and multi-party computation" can be applied to further secure algorithm privacy. Data are encrypted and can only be decoded with an access key possessed by a validated owner. Although this method enables proven

data and algorithm privacy, it requires massive computational resources in real-life scenarios. The third method to secure a high performance while protecting the data privacy is to apply "federated learning." Federated learning can also be reinforced with methods like "algorithm or model compilation, secure aggregation (MPC – HE), and trustless traceability."[17,59] These reinforcement methods ensure a comparable level of algorithm privacy to that of homomorphic and multi-party computation approaches, whereas achieving the high-performance levels of a standard federated learning approach.

In addition, federated learning possesses unique qualities that facilitates healthcare data sharing. First, it is fundamentally data agnostic, meaning that it can be applied to analyze a variety of data modalities. Second, it is framework and infrastructure agnostic, indicating that it can be applied to any kind of machine or technology analytics technology and can be adapted to different information technology infrastructures.

Consequently, federated learning has been increasingly applied to advance AI in medical research. By enabling pharmaceutical companies, hospitals, and academics to bring their data together securely, it accelerates the formation of new international collaboration and cooperative consortia on discovering complex disease mechanisms and biologies. It also opens up new opportunities for data exchange and aggregation, such as combining heterogeneous, multimodal data and ensuring a critical mass of patient data for rare disease R&D. Increasing evidence has also shown that it is a promising approach in advancing precision medicine by accelerating the identification of biomarkers, patient subtypes, and drug targets for complex diseases such as cancers.

Use of synthetic data is a lucrative strategy to preserve data privacy while improving the efficiency and innovation of using AI/ML to advance precision medicine and medical research.[60,61] To protect patient privacy, Giuffre and Shung have discussed the utilization of practical differential privacy methods to help trade-off privacy and utility, thereby increasing practical applications of AI/ML-based synthetic data generation (SDG) tools.[60] Although data synthesis is not new, there has been significant advancements in SDG methods. Traditionally, stochastic Monte-Carlo simulations on differential equations of dynamic systems or statistical models have been used for SDG. In the past decade, scalable and generalizable AI/ML-based SDG methods, such as generative adversarial neural networks, variational auto-encoders, and auto-regressive models were developed and have been successfully exploited for the recent AI revolution of large-language models.[62–64]

Use of synthetic data can not only help preserve patient privacy but also help generate hypotheses in the process of obtaining real datasets, augment real datasets (partial synthetic data), ease sharing data to verify analyses and improve reproducibility, and pre-train models to be used for application in specific populations.[65] AI/ML models for SDG have shown to emulate real data characteristics in various therapeutic areas, including but not limited to hematology, oncology, infectious diseases, medical imaging, and endocrinology for lucrative applications, such as estimation of treatment effect and survival, use as a proxy for clinical trial datasets to perform secondary analyses, generate large datasets for development of image segmentation models, predict future patient outcomes (e.g., glycemic change), and data augmentation to develop disease diagnosis models.[66–70] Additionally, open source tools, such as Synthea, have been developed for generating EHRs data of patient disease progression and clinical workflow.[71] These examples and open-source software highlight the utility of synthetic data while maintaining privacy of patient data.

Artificial data, however, are not immune to pitfalls. These data can perpetuate and/or accentuate biases underlying the original data used to create the data generation model, might lack interpretability due to the black-box nature of underlying algorithms leading to lack of trust in using for real applications, might reveal confidential information in an adversarial attack, and lack consensus on evaluation of data quality.[60,61,72] These challenges are being carefully addressed by development of regulatory policies around using these data for improving patient outcomes.[60]

Regulatory considerations (with a focus on fit-for-purpose risk-based framework) for the use of AI/ML for precision medicine

The diverse uses of AI in drug development highlight the need for a careful regulatory assessment of both benefits and risks and underscores the importance of adopting a risk-based management approach that is proportional with measures commensurate with the level of risk posed by the specific context of use. For any specific AI application in drug development, model risk calculations will be determined by model influence and the decision consequence based on the context of use. For example, high-risk models may require more evidence of credibility than low-risk models, and the regulatory approach may differ accordingly.

As with any innovation, AI and ML creates opportunities, and new and unique challenges. To meet these challenges, the FDA has accelerated its efforts to create an agile regulatory ecosystem that can facilitate innovation and adoption while safeguarding public health. As the FDA continues to refine the regulatory approaches around the use of AI to facilitate the generation of reliable evidence and to support decision making, the evidentiary standards needed to support drug approvals remain the same regardless of the technological advances involved. AI and ML will undoubtedly play a critical role in drug development, and the FDA remains committed to robust policy development that both protects and promotes public health.

CALL FOR ACTIONS

Sharing of data, algorithms, and experiences among the community

The call to action to make data collected with National Institutes of Health (NIH) grants available to the community has finally become a funded mandate, with the NIH Data management policy coming into effect January 25, 2023, and allowing researchers to write in their grants the cost for long-term data sharing (https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html). But sharing does not necessarily mean re-use and therefore reproducibility; many datasets, especially in the medical world, have been recently deemed "Open In Appearance Only."[73] Furthermore, even when data are successfully re-used, reproducibility of analysis and findings are still not guaranteed if

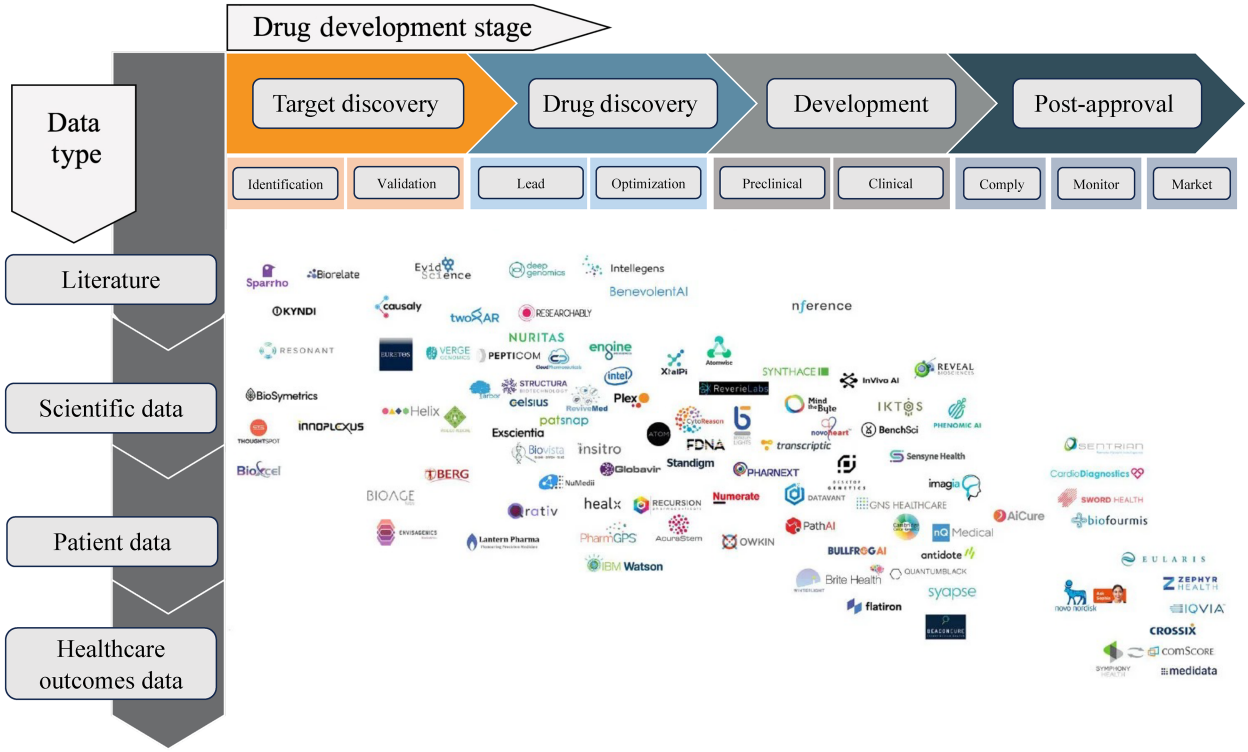

Figure 7 AI companies come in many forms (image from LEK – https://www.lek.com/insights/ei/artificial-intelligence-life-sciences-formula-pharma-success-across-drug-lifecycle). AI, artificial intelligence.

code is not released alongside data.[43] For this reason, health analysis and algorithms could be made available to all potential future subjects of those algorithms' decisions (in the health world, that means all of us). Due to privacy, cost and IP implication this may not be always possible (especially when it comes to regulatory bodies) but it is worth pushing for at least some form of publicly auditability, when data/models may not be made publicly available directly, but performances and scores across domains of interest (e.g., representatives for datasets) are. This has been described as "federated benchmarking," and the recent release of MedPerf (medperf.org) by MLCommons provides a privacy-preserving open-source platform for benchmarking AI models to deliver clinical efficacy.[74]

Development of best practices and regulatory guidances

The FDA continues to engage sponsors, AI tool developers, data service providers, ethicists, academia, patients and patient groups, and other international regulatory authorities interested in AI and ML in drug and biologic development through discussion papers, white papers, workshops, and meetings, etc. These engagements serve to better understand any gaps in regulatory scientific methods that may need to be developed and aims to pave the way for meaningful regulatory clarity as these technologies continue to evolve.

Work with broad stakeholders

Precision medicine, is an important field with profound implications for drug development and health care, stands on the brink of a significant evolution with the advent of AI/ML. A considerable amount of activity has built up over the past decade applying AI toward all aspects of health care. A concise visualization this activity was presented by the consulting firm LEK in 2018 (**Figure 7**). Companies are shown organized by their activity in terms of primary data types used and the drug development stage their technologies are principally applied toward. The key point of this illustration is that there is a vast variety of AI-driven research in toward health care, and many companies are using diverse and complementary approaches. Clearly in this ecosystem, it is essential to find effective partnerships between academia, biotechnology companies, and big pharma. Thus, successful integration of AI/ML in precision medicine requires collaborations across a wide range of stakeholders, including but not limited to patient groups, academic institutions, non-profit organizations, industry, and regulatory bodies. Each stakeholder group possesses unique expertise and perspectives that are critical to the successful application of AI/ML in precision medicine. We envision this collaboration will involve knowledge sharing, joint problem-solving, and creation of standards and solutions. In this concerted effort, we need to ensure the applications of AI/ML in precision medicine is technologically sound, clinically relevant, ethically appropriate, and patient-centered. By fostering an ecosystem of collaboration, we can collectively address the challenges and unlock the potential these technologies hold for facilitating drug development and improving patient outcomes.

ACKNOWLEDGMENTS

The authors thank Jeffrey Florian for critical review of the manuscript. The authors thank Paul Schuette, Alan Edelman, Matthew Diamond,

and Elham Tabassi for their contributions as presenters and roundtable discussion panelists for the FDA/MCERSI February 2023 public workshop, "Application of Artificial Intelligence & Machine Learning for Precision Medicine."

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

L.F. is a paid employee of Sage Bionetworks. J.L. is a paid employee of Genentech Inc. J.L. and C.W.C. are paid employees of Owkin. N.T. is a paid employee of Ares Trading S.A. N.K. and J.G. are paid employees of Pumas AI Inc. As a Guest Editor of this themed issue of *Clinical Pharmacology & Therapeutics*, Q.L. was not involved in the review or decision process of this paper. All other authors declared no competing interests for this work.

DISCLAIMER

The contents of this article reflect the views of the authors and should not be construed to represent the FDA's views or policies. No official support or endorsement by the FDA is intended or should be inferred. The participation of these individuals or the materials should not be interpreted as representing the official viewpoint of the US Department of Health and Human Services or the National Institutes of Health, except where noted.

1. Smuck, M., Odonkor, C.A., Wilt, J.K., Schmidt, N. & Swiernik, M.A. The emerging clinical role of wearables: factors for successful implementation in healthcare. *NPJ Digit Med.* **4**, 45 (2021).
2. Acosta, J.N., Falcone, G.J., Rajpurkar, P. & Topol, E.J. Multimodal biomedical AI. *Nat. Med.* **28**, 1773–1784 (2022).
3. Terry, S.F. Obama's precision medicine initiative. *Genet. Test. Mol. Biomarkers* **19**, 113–114 (2015).
4. FDA-NIH Biomarker Working Group. *BEST (Biomarkers, EndpointS, and Other Tools) Resource* (Food and Drug Administration (US), Silver Spring (MD), 2016). http://www.ncbi.nlm.nih.gov/books/NBK326791/ Accessed August 13, 2023.
5. Torres, C. & Grippo, P.J. Pancreatic cancer subtypes: a roadmap for precision medicine. *Ann. Med.* **50**, 277–287 (2018).
6. Marra, A., Trapani, D., Viale, G., Criscitiello, C. & Curigliano, G. Practical classification of triple-negative breast cancer: intratumoral heterogeneity, mechanisms of drug resistance, and novel therapies. *NPJ Breast Cancer.* **6**, 54 (2020).
7. Garraway, L.A., Verweij, J. & Ballman, K.V. Precision oncology: an overview. *J. Clin. Oncol.* **31**, 1803–1805 (2013).
8. Weinstein, J.N. *et al*. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
9. Schreier, G. The internet of things for personalized health. *Stud. Health Technol. Inform.* **200**, 22–31 (2014).
10. U.S. Department of Health and Human Services. Artificial Intelligence (AI) Strategy <https://www.hhs.gov/sites/default/files/final-hhs-ai-strategy.pdf>. Accessed October 5, 2023.
11. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
12. Petersen, B.K. *et al*. Deep reinforcement learning and simulation as a path toward precision medicine. *J. Comput. Biol.* **26**, 597–604 (2019).
13. Meskó, B. & Görög, M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med.* **3**, 126 (2020).
14. Lehne, M., Sass, J., Essenwanger, A., Schepers, J. & Thun, S. Why digital medicine depends on interoperability. *NPJ Digit Med.* **2**, 79 (2019).
15. Kiernan, D. *et al*. Establishing a framework for privacy-preserving record linkage among electronic health record and administrative claims databases within PCORnet®, the National Patient-Centered Clinical Research Network. *BMC. Res. Notes* **15**, 337 (2022).
16. Baumfeld Andre, E. *et al*. The current landscape and emerging applications for real-world data in diagnostics and clinical decision support and its impact on regulatory decision making. *Clin. Pharmacol. Ther.* **112**, 1172–1182 (2022).
17. Rieke, N. *et al*. The future of digital health with federated learning. *NPJ Digit Med.* **3**, 119 (2020).
18. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
19. Sanchez, P., Voisey, J.P., Xia, T., Watson, H.I., O'Neil, A.Q. & Tsaftaris, S.A. Causal machine learning for healthcare and precision medicine. *R. Soc. Open Sci.* **9**, 220638 (2022).
20. *Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products* (United States Food and Drug Administration) (2023) <https://www.fda.gov/media/167973/download>. Accessed August 13, 2023.
21. Terranova, N. *et al*. Artificial intelligence for quantitative modeling in drug discovery and development: An innovation and quality consortium perspective on use cases and best practices. *Clin Pharma Therap.* **15**, 658–672 (2024).
22. Liu, Q. *et al*. Landscape analysis of the application of artificial intelligence and machine learning in regulatory submissions for drug development from 2016 to 2021. *Clin Pharma Therap.* **113**, 771–774 (2023).
23. Liu, R. *et al*. Systematic pan-cancer analysis of mutation–treatment interactions using large real-world clinicogenomics data. *Nat. Med.* **28**, 1656–1661 (2022).
24. Assaf, Z.J.F. *et al*. A longitudinal circulating tumor DNA-based model associated with survival in metastatic non-small-cell lung cancer. *Nat. Med.* **29**, 859–868 (2023).
25. Lu, J., Bender, B., Jin, J.Y. & Guan, Y. Deep learning prediction of patient response time course from early data via neural-pharmacokinetic/pharmacodynamic modelling. *Nat Mach Intell.* **3**, 696–704 (2021).
26. Lu, J., Deng, K., Zhang, X., Liu, G. & Guan, Y. Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience.* **24**, 102804 (2021).
27. Laurie, M. & Lu, J. Explainable deep learning for tumor dynamic modeling and overall survival prediction using neural-ODE (2023) https://doi.org/10.48550/ARXIV.2308.01362.
28. Netterberg, I. *et al*. Tumor time-course predicts overall survival in non-small cell lung cancer patients treated with Atezolizumab: dependency on follow-up time. *CPT Pharmacometrics Syst. Pharmacol.* **9**, 115–123 (2020).
29. Sundrani, S. & Lu, J. Computing the Hazard ratios associated with explanatory variables using machine learning models of survival data. *JCO Clinical Cancer Informatics.* **5**, 364–378 (2021).
30. Terranova, N. *et al*. Pharmacometric modeling and machine learning analyses of prognostic and predictive factors in the JAVELIN gastric 100 phase III trial of avelumab. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 333–347 (2022).
31. Courlet, P. *et al*. Modeling tumor size dynamics based on real-world electronic health records and image data in advanced melanoma patients receiving immunotherapy. *CPT Pharmacometrics Syst. Pharmacol.* **12**, 1170–1181 (2023).
32. Terranova, N., Girard, P., Ioannou, K., Klinkhardt, U. & Munafo, A. Assessing similarity among individual tumor size lesion dynamics: the CICIL methodology. *CPT Pharmacometrics Syst. Pharmacol.* **7**, 228–236 (2018).
33. Basu, S., Munafo, A., Ben-Amor, A.F., Roy, S., Girard, P. & Terranova, N. Predicting disease activity in patients with multiple sclerosis: An explainable machine-learning approach in the Mavenclad trials. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 843–853 (2022).
34. Brunton, S.L., Proctor, J.L. & Kutz, J.N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **113**, 3932–3937 (2016).

35. Rackauckas, C. *et al*. Universal differential equations for scientific machine learning (2020) https://doi.org/10.48550/ARXIV.2001.04385.
36. *US Package Insert of Kinert®* (Amgen Inc) <https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/103950s5189lbl.pdf>. Accessed May 10, 2023.
37. *US Fact Sheet of Kinert®* (Amgen Inc) <https://www.fda.gov/media/163075/download>. Accessed May 10, 2023.
38. Anakinra EUA Review <https://www.fda.gov/media/163546/download>. Accessed May 10, 2023.
39. Kung, T.H. *et al*. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. Dagan a, ed. *PLOS Digit Health* **2**, e0000198 (2023).
40. Hu, K. *ChatGPT sets record for fastest-growing User Base – analyst note* (Reuters, 2023) <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. Accessed October 5, 2023.
41. Goddard, J. Hallucinations in ChatGPT: a cautionary tale for biomedical researchers. *Am. J. Med.* **136**, 1060 (2023).
42. Tabassi, E. *AI Risk Management Framework: AI RMF (1.0)* (National Institute of Standards and Technology, Gaithersburg, MD, 2023).
43. McDermott, M.B.A., Wang, S., Marinsek, N., Ranganath, R., Foschini, L. & Ghassemi, M. Reproducibility in machine learning for health research: still a ways to go. *Sci. Transl. Med.* **13**, eabb1655 (2021).
44. Eykholt, K. *et al*. Robust physical-world attacks on deep learning models (2018) <http://arxiv.org/abs/1707.08945>. Accessed September 5, 2023.
45. Han, X., Hu, Y., Foschini, L., Chinitz, L., Jankelson, L. & Ranganath, R. Deep learning models for electrocardiograms are susceptible to adversarial attack. *Nat. Med.* **26**, 360–363 (2020).
46. Mitratza, M. *et al*. The performance of wearable sensors in the detection of SARS-CoV-2 infection: a systematic review. *Lancet Digital Health* **4**, e370–e383 (2022).
47. Parikh, R.B. *et al*. Performance drift in a mortality prediction algorithm among patients with cancer during the SARS-CoV-2 pandemic. *J. Am. Med. Inform. Assoc.* **30**, 348–354 (2023).
48. Chen, I., Johansson, F.D. & Sontag, D. Why is my classifier discriminatory? (2018) <http://arxiv.org/abs/1805.12002>. Accessed August 27, 2023.
49. 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).
50. Fawzy, A. *et al*. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. *JAMA Intern. Med.* **182**, 730–738 (2022).
51. Kearns, M. & Roth, A. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (Oxford University Press, New York, 2020).
52. Ouyang, L. *et al*. Training language models to follow instructions with human feedback (2022) https://doi.org/10.48550/ARXIV.2203.02155.
53. Carlini, N. *et al*. Extracting training data from large language models (2020) https://doi.org/10.48550/ARXIV.2012.07805.
54. Gebru, T. *et al*. Datasheets for datasets (2018) https://doi.org/10.48550/ARXIV.1803.09010.
55. Holland, S., Hosny, A., Newman, S., Joseph, J. & Chmielinski, K. The dataset nutrition label: a framework to drive higher data quality standards (2018) https://doi.org/10.48550/ARXIV.1805.03677.
56. Nguyen, T.V. *et al*. A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. *Sci. Rep.* **12**, 8888 (2022).
57. Sadilek, A. *et al*. Privacy-first health research with federated learning. *NPJ Digit Med.* **4**, 132 (2021).
58. Ogier du Terrail, J. *et al*. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat. Med.* **29**, 135–146 (2023).
59. Jordan, S., Fontaine, C. & Hendricks-Sturrup, R. Selecting privacy-enhancing Technologies for Managing Health Data use. *Front. Public Health* **10**, 814163 (2022).
60. Giuffrè, M. & Shung, D.L. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *NPJ Digit Med.* **6**, 186 (2023).
61. Tucker, A., Wang, Z., Rotalinti, Y. & Myles, P. Generating high-fidelity synthetic patient data for assessing machine learning healthcare software. *NPJ Digit Med.* **3**, 147 (2020).
62. Goodfellow, I. *et al*. Generative adversarial networks (2014) https://doi.org/10.48550/ARXIV.1406.2661.
63. Kingma, D.P. & Welling, M. Auto-encoding variational bayes (2013) https://doi.org/10.48550/ARXIV.1312.6114.
64. Gregor, K., Danihelka, I., Mnih, A., Blundell, C. & Wierstra, D. Deep AutoRegressive Networks (2013) https://doi.org/10.48550/ARXIV.1310.8499.
65. Gonzales, A., Guruswamy, G. & Smith, S.R. Synthetic data in health care: a narrative review. *PLOS Digit Health.* **2**, e0000082 (2023).
66. D'amico, S. *et al*. Synthetic data generation by artificial intelligence to accelerate research and precision medicine in hematology. *JCO Clinical Cancer Informatics* **7**, e2300021 (2023).
67. Azizi, Z., Zheng, C., Mosquera, L., Pilote, L. & El Emam, K. Can synthetic data be a proxy for real clinical trial data? A validation study. *BMJ Open* **11**, e043497 (2021).
68. Cronin, N.J., Finni, T. & Seynnes, O. Using deep learning to generate synthetic B-mode musculoskeletal ultrasound images. *Comput. Methods Prog. Biomed.* **196**, 105583 (2020).
69. Ngufor, C., Van Houten, H., Caffo, B.S., Shah, N.D. & McCoy, R.G. Mixed effect machine learning: a framework for predicting longitudinal change in hemoglobin A1c. *J. Biomed. Inform.* **89**, 56–67 (2019).
70. Karbhari, Y., Basu, A., Geem, Z.W., Han, G.T. & Sarkar, R. Generation of synthetic chest X-ray images and detection of COVID-19: a deep learning based approach. *Diagnostics* **11**, 895 (2021).
71. Walonoski, J. *et al*. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J. Am. Med. Inform. Assoc.* **25**, 230–238 (2018).
72. Kokosi, T. & Harron, K. Synthetic data in medical research. *BMJ* **1**, e000167 (2022).
73. Watson, H. *et al*. Delivering on NIH data sharing requirements: avoiding open data in appearance only. *BMJ Health Care Inform.* **30**, e100771 (2023).
74. Karargyris, A. *et al*. Federated benchmarking of medical artificial intelligence with MedPerf. *Nat Mach Intell.* **5**, 799–810 (2023).