Artificial Intelligence for Quantitative Modeling in Drug Discovery and Development: An Innovation and Quality Consortium Perspective on Use Cases and Best Practices

Nadia Terranova[1,†], Didier Renard[2,†], Mohamed H. Shahin[3,†], Sujatha Menon[3,†], Youfang Cao[4,†], Cornelis E.C.A. Hop[5,†], Sean Hayes[6], Kumpal Madrasi[7], Sven Stodtmann[8], Thomas Tensfeldt[3], Pavan Vaddady[9], Nicholas Ellinwood[10,*,†] and James Lu[11,*,†]

Recent breakthroughs in artificial intelligence (AI) and machine learning (ML) have ushered in a new era of possibilities across various scientific domains. One area where these advancements hold significant promise is model-informed drug discovery and development (MID3). To foster a wider adoption and acceptance of these advanced algorithms, the Innovation and Quality (IQ) Consortium initiated the AI/ML working group in 2021 with the aim of promoting their acceptance among the broader scientific community as well as by regulatory agencies. By drawing insights from workshops organized by the working group and attended by key stakeholders across the biopharma industry, academia, and regulatory agencies, this white paper provides a perspective from the IQ Consortium. The range of applications covered in this white paper encompass the following thematic topics: (i) AI/ML-enabled Analytics for Pharmacometrics and Quantitative Systems Pharmacology (QSP) Workflows; (ii) Explainable Artificial Intelligence and its Applications in Disease Progression Modeling; (iii) Natural Language Processing (NLP) in Quantitative Pharmacology Modeling; and (iv) AI/ML Utilization in Drug Discovery. Additionally, the paper offers a set of best practices to ensure an effective and responsible use of AI, including considering the context of use, explainability and generalizability of models, and having human-in-the-loop. We believe that embracing the transformative power of AI in quantitative modeling while adopting a set of good practices can unlock new opportunities for innovation, increase efficiency, and ultimately bring benefits to patients.

In recent decades, the utilization of quantitative modeling based on diverse computational approaches has gained significant prominence in the field of drug discovery and development. These approaches encompass a range of methodologies, including *in silico* prediction of absorption, distribution, metabolism, and excretion (ADME) properties, as well as model-informed drug discovery and development (MID3)[1] strategies. MID3 includes providing quantitative predictions for aspects such as pharmacokinetics, pharmacodynamics, efficacy and safety end points, and disease progression. By leveraging these models, researchers can optimize dosing strategies, inform clinical trial designs, and obtain robust quantitative assessments regarding drug efficacy and safety. This paradigm shift toward quantitative modeling has substantially enhanced the decision-making process, driving more efficient and effective drug development processes. Although the modeling of data has been the exclusive domain of human intelligence, in recent years, the fields of artificial intelligence (AI) and machine learning (ML) have made significant methodological advancements[2–4] such that these algorithms have the capability to automatically generate predictive computational models, and are being increasingly utilized to support drug discovery and development.[5] For instance, pharmacometrics (PMx) as a discipline has matured over the past 4 decades,[6] and standard pharmacokinetic (PK), PK/pharmacodynamic (PK/PD), and exposure-response analyses have largely become standardized with well-established quality and reporting guidelines.[7–9] However, with multidimensional, multisource, and multimodal data being generated,[10] there is a need to embrace modern predictive and computationally powerful analytics as a next step in the evolution of PMx methodologies.[6]

Just as the significant increase in the adoption of Quantitative Systems Pharmacology (QSP) in regulatory submissions since 2013[11] prompted a consortium collaboration to evaluate the state

[1]Quantitative Pharmacology, Merck KGaA, Lausanne, Switzerland; [2]Full Development Pharmacometrics, Novartis Pharma AG, Basel, Switzerland; [3]Clinical Pharmacology, Pfizer Inc., Groton, Connecticut, USA; [4]Clinical Pharmacology and Translational Medicine, Eisai Inc., Nutley, New Jersey, USA; [5]DMPK, Genentech Inc., South San Francisco, California, USA; [6]Quantitative Pharmacology & Pharmacometrics, Merck & Co. Inc., Rahway, New Jersey, USA; [7]Modeling & Simulation, Sanofi, Bridgewater, New Jersey, USA; [8]Pharmacometrics, AbbVie Deutschland GmbH & Co. KG, Ludwigshafen, Germany; [9]Quantitative Clinical Pharmacology, Daiichi Sankyo, Inc., Basking Ridge, New Jersey, USA; [10]Global PK/PD & Pharmacometrics, Eli Lilly, Indianapolis, Indiana, USA; [11]Clinical Pharmacology, Genentech Inc., South San Francisco, California, USA. *Correspondence: Nicholas Ellinwood (n.ellinwood@lilly.com) and James Lu (lu.james@gene.com)

[†]These authors contributed equally to this work.

Received June 28, 2023; accepted September 11, 2023. doi:10.1002/cpt.3053

of model assessment in the pharmaceutical/biotech industry and provide recommendations for its advancement,[12] we have recognized a similar emerging need within the scientific communities for AI/ML. There is a growing demand to emphasize the potential of AI/ML in the realm of drug discovery and development, whereas also acknowledging and addressing the associated challenges. Additionally, there is a strong impetus to establish a set of best practices tailored specifically to the unique characteristics of the healthcare domain,[13] encompassing relevant use cases and scenarios.

Hence, the Innovation and Quality (IQ) working group in AI/ML was formed in May 2021 within the International Consortium for Innovation and Quality in Pharmaceutical Development (also known as the IQ Consortium). The IQ Consortium[14] is a not-for-profit organization of pharmaceutical and biotechnology companies with the mission of advancing science and technology to augment the capability of member companies to develop transformational solutions that benefit patients, regulators, and the broader research and development community. The IQ AI/ML working group is an expert-based collaborative group consisting of 14 IQ member companies,[15] with the following 3 main aims:

- Identify appropriate AI/ML model assessment methodologies and promote the adoption of good practices for scientifically sound and robust adoption of these approaches within drug discovery and development.
- Identify opportunities for collaborative efforts leveraging AI/ML techniques for creation and mining of "big data."
- Increase awareness of AI/ML applications within the pharmaceutical industry, academia, and regulatory agencies to foster their fit-for-purpose use.

Upon its inception, the IQ AI/ML working group conducted a survey in 2021 to prepare for a US Food and Drug Administration (FDA) scientific exchange on AI/ML use in the field of clinical pharmacology. Members were asked to give a high-level overview of AI/ML activities at their respective companies, including how AI/ML was used for internal decision making or regulatory interactions, how AI/ML can improve upon existing methodologies, and the potential challenges for deploying AI/ML more broadly. The results of this survey were shared with the FDA in October of 2021, and are summarized in **Table 1**, including a list of key references[5,16–28] identified by the working group.

Although there is a wide array of ML algorithms and technical concepts of relevance to drug discovery and development, we refer the reader to the primer.[29] Additionally, several review papers have discussed opportunities and challenges of applying and integrating AI/ML with the related disciplines of PMx and QSP.[2,30,31] In this white paper, we outline and discuss the methodologies and applications that were covered in an IQ virtual workshop on Machine Intelligence for Quantitative Modeling in Drug Discovery and Development Applications,[32] as well as an American Conference on Pharmacometrics on Explainable Machine Learning for Disease Progression Modeling and Digital Twins, in November 2022.[33]

By providing an overview of a broad spectrum of concrete applications within the biopharmaceutical industry, along with discussions on identified opportunities and best practices, this white paper serves both as a summary of ongoing efforts, as well as to stimulate new applications and advancements in quantitative modeling and analytics across the pharmaceutical domain. As illustrated in **Figure 1**, this paper is divided into 4 main parts, covering the following thematic areas: (i) AI/ML-enabled analytics and pharmacometrics workflows; (ii) explainable artificial intelligence and its application in disease progression modeling; (iii) Natural Language Processing (NLP) in Quantitative Pharmacology Modeling; and (iv) AI/ML utilization in drug discovery. Finally, we end with overall conclusions and a set of recommendations.

PART 1: AI/ML-ENABLED ANALYTICS FOR PHARMACOMETRICS AND QSP WORKFLOWS

The integration of AI/ML into MID3 approaches like PMx and QSP provides new opportunities for addressing quantitative and clinical pharmacology questions, in a totality of evidence mindset leveraging large and diverse data along with new computational resources.[30,34] **Figure 2** illustrates how AI/ML can help advance MID3.[19] At level 1, the machine is used as a tool to save time and labor; it aids carrying out well-defined tasks more efficiently. At level 2, the machine serves as an assistant to raise the bar of what conventional models can deal with, thus enabling improvement of models compared with human intellect alone. At level 3, the machine is a partner and acts as an innovator pushing at the boundaries of what is possible to model. In this section, we review AI/ML-enabled MID3 analytics and workflows with examples across these levels.

AI/ML-enabled PMx modeling

The development of PMx models is a step-by-step process towards a model that is fit-for-purpose. This process generally involves many repetitive and time-consuming tasks, for instance, identifying the structural and statistical components of the model. Sibieude *et al.*[17] explored how model selection could benefit from AI/ML and compared a hybrid genetic algorithm (GA) and artificial neural network (NN) models for classification or regression in different scenarios based on simulated PK data. The NN classification model achieved the most accurate results and the GA was also successful at selecting plausible models. For the latter, the importance of appropriately defining the fitness function for optimal model selection has emerged, as well as opportunities to explore combinations of key metrics for model evaluation beyond the objective function, usually considered by a modeler during the model building steps (e.g., the number of parameters, high parameter correlation values, failed convergence, missing covariance step, and shrinkage). Computational gains were substantial, especially for NN models which, however, suffered by overfitting in certain scenarios due to the limited training set used in the study. This work aimed at first establishing a proof-of-concept that ML could be used for fast initial selection of models, followed by conventional PMx modeling for a more efficient workflow. Other works demonstrating the benefit of AI/ML use for PMx model selection have been reported in literature.[35]

Another important step in PMx model development is covariate selection, where standard methods are not well-suited to handle high-dimensional datasets. Sibieude *et al.*[16] compared classical

Table 1 Summary of 2021 AI/ML working group survey results

Purpose category	AI/ML Methodology	Added value	Challenges	Key References
Automation of PK/PD modeling	• Tree-based models. • DL language models. • Neural-ODEs.	• Increase automation of PK/PD modeling including selection of model structure and covariates, and analysis report drafting. • Save time and effort by extracting PK and DDI data from publications and reports using NLP.	• Validation of methodology. • Acceptance by scientific community and regulatory agencies.	[5,16–20]
Precision medicine and optimizing treatment regimens	• Tree-based models.	• Enable the utilization of high dimensional, complex data to identify key biomarkers, covariates, optimal treatment regimens, and patient subgroups.	• Limited patient population and insufficient high-quality data. • Difficult to generalize to different patient populations. • Impact of unmeasured confounders. • Acceptance by the scientific community and regulatory agencies.	[21,22]
Disease progression modeling and digital twins	• Tree-based models. • DL. • Neural-ODEs.	• Enable improved precision and the utilization of high-dimensional, complex data.	• Explainability of complex models. • Generalizability. • Acceptance by scientific community and regulatory agencies.	[20,23–26]
Causal inference	• Tree-based models. • SHAP analysis. • Causal forest. • Neural networks.	• Adjust for confounders that may affect dose/exposure-response relationships in complex, nonlinear manners	• Validation of methodology	[27,28]

AI, artificial intelligence; DDI, drug-drug interaction; DL, deep learning; ML, machine learning; NLP, Natural Language Processing; ODE, ordinary differential equation; PK/PD, pharmacokinetic/pharmacodynamic; SHAP, SHapley Additive exPlanations.

methods, such as stepwise covariate modeling (SCM) and conditional sampling for stepwise approach based on correlation tests (COSSAC), with ML methods (including Random Forest (RF), NN, and Support Vector Regression). Different scenarios of covariate influence were tested based on simulated PK data. Overall, ML performed similarly to, or better than SCM and COSSAC, and covariate effect size was the factor that had the most impact on the method performance. Significant differences were also found in computational speed, with ML being 30–100 times faster and able to provide results in a few minutes or hours, depending on the complexity of the explored scenarios. Hence, in this context, ML could be useful to provide fast initial screening of high dimensional covariates sets, followed by conventional approaches to assess clinical relevance of selected covariates and develop the final model.

As an example, this ML-PMx setting was adopted in the assessment of prognostic and predictive factors of long-term overall survival (OS) and tumor growth dynamics (TGDs) for the Javelin Gastric 100 phase III trial of avelumab.[36] In this analysis, RF and SIDEScreen were used to assess baseline and time-varying prognostic and predictive factors for OS (89 covariates) and TGD (52 covariates). Variable importance was assessed based on Boruta, permutation, random splits, and Shapley, and effectively informed the integration of relevant baseline and time-varying factors into PMx models for OS and TGD. Another recently published practical use case leverages ML methods to assess high dimensional images-derived radiomics features for integration into modeling of real-world tumor dynamics in patients with melanoma.[37,38]

These works demonstrate the successful use of AI/ML methods to address PMx model selection and covariate assessment. Furthermore, ML workflows are flexible enough to combine and handle these model building steps together: for instance, a GA implemented in an R-based NONMEM workbench for identification of near optimal models has been recently made available to the scientific community.[39]

Deep learning-enabled PK/PD modeling

The mainstay of modeling activities for drug development includes empirical compartmental models built from sparsely sampled PK/PD datasets. In this respect, AI/ML provides new ways for pharmacometricians to think about their models. There have been a number of approaches proposed in using feed-forward NNs[40–43] for modeling of PK(/PD) data. However, these did not tackle the more complex problem of extrapolating outside the range of observed data. In fact, the main limitation of such models is that

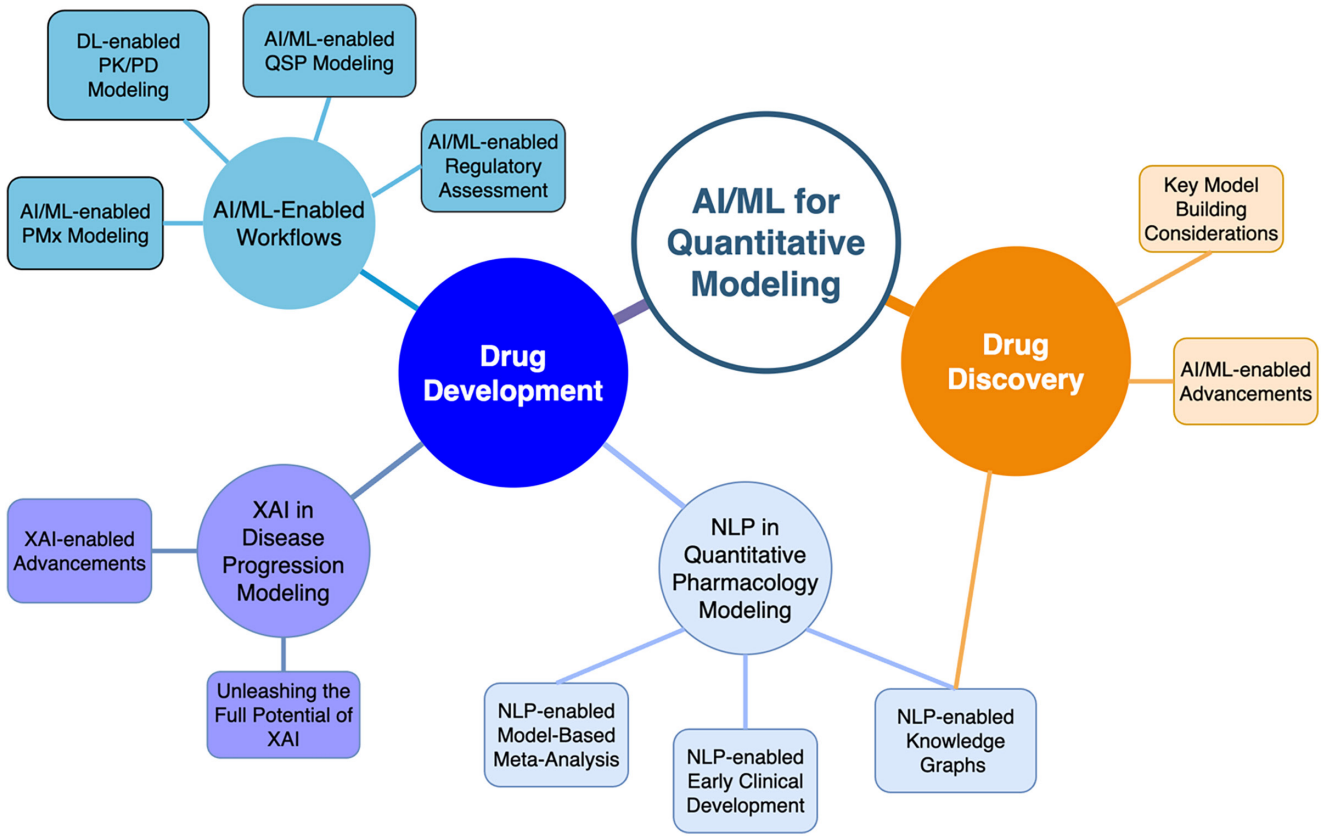

Figure 1 Overview of AI/ML for quantitative modeling in drug discovery and development. The nodes indicate the thematic areas and topics covered by this white paper, whereas the edges indicate the relationships between them. AI, artificial intelligence; DL, deep learning; ML, machine learning; NLP, natural language processing; PD, pharmacodynamic; PK, pharmacokinetic; PMx, pharmacometrics; QSP, quantitative systems pharmacology; XAI, explainable artificial intelligence.

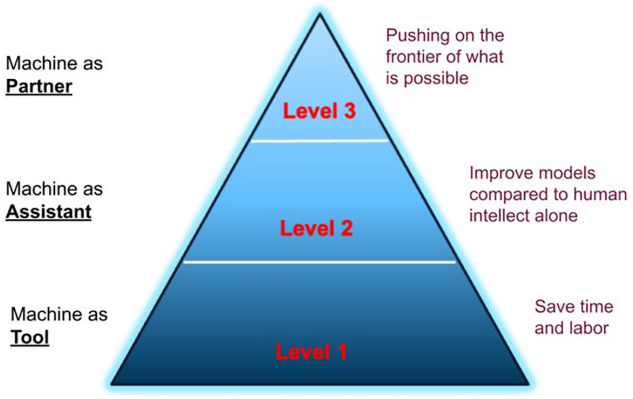

Figure 2 Machine intelligence to advance MID3 at various levels. MID3, model-informed drug discovery and development.

they did not explicitly encode causality relationships among dose, PKs, and PDs and, hence, could not enable robust predictions of new dosing regimens.

More recent research works tried to tackle this issue by integrating pharmacological aspects into deep learning (DL) architectures. For instance, Liu *et al.*[18] relied on long short-term memory (LSTM) recurrent neural networks (RNNs) to analyze simulated PK/PD data. Data from a single dosing regimen were used to train the model which was then used to predict the individual PK/PD data for other dosing regimens. Results suggested that the model could capture temporal dependencies and accurately predict PD profiles in the new settings. It is worth noting, however, that the authors simulated highly rich sampling profiles (336 timepoints), which is not a realistic data setting in standard clinical research. In addition to using RNNs, Braem *et al.*[44] implemented a pharmacologically reasonable network architecture to improve PK extrapolation to different dosing schemes. The model was trained on simulated data and transfer learning was used to adapt the predictions to new patient groups. The model was also applied to real clinical data for extrapolation to different dosing schemes. Following a slightly different approach, Lu *et al.*[19,20] explored neural ordinary differential equations (neural-ODEs) for PK/PD modeling. This is an attempt to merge DL with dynamic systems by building a pharmacology-informed DL architecture. The key concept relies on developing a pharmacology-informed encoder-decoder architecture that encapsulates the fundamental dose-concentration-effect principle. Such an architecture can enable model predictions for counterfactual dosing regimens (that is, simulating dosing regimens different from what was given to the patient), thus ensuring its generalizability. The importance of the architecture choice on the latter was further studied by Lu *et al.*[19] who compared neural-ODE with alternative approaches,

including LSTM RNN and nonlinear mixed effect models. All methods performed similarly when the training and test sets came from the same dosing regimen. However, for predicting a new treatment regimen, the neural-ODE model outperformed the other models. Further work is needed to identify the best architecture for handling covariates in neural-ODE models.

AI/ML-enabled QSP modeling

The integration of QSP and ML has recent been reviewed in a white paper.[31] Within it, the authors identified four categories of on-going research activity: (i) parameter estimation and extraction, (ii) model structure, (iii) dimension reduction, and (iv) stochasticity and virtual populations. The working group concluded that the integration of QSP and ML is still in its early stages of moving from evaluating available technical tools to building case studies.

In QSP models, the approach is often used to describe complex physiological phenomena with differential equations. However, often there are additional mechanisms which are needed to better describe the available dynamic data. In such scenarios, the adoption of universal differential equations (UDEs)[45] can prove highly beneficial, as they offer a versatile mathematical framework that allows for the integration of information derived from physical laws and scientific models, along with data-driven ML approaches. For example, Poels *et al*.[46] showed an application for toxicity predictions in immuno-oncology, which revolves around predicting the risk of cytokine release syndrome (CRS) following bispecific antibody treatment of patients with cancer. A QSP model was developed to predict CRS with a priming dose strategy. Automated model discovery was investigated, using data to learn missing terms of a system of ODEs. They used UDEs as a framework to explore this question by adding an NN component to the model. The NN component acts as a function approximator, thereby enabling the encapsulation of complex patterns from data. Sparse regression can be used to recover the equations of the additional term needed to reproduce the data. Although current methods focus on identifying empirical terms to supplement existing mechanistic equations, there are substantial future prospects in harnessing the vast biomedical knowledge present in the literature to directly generate these mechanistic equations (e.g., see Part 3 of the paper). The integration of automated model discovery and mechanistic modeling has the potential to contribute to more robust and comprehensive analyses, enabling the extraction of intricate data-driven insights and enhanced predictivity via QSP models.

AI/ML-enabled regulatory assessment

From a regulatory perspective, AI/ML approaches can be leveraged across several areas to support overall drug development and regulatory efficiency. This includes, but is not limited to: causal inference[47]; automation tools for bioequivalence assessment[48] or facilitating product specific guidance[49]; business intelligence to predict submissions of abbreviated new drug applications[50,51]; regulatory equivalence assessment for complex particle size distribution[52]; and multivariate analysis methods to facilitate active pharmaceutical ingredient sameness assessment.[53]

Of note, the value of adopting AI/ML approaches to mine large and heterogenous datasets has been shown in recent regulatory applications focusing on the assessment of heterogeneous treatment effect (HTE). The HTE analyses focus on examining varying treatment effects for individuals or subgroups in a population (e.g., for personalized medicine). For example, Gong *et al*.[47] developed a causal forest HTE method and evaluated its performance against the conventional two-step method by simulating scenarios with different levels of complexity. Causal forest outperformed the conventional method, especially when data were complex (e.g., nonlinear) and high dimensional, thus revealing a promising venue to advance analytical solutions for real-world HTE analyses.

PART 2: EXPLAINABLE ARTIFICIAL INTELLIGENCE AND ITS APPLICATIONS IN DISEASE PROGRESSION MODELING

In the previous section, we highlighted the evolution of ML as a tool, assistant, and partner for supporting decision making and advancing drug development at multiple levels. However, as ML models become more complex, it can become increasingly difficult to understand how certain decisions or predictions are made. Explainable artificial intelligence (XAI) aims to address ML models interpretability[54] by incorporating human-understandable explanations to output results. As shown in **Figure 3**, XAI can be classified into model-specific and model-agnostic approaches.[54] Model-specific approaches are tailored to a specific ML model and leverage its structure or logic to generate explanations that are then dependent on the model's design or implementation. For example, decision trees are model-specific methods that provide interpretable predictions based on a series of binary rules that split the input space into regions.[55] On the other hand, model-agnostic approaches can be applied to any ML model, regardless of its architecture or structure. For example, Local Interpretable Model-agnostic Explanations (LIME)[56] and SHapley Additive exPlanations (SHAP)[57] are well known and commonly used model-agnostic methods that provide measures of the importance or contribution of input features to the model predictions. Another key aspect of XAI is the use of visualization to help the users comprehend how ML model predictions are made. As an example, heat maps can be used to visualize the important features of a model, and decision trees can show how a model makes decisions based on different inputs.

Applications of XAI are of great importance in the field of drug development and clinical quantitative pharmacology, because model-informed decisions can have significant impact. Thus, having interpretable and explainable predictions becomes critical in order to build trust and meet fair and ethical principles. In this section, we present a few examples recently presented in the literature.

XAI-enabled advancements in disease progression modeling

Disease progression modeling (DPM) focuses on using patient characteristics and pathophysiologic information to quantitatively describe longitudinal changes in the disease trajectory as a function of time.[58] Given the increasing importance of DPM in supporting drug development, the IQ DPM working group has performed an industry-wide survey and published a white paper to summarize findings on its use.[59] Currently, DPM is mainly used for internal decision making and helping to inform clinical trial design, rather than in regulatory decision making where guidance

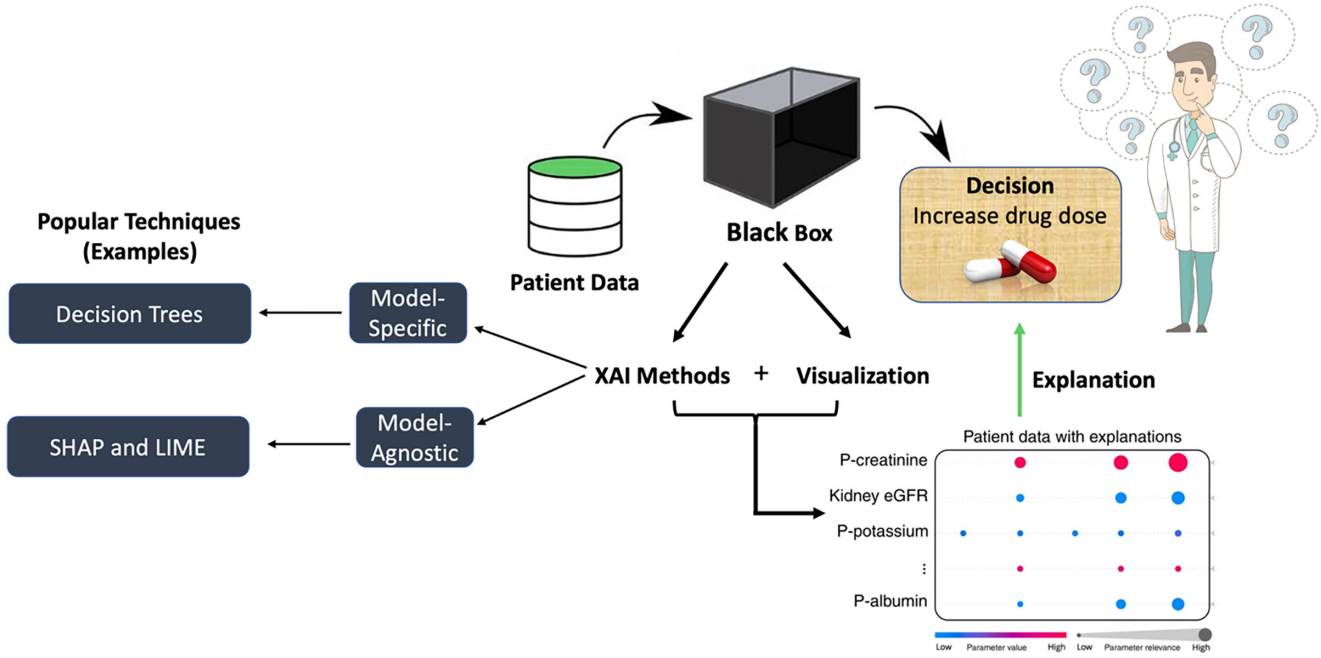

Figure 3 Example of utility of XAI methods to provide interpretable explanations of ML model predictions in order to more transparently inform clinical pharmacology decisions. eGFR, estimated glomerular filtration rate; LIME, Local Interpretable Model-agnostic Explanations; SHAP, SHapley Additive exPlanations; XAI, Explainable Artificial Intelligence.

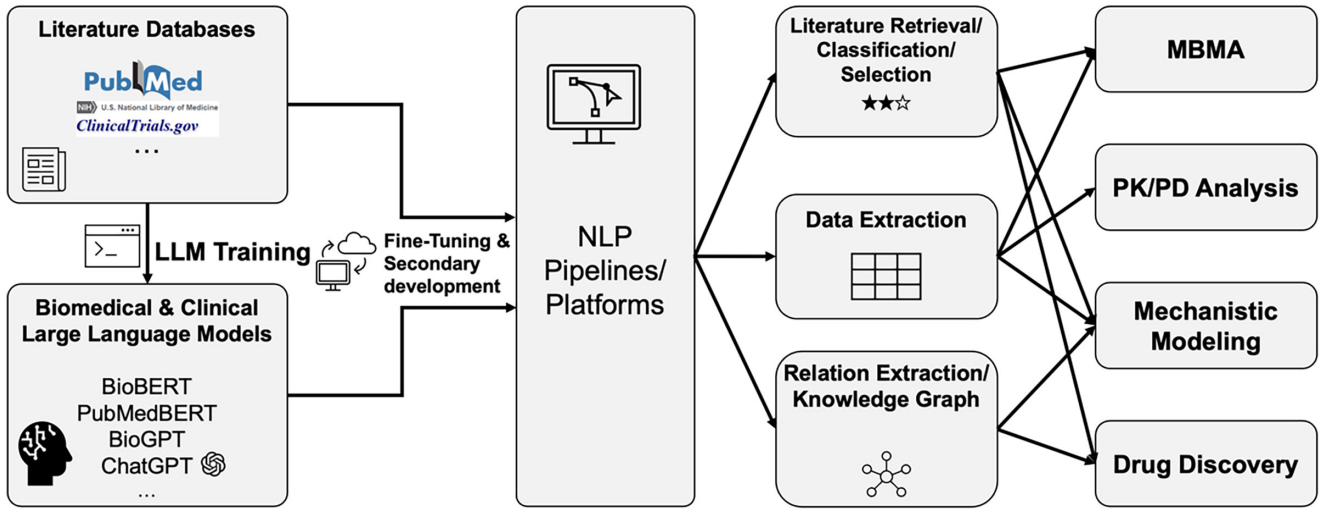

Figure 4 Leveraging modern NLP techniques to streamline and improve the efficiency of knowledge discovery and data extraction from biomedical literature, in order to aid MID3. LLM, large language model; MBMA, Model-Based Meta-Analysis; MID3, model-informed drug discovery and development; NLP, Natural Language Processing; PD, pharmacodynamic; PK, pharmacokinetic.

on best practices is sought. It has also been highlighted that AI/ML offers an exceptional opportunity to integrate large, multi-dimensional data and enable precision medicine development. Additionally, applying XAI like SHAP can help by providing novel insights into the underlying mechanisms and factors of disease progression, including latent factors and temporal dynamics of disease stages.

For instance, Basu *et al.*[60] used an explainable ML approach to predict future disease activity in patients with multiple sclerosis (MS) and identify the most predictive covariates. The analysis was conducted on a pooled population of 1,935 patients enrolled in 3 cladribine phase III clinical trials with different outcomes. Gradient boosting and SHAP methods were used to identify patients' covariates for the early prediction of disease activity, including patient baseline characteristics, longitudinal magnetic resonance imaging readouts, neurological, and laboratory measures. The value of incorporating XAI, such as SHAP, with complex nonlinear ML models clearly stands out in this work to enable an efficient assessment of covariates importance and contribution to model predictions at the population and individual patient levels,

by exploiting typical SHAP plots (e.g., feature importance, summary plot, and dependence plots). The most predictive covariates for early identification of disease activity in patients were found to be treatment, higher number of new combined unique active lesion count, higher number of new T1 hypointense black holes, and higher age-related MS severity score. Interestingly, investigations of SHAP dependence plot for treatment revealed an exact match with cladribine exposure-response relationship derived from a population repeated time-to-event model of qualifying relapsed previously developed in a more conventional PMx setting.[61] These results are supportive of the use AI/ML to address model-informed drug development and clinically focused questions integrating multimodal and heterogenous data as well multiple end points. This analysis improves understanding of the mechanism of onset of disease activity in patients with MS by allowing early identification in clinical settings and additionally enabling better patient monitoring and treatment planning.

Although various approaches exist for DPM (ranging from empirical to quantitative systems modeling) informed either by data alone and/or underlying disease biology,[59] AI/ML offers the potential to improve the predictivity of DPM models.[62,63] Although AI/ML approaches are well-suited for identifying patterns from complex data, it has been argued that they lack the ability to incorporate pharmacologic principles and drug-specific information.[58] However, as discussed earlier in the paper (part 1), recent developments in neural-ODE[64] have demonstrated the construction of pharmacology-informed neural network (PINN) architectures[20] and how they can be applied to DPM in geographic atrophy and oncology.[62] In particular, rather than identifying models that simply describe the longitudinal data, these neural-ODE models use the concept of learning autonomous dynamic systems from the disease trajectories. Furthermore, such PINN architectures consist of an encoder and a decoder network, with an information bottleneck in between. It has been demonstrated that dynamic system techniques can be used to visualize and understand the decoder that has been learned from data; furthermore, in oncology DPM, the individual patient "metrics" available at the model bottleneck can enable the interpretation of which aspects of the tumor dynamics profile are used for survival prediction.[62]

Unleashing the full potential of XAI

Although XAI has the potential to alleviate the black-box nature of complex ML models, several challenges still lie ahead to unleash XAI's full potential. The need for standardized definitions of explainability, communicating the results to non-technical audiences, as well as integrating explainability into the design of ML systems are among the major areas for further work.[65,66] In fact, XAI requires combining human intuition and systematic thinking with the ability of ML to process vast amounts of data. Scientific machine learning is one such approach where domain knowledge is coupled to flexible ML techniques in the initial framework design (also termed glass-box) to improve both accuracy and explainability,[67] but it requires more expertise to create. The transparency of ML algorithms is closely linked to their explainability, and by providing clarity to the model's internal workings it can instill greater confidence among stakeholders in the reliability and validity of the model's outputs.[68] However, explainable ML also involves ensuring ethical and legal principles are met.[69] Collaboration among data scientists, clinical pharmacologists, clinicians, legal, and ethical experts is necessary to develop accurate and XAI systems.

PART 3: NLP IN QUANTITATIVE PHARMACOLOGY MODELING

The exponential growth of biomedical and clinical knowledge stored in natural language can be overwhelming for scientists, hindering their ability to utilize the information effectively and efficiently. NLP is a powerful tool that can revolutionize drug development by extracting and analyzing information from the vast amount of biomedical literature. Traditional NLP heavily focuses on methods that analyze texts based on key words, such as the tools that were developed to drive PubMed searches. In the past few years, NLP has been revolutionized by DL methods, such as the transformer architecture[70] and by the very recent development of large language models (LLMs), such as the ChatGPT and GPT-4. These modern NLP tools can automate the identification of relevant papers, extract key information and causal relationships, generate natural-sounding text almost indistinguishable from human-written text, and summarize structured data from text.

Biomedical literature requires domain-specific models to be trained on specialized corpus and text data. Several biomedical and clinical domain-specific LLMs have been developed in the past few years, such as PubMedBERT,[71] BioBERT,[72] Med-BERT BERT,[73] and ClinicalBERT.[74] Recently, Microsoft released the BioGPT,[75] which is a more advanced domain-specific generative transformer language model pretrained on large scale biomedical literature, and represents the state-of-the-art development in the field. These LLMs have demonstrated exceptional performance on various biomedical NLP tasks, such as relation extraction,[75,76] question answering,[77] and document classification.[78] These AI models have great potential in drug research and development applications, including understanding underlying biological mechanisms for drug efficacy and toxicity and identifying drug targets or predicting drug interactions. By leveraging these advancements, researchers can more efficiently implement rational drug designs and increase the probability of success.

Using NLP, particularly LLMs, in drug development applications presents practical challenges, because the large size of the models come with high costs (e.g., the training of the GPT-4 model with hundreds of billions of parameters costs over $100 M). In practice, LLMs are mostly used as foundational models to power many specific applications and can be fine-tuned using in-house data or external plugins. However, data security risks must be considered when exposing such models or external plugins to internal proprietary patient data. To address data security concerns, in-house implementations of such LLMs or their smaller specialized versions may provide feasible alternative options. In-house models can take full advantage of the proprietary data behind the firewall to generate more specific outputs to the internal scientists. Furthermore, as evident from recent development of AutoGPT, complete automated use of these "intelligent" AI machines without human supervision can generate wrong results which can be

risky in drug development. We believe the "human-in-the-loop" concept[79] should be considered in any AI-aided drug development.

We illustrate in **Figure 4** the methodologies used in NLP-enabled literature search and selection to showcase their potential application areas within the context of MID3. In this section, we review the methodologies and provide examples of NLP-enabled literature search and selection for model-based meta-analysis (MBMA) and data extraction for PK/PD analysis and highlight the use of NLP to construct a detailed knowledge graph of disease biology from public literature.

NLP-enabled MBMA

Identifying and extracting relevant data from the biomedical literature for MBMA is a key PMx task in drug development; however, its workflow is typically a manual, labor intensive, and disease domain-expert dependent process. It involves initial keyword-based searches on public literature databases, such as PubMed, followed by selecting the most relevant papers and extracting data according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[80]; see **Figure 5** for an illustration. The final pool may only contain ~10–20% literature from the initial search results, thus, resulting in an inefficient and unscalable process. As indicated in **Figure 5**, NLP approaches present several opportunities for improving MBMA.

One recent approach proposes a pipeline[81] based on the PubMedBERT,[71] which is a transformer-based biomedical LLM trained on the whole PubMed dataset. PubMedBERT generates a tokenized vector representation for each input paper abstract. An additional in-house model[81] of three-layer ranking NN was trained on top of the PubMedBERT output vector to rank each paper for its relevance to MBMA. Whereas the PubMedBERT model parameters are untouched, the 3-layer ranking NN was trained using an internally labeled MBMA dataset consisting of 14 different diseases, with over 28,000 papers from initial PubMed searches, and around 3,000 human selected papers in the final MBMA analyses. The ranking NN was trained to distinguish the human-selected papers from the rest and generalize the ranking to unseen diseases and future publications. The pipeline achieved an overall mean recall rate of 85% and 77% along with an overall mean precision of 31% and 28% on the task of predicting unseen diseases and future publications, respectively. Similar performance was achieved on a new MBMA effort for severe acute respiratory syndrome-coronavirus 2 drug development, a disease area that was not represented in the initial dataset. The authors suggest that such an NLP-MBMA pipeline can dramatically reduce the cost (from ~5 FTE months to a few dollars of computing cost) and increase the efficiency (from months to a few minutes) of the MBMA process by automatizing literature selection and streamlining the whole process. Higher performance could be expected with a tool built into a "human-in-the-loop" system and with the integration of newly available AI tools, such as ChatGPT.

NLP-enabled early clinical development

NLP technology can help extract PK/PD and clinical related data from biomedical literature. A web-based tool using NLP techniques has been implemented[82] to extract PK/PD data from published literature[83] with apps for Named Entity Recognition

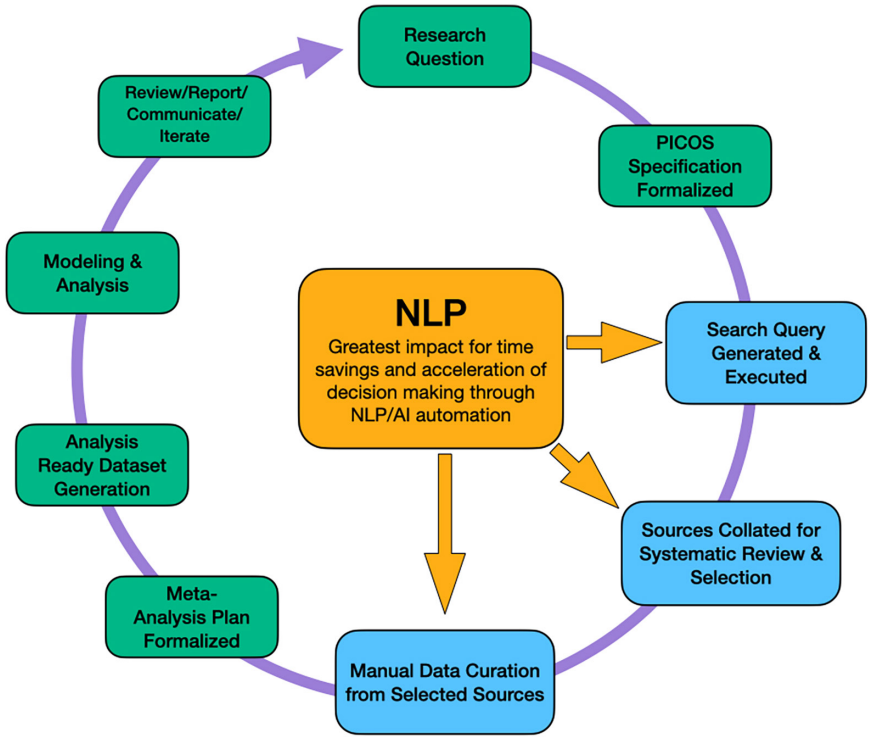

Figure 5 The workflow for MBMA consists of a number of manual, labor intensive, and disease domain dependent tasks to identify and select articles from scientific and medical literature. The arrows shown in orange indicate steps with NLP that have the largest potential impacts. AI, artificial intelligence; MBMA, Model-Based Meta-Analysis; NLP, Natural Language Processing.

(NER) relationship extraction available online. In this study, the authors developed an ML-based method to automatically identify and characterize scientific publications containing *in vivo* PK parameters, with a dataset of 4,792 PubMed publications labeled by experts. The final pipeline utilized unigram features and mean pooling of BioBERT embeddings, achieving an F1 score of 83.8% on the test set, and identified over 121,000 relevant PubMed publications. The resulting repository is accessible via a public web interface (https://app.pkpdai.com) and aims to expedite PK data search and comparison, thus aiding in ADME dataset curation.

Population, Intervention, Comparison, Outcome and Study Design (PICOS) data and clinical information are identified and extracted from textual sources. The effectiveness of NLP techniques has been demonstrated in automatically extracting PICOS elements from unstructured text.[84] In addition, NLP has been applied to clinical data extraction, including medication and adverse event extraction, to assist pharmacovigilance and adverse drug event monitoring.[85,86]

Although the use of NLP to extract clinical data from published literature has the potential to significantly enhance the efficiency and accuracy of evidence-based medicine and clinical research, significant challenges remain, including the need for high-quality annotated data, domain-specific knowledge, the potential for bias in the training data and ultimately the need to identify and extract the relevant data for subsequent modeling. Further developments in NLP techniques addressing these challenges can enable the widespread adoption of NLP in biomedical and drug research.

NLP-enabled knowledge graphs

In recent years, the integration of NLP techniques has revolutionized the construction of biomedical knowledge graphs (KGs), paving the way for its diverse applications in drug discovery and development. Santos *et al*.[87] used a KG to interpret clinical proteomics data for drug target identification and drug repurposing. Erdengasileng *et al*.[88] proposed an approach to identify potential drug–drug interactions with high accuracy. Zhang *et al*.[89] developed MatchMixeR, a cross-platform normalization method for gene expression data integration to identify new drug targets and potential drug combinations. BioKDE, a KG-based biomedical search engine and knowledge discovery platform that integrates data from various biomedical databases, including PubMed, Gene Ontology, and Reactome, can be used to identify potential drug targets based on their biological functions and interactions with other molecules, as well as drug repurposing for different diseases.[90] Additionally, NLP-based KGs can be used to identify potential off-target effects of drugs, thereby helping to develop safer and more effective drugs.

Construction of biomedical KGs requires accurate NER and reliable relation extraction. Recent advancements in NLP enabled the extraction of valuable information from biomedical text with high accuracy, despite the challenges in identifying and classifying different entities (e.g., genes, diseases, drugs, and proteins). Tian *et al*.[91] proposed a transformer-based approach for NER in clinical trial eligibility criteria, which outperformed traditional ML approaches.

NLP is important for relation extraction in constructing KGs reflective of the underlying biological mechanisms in a structured manner (e.g., drug-disease and gene-disease relationships, and protein–protein interactions). Yu *et al*.[92] proposed a Bayesian network structure learning method called GRASP, which uses an adaptive sequential Monte Carlo approach to infer the causal relationships between genes. GRASP was able to identify causal relationships between genes that were not previously known, demonstrating its potential in constructing the biomedical KG. Looking ahead, LLM NLP applications are poised to become increasingly important in future drug development. However, it is important to note that these models should be integrated with a "human-in-the-loop" approach, where human scientists are strategically placed to validate and make crucial decisions. This point will be further emphasized in the Conclusion and Recommendation section of this paper.

PART 4: AI/ML UTILIZATION IN DRUG DISCOVERY

In the 1990s, the availability of biological reagents and liquid chromatography mass spectrometry dramatically reduced the attrition of small molecule drugs due to PK considerations. Currently, attrition due to poor clinical exposure is rare, with preclinical toxicology, clinical intolerability, or insufficient efficacy being the major sources of attrition. Reagents, such as microsomes, cryopreserved hepatocytes, recombinant drug metabolizing enzymes, and cells overexpressing specific transporters, have enabled drug metabolism and PK departments to generate large quantities of *in vitro* ADME data over the last 15–20 years. These data serve two specific functions: first, *in vitro* data related to metabolic stability, plasma protein binding, permeability, efflux, and CYP inhibition can be used for the design (i.e., prior to synthesis) of small molecules with superior ADME properties (**Figure 6**, *design cycle*), along with other parameters, such as biochemical and cellular potency and selectivity data; second, archived data can be used to build ML models to predict these properties (**Figure 6**, *In silico* optimization).

In silico ML models to predict these *in vitro* ADME properties, as well as physicochemical properties, such as lipophilicity and solubility, have been available and impactful for > 15 years (**Figure 6**, *multi-parameter optimization*); generally, the models use RF or support vector machine approaches and, more recently, deep NNs.

A recent publication by the IQ *In Silico* Working Group[93] showed that the availability of a metabolic stability model at Genentech more than doubled the percentage of compounds that are metabolically stable. Similarly, the availability of a solubility model at AstraZeneca and a time-dependent CYP inhibition model at Eli Lilly significantly increased the percentage of compounds with desirable properties.[93] As the amount of data continues to increase steadily, the quality of the predictions as well as the domain of applicability (DA) will improve, and models to predict *in vivo* PK in preclinical species as well as properties of large molecules have become available recently. Progress has also been made in the prediction of potency and toxicity as well. Moreover, many of these models can be used prior to synthesis to increase the odds of success and the efficiency of the drug discovery process.

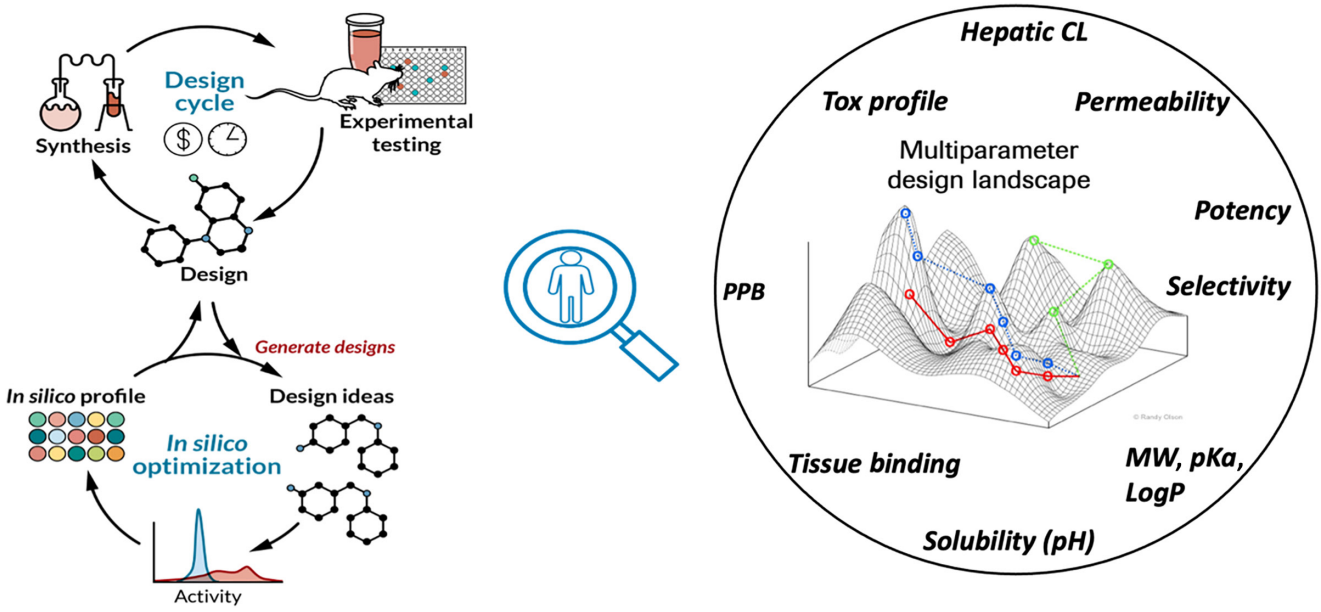

Figure 6 How *in silico* models can enhance the design cycle in drug discovery (left) resulting in better and quicker multi-parameter optimization (right). CL, clearance; PPB, plasma protein binding.

Key considerations for building models in drug discovery

There are more abundant and richer datasets now than ever before to make improved predictions about a compound's ADME, PK, and toxicity properties. However, when building and using AI/ML models, it is critical to understand what type of *in silico* model should be used to solve a particular critical issue in drug discovery, and what confidence is attached to the model's predictions based on the DA and the quality of the input data.

Open access, online databases, such as ChEMBL and PubChem, can be used to augment experimental data, but these may contain errors, including chemical structures and biological properties, or show high variability due to the use of different experimental protocols. Thus, data curation is needed to ensure high-quality input data and should include chemical structure and experimental settings. The databases of large pharmaceutical companies are often reliable sources of high-quality data.

Once an ML model has been built to predict a particular *in vitro* or *in vivo* end point of relevance to a drug discovery program, it is necessary to define the DA. The DA is a highly discussed and well-studied theoretical region of physicochemical, structural, or biological space that surrounds a model's descriptors and response, and it is used to estimate the uncertainty in a model's prediction of new compound properties based on the similarity to the compounds used in the training/test set. New methods have been developed to help with DA analysis, but no AI/ML model should be a static model, and the DA will change or broaden over time.

A clear understanding of the limitations and variability inherent in the experimental data used to build an ML model, and the associated DA, should help determine what model to use to help answer a specific question in a drug discovery program. The decision to use a local or global *in silico* model[94] depends on how much data are available for model construction and how generalizable the problem or end point is. Global models attempt to include information from later stages of drug discovery into earlier stages, are more practical, and enable extrapolation beyond the current data. If the global model is updated regularly, it is also possible to incorporate relevant local data in a timely manner. Currently, global over local models are preferred given comparable performance.[94] Choosing a local model can be impractical in the fast-moving drug discovery process.

When it comes to decision making, the impact and performance of AI/ML models can vary greatly, and their outcome is usually combined with additional evidence generated in early drug discovery programs. It is essential to carefully consider the DA and the quality of data used to the build the *in silico* model when determining the weight given to the model in the decision process. The success metric for these models should not be their accuracy, but their ability to increase the probability of success for advancing the drug discovery project and filling a gap in the decision-making process. Collaboration with ADME scientists, medicinal chemists, and toxicologists is essential in understanding the gap in knowledge and the model applicability. The goal is to bring together different relevant models across disciplines and refine optimal compound properties all the way into the clinic.

Advancements in AI/ML-enabled drug discovery

AI/ML models used for predicting ADME properties have dramatically improved and they are now being used to predict *in vivo* PK in preclinical species as well as *in vitro* ADME end points. This has led to a preference for advancing candidate molecules with desirable *in silico* predicted properties as opposed to solely relying on more resource-intensive experimental studies (e.g., P-gp efflux). New methods, such as multitask deep NNs and transfer learning, are likely to further improve *in vitro* and *in vivo* end

point predictions and shorten the time required to reach key compound decision points. In recent years, progress has been made in utilizing AI/ML for toxicity and large molecule drug development, which have lagged the use of AI/ML for predicting ADME properties for small molecular entities.

For toxicity predictions, it is key to understand the therapeutic index of a compound, and this requires both improved human PK and PD (related to the pathology of a disease) drug discovery predictions. Advances in AI/ML image analysis are leading to automation in pathology and an increased application of more advanced imaging techniques to understand biology. For human PK predictions, recent proof-of-concept work[95] has shown how AI/ML can be used to predict specific *in vivo* human PK parameters, and can eventually be used to help design compounds with more optimal PK properties. Likely, more AI/ML model development and use for predicting toxic end points will increase in the near future given new, faster, and more data rich technologies, such as multiplexed assays and multi-omics, as well as the regulatory requirement for SEND-compliant data. AI/ML methods are likely the best fit for rapidly integrating and analyzing these different data types and ultimately better understanding toxic drug effects.

Although antibody drug development has historically relied on more laboratory-based methods, recent advances in microfluidics technologies and next-generation sequencing have increased the data available for antibody identification, optimization, and *de novo* design in recent years.[96–98] For example, work led by Prescient Design (Genentech) has shown how large self-supervised "deep manifold sampling" can help produce antibody binder sequences that are stable, well-expressed, and with good drug-like properties.[99] In their work, as sequences are generated, data are fed back into an active learning framework, which selects sequences that balance model improvement and model exploitation. Future work in generating synthetic antibodies using DL can include integration with high throughput biology methods tuned to antibody discovery, as well as new methods to integrate structure and ML frameworks.

In silico models have been successfully incorporated in the drug discovery process due to their improved quality and DA, with further possibilities to progress through deep NNs in combination with transfer learning and a multitask architecture. Nevertheless, there are limitations if the synthetic efforts expand into previously unexplored chemical space associated with, for example, bifunctional degraders and macrocyclic peptides, and reliable prediction of potency (e.g., virtual screening) and (*in vivo*) toxicology is still evolving. Multiparameter optimization tools have been developed, but the scoring function still requires user input. The adoption of *in silico* models is variable, and a user-friendly interface that is incorporated effectively in the corporate computational infrastructure will aid its implementation. Collaboration between computational scientists and experimentalists is necessary to enable "augmented design" to enhance the drug discovery process.

CONCLUSIONS AND RECOMMENDATIONS

The field of AI/ML is rapidly evolving and there are significant advancements being made to support quantitative modeling for drug discovery and development at various levels (see **Figure 2**). Effective collaboration among industry partners, academia, and regulatory agencies is essential to fully understand and harness its potential. In pursuit of this goal, we have organized workshops to enable scientific exchange between practitioners and compiled our findings in this white paper. Although AI/ML can unlock many opportunities, it is important to be cautious when using these advanced algorithms to avoid deriving biased and nongeneralizable conclusions from data. Drawing on valuable inputs from industry, academia, and the FDA, we offer a set of guidelines and recommendations for the appropriate utilization of AI/ML in quantitative modeling:

- Define the context-of-use (COU) and utilize risk-informed credibility assessment framework for AI/ML applications. Similar to other quantitative models used to support drug discovery and development, the use of AI/ML should also undergo credibility assessment depending on the COU,[100] whether that be replacement of a computationally expensive covariate search, which has low decision consequence, or as part of the patient enrichment strategy which would have higher decision consequence. The level of model validation would depend on the model risk entailed, and the appropriate performance metrics may depend on the COU as well.
- Beware of potential overfitting and hence the difference in predictions and/or estimates obtained using the training versus the test sets. Although AI/ML models have high expressive power, this comes at the expense of overfitting or memorization. The commonly applied approach of evaluating PMx models on the whole dataset and drawing inferences thereof should be reconsidered in the AI/ML setting. In particular, even if the training and test sets come from the same distribution (for instance, in a cross-validation setting), overfitting may result in disparate findings between the training and test sets.[101]
- Ensure the AI/ML model exhibits sufficient generalizability outside of the training distribution, depending on its COU. Due to the expressive power (i.e., the ability to describe a wide variety of quantitative relationships) of AI/ML models, one can provide an accurate description of the existing data distribution while losing the ability to perform well for an external test set that is outside of the training distribution. For high-risk applications where generalizability outside of the training distribution is important, the ability of the model to predict outside of the training domain should be appropriately assessed and defined.
- To ensure the reproducibility of AI/ML models, it is essential to implement version control throughout the model's lifecycle. A defining feature of AI/ML models is their capacity to enhance performance through continuous learning from the accumulation of data. Consequently, these models often require periodic or even continuous updates, making version control a critical aspect of the development process. It is important to note that version control goes beyond merely creating snapshots of the model architecture and hyperparameters; it also involves referencing the training data used to generate the model.
- If possible, apply XAI methods and/or choose model formulations that enhance transparency and explainability. AI/ML models may entail a large number of decision trees or trainable

weights, which are not easily interpretable. However, there are various approaches to improve model explainability. One path is to use XAI techniques like SHAP[57] and LIME[56] to quantify how the input features impact model predictions. By evaluating which are the most predictive features and how they impact the predictions, one can eliminate spurious effects from the analysis. For instance, if data has missing values that fall under the category of Missing Not At Random,[102] one needs to ensure that the missingness pattern is not used unintentionally in the ML model. There are also explicit ways to improve explainability via the choice of model formulation: for instance, encoder-decoder NN architectures perform data abstraction by compressing them through a "bottleneck" layer.[62] By creating low dimensional embeddings of data that exhibit parsimony in explaining patient variability, such methodologies can enhance the ability of the model to be comprehended.

- If possible, incorporate relevant domain concepts into the AI/ML formalisms to enhance its generalizability. One key difference in human constructed models versus AI/ML models is that the former often incorporates key principles that are well-accepted within the scientific domain of interest, for instance, physics or pharmacology, whereas the latter is often purely data driven. However, building AI/ML models that are physics- and/or pharmacology-informed can significantly improve its generalizability. An example is geometric DL,[103] which leverages concepts, such as invariance and equivariance, to ensure that the AI/ML model exhibits symmetries present in the physical tasks at hand. Approaches such as scientific ML[104] and pharmacology-informed neural networks[62] are other proposals that attempt to reconcile domain concepts with the data-driven nature of AI/ML.
- Quantify the uncertainty of the AI/ML model predictions via performing appropriate bootstrap. Although parameter and prediction uncertainty may often be easily quantifiable in an analytical fashion in empirical PMx and statistical models, it is not so for AI/ML models. Nevertheless, confidence intervals of AI/ML model predictions can still be computed via bootstraps (i.e., sampling-with-replacement) or other approximations, such as performing dropouts in the context of NNs.[105] Quantifying uncertainty can help better assess the quality of the predictions and how well they are supported by the existing data.
- If possible, encode causality relationships into the AI/ML model. In contrast to applications of AI/ML in other technical fields, for pharmacology/toxicology applications there are often explicit causal assumptions being made among dose, PK, and efficacy/safety. If the model architecture does not explicitly take these into account, it is not guaranteed that the AI/ML model would extrapolate well outside of the training domain.[19,62]
- Draw a causal diagram to determine which variables should go into the AI/ML model and which should not. Whereas AI/ML models can incorporate many more explanatory variables than alternative approaches, using them can create the temptation of incorporating all available variables. However, for causal inference applications, it is well-recognized that the use of AI/ML algorithms by themselves is not a replacement for the need to consider which variables should be included (e.g., as confounders) or left out (e.g., as colliders).[101,106,107]
- If possible, use synthetic dataset to demonstrate the soundness of the proposed AI/ML workflow. In comparison to workflows for PMx and statistical models, AI/ML methodologies can entail multiple computational steps, including hyperparameter tuning, model training, model evaluation with validation and test sets, estimation of confidence intervals, and feature importance attributions. If proper care is not taken in the sequence of steps, one can unintentionally introduce biases into the model predictions and/or inferences, or over- and underestimate the confidence intervals.[101] By testing the planned AI/ML workflow on appropriate synthetic data, one may uncover potential flaws within the model generation process.
- Involve "human-in-the-loop" where relevant. Depending on the COU, consideration should be given as to whether and how human scientists should be strategically placed to validate AI/ML model findings and make crucial decisions.

ACKNOWLEDGMENTS

The authors acknowledge members of the IQ AI/ML Working Group as well the inputs from the following speakers and panelists who participated in the workshops organized by the working group (listed alphabetically): Jiang Bian, Richard Bonneau, Peter Bonate, Fabio Broccatelli, Narayan Cheruvu, Prashant Desai, Jenny Ding, Gaurav Dwivedi, Jonathan L. French, Nigel Greene, Meng Hu, Elizabeth Joshi, Sarah Kim, Karthik Lingineni, Dan Lu, Frank Kloprogge, James Kozloski, Qi Liu, Steven Liu, Kamrine Poels, Hoifung Poon, John Sanders, Shinichi Tsuchiwata, and Jinfeng Zhang. We could also like to acknowledge the IQ Secretariat Lee Nagao, the invaluable help provided by Maja Leah Marshall in the organization of the IQ virtual workshop, as well as the leadership of IQ CPLG.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

All authors were employees and additionally may be shareholders of their respective companies at the time of writing.

1. Marshall, S.F. *et al*. Good practices in model-informed drug discovery and development: practice, application, and documentation. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 93–122 (2016).
2. McComb, M., Bies, R. & Ramanathan, M. Machine learning in pharmacometrics: opportunities and challenges. *Br. J. Clin. Pharmacol.* **88**, 1482–1499 (2022).
3. Badillo, S. *et al*. An introduction to machine learning. *Clin. Pharmacol. Ther.* **107**, 871–885 (2020).
4. Liu, Q. *et al*. Application of machine learning in drug development and regulation: current status and future potential. *Clin. Pharmacol. Ther.* **107**, 726–729 (2020).
5. 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. Pharmacol. Ther.* **113**, 771–774 (2023).

6. Barrett, J.S. Time to expect more from Pharmacometrics. *Clin. Pharmacol. Ther.* **108**, 1129–1131 (2020).
7. Bonate, P.L. *et al*. Guidelines for the quality control of population pharmacokinetic-pharmacodynamic analyses: an industry perspective. *AAPS J.* **14**, 749–758 (2012).
8. Dykstra, K. *et al*. Reporting guidelines for population pharmacokinetic analyses. *J. Pharmacokinet. Pharmacodyn.* **42**, 301–314 (2015).
9. FDA Guidance for Industry. *Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications* <https://www.fda.gov/media/71277/download> (April 2003). Accessed May 30, 2023.
10. Acosta, J.N., Falcone, G.J., Rajpurkar, P. & Topol, E.J. Multimodal biomedical AI. *Nat. Med.* **28**, 1773–1784 (2022).
11. Bai, J.P.F. *et al*. Quantitative systems pharmacology: landscape analysis of regulatory submissions to the US Food and Drug Administration. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 1479–1484 (2021).
12. Chan, J.R. *et al*. Current practices for QSP model assessment: an IQ consortium survey. *J Pharmacokinet Pharmacodyn*, **11**, 1–13 (2022).
13. Zame, W.R. *et al*. Machine learning for clinical trials in the era of COVID-19. *Stat. Biopharm. Res.* **12**, 506–517 (2020).
14. International Consortium for Innovation & Quality in Pharmaceutical Development <https://iqconsortium.org/> (2023). Accessed May 30, 2023.
15. *IQ Consortium Working Group for Artificial Intelligence & Machine Learning* <https://iqconsortium.org/initiatives/working-groups/artificial-intelligence-and-machine-learning/> (2022). Accessed May 30, 2023.
16. Sibieude, E., Khandelwal, A., Hesthaven, J.S., Girard, P. & Terranova, N. Fast screening of covariates in population models empowered by machine learning. *J. Pharmacokinet. Pharmacodyn.* **48**, 597–609 (2021).
17. Sibieude, E., Khandelwal, A., Girard, P., Hesthaven, J.S. & Terranova, N. Population pharmacokinetic model selection assisted by machine learning. *J. Pharmacokinet. Pharmacodyn.* **49**, 257–270 (2022).
18. Liu, X. *et al*. Long short-term memory recurrent neural network for pharmacokinetic-pharmacodynamic modeling. *Int. J. Clin. Pharmacol. Ther.* **59**, 138–146 (2021).
19. 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).
20. 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. Machine Intell.* **3**, 696–704 (2021).
21. Fang, C., Xu, D., Su, J., Dry, J.R. & Linghu, B. DeePaN: deep patient graph convolutional network integrating clinico-genomic evidence to stratify lung cancers for immunotherapy. *NPJ Digit Med* **4**, 14 (2021).
22. Cui, Z.L., Kadziola, Z., Lipkovich, I., Faries, D.E., Sheffield, K.M. & Carter, G.C. Predicting optimal treatment regimens for patients with HR+/HER2- breast cancer using machine learning based on electronic health records. *J Comp Eff Res* **10**, 777–795 (2021).
23. Fisher, C.K., Smith, A.M. & Walsh, J.R. Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Sci. Rep.* **9**, 13622 (2019).
24. Young, A.L. *et al*. Disease progression modeling in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **201**, 294–302 (2020).
25. Pinto, M.F. *et al*. Prediction of disease progression and outcomes in multiple sclerosis with machine learning. *Sci. Rep.* **10**, 21038 (2020).
26. Qian, Z.Z., William, R., Fleuren Lucas, M., Elbers, P. & van der Schaar, M. *Integrating Expert ODEs into Neural ODEs: Pharmacology and Disease Progression*. <https://arxiv.org/abs/2106.02875 ttps://arxiv.org/pdf/2106.02875.pdf> (2021). Accessed May 30, 2023.
27. Bica, I., Alaa, A.M., Lambert, C. & van der Schaar, M. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. *Clin. Pharmacol. Ther.* **109**, 87–100 (2021).
28. Wager, S. & Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **113**, 1228–1242 (2018).
29. Talevi, A. *et al*. Machine learning in drug discovery and development part 1: a primer. *CPT Pharmacometrics Syst. Pharmacol.* **9**, 129–142 (2020).
30. Terranova, N., Venkatakrishnan, K. & Benincosa, L.J. Application of machine learning in translational medicine: current status and future opportunities. *AAPS J.* **23**, 74 (2021).
31. Zhang, T. *et al*. Two heads are better than one: current landscape of integrating QSP and machine learning: an ISoP QSP SIG white paper by the working group on the integration of quantitative systems pharmacology and machine learning. *J. Pharmacokinet. Pharmacodyn.* **49**, 5–18 (2022).
32. *IQ Machine Intelligence for Quantitative Modeling in Drug Discovery & Development Applications Workshop* <https://iqconsortium.org/initiatives/working-groups/artificial-intelligence-and-machine-learning/> (September 2022). Accessed May 30, 2023.
33. *Explainable Machine Learning for Disease Progression Modeling & Digital Twins* <https://www.go-acop.org/default.asp?id=46&keuze=meeting&mid=21> (November 2022). Accessed May 30, 2023.
34. Venkatakrishnan, K. & van der Graaf, P.H. Model-informed drug development: connecting the dots with a totality of evidence mindset to advance therapeutics. *Clin. Pharmacol. Ther.* **110**, 1147–1154 (2021).
35. Sherer, E.A. *et al*. Application of a single-objective, hybrid genetic algorithm approach to pharmacokinetic model building. *J. Pharmacokinet. Pharmacodyn.* **39**, 393–414 (2012).
36. 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).
37. Abler, D. *et al*. Semiautomated pipeline to quantify tumor evolution from real-world positron emission tomography/computed tomography imaging. *JCO Clin Cancer Inform* **7**, e2200126 (2023).
38. 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**, 506–517 (2023).
39. Ismail, M. *et al*. Development of a genetic algorithm and NONMEM workbench for automating and improving population pharmacokinetic/pharmacodynamic model selection. *J. Pharmacokinet. Pharmacodyn.* **49**, 243–256 (2022).
40. Gobburu, J.V. & Chen, E.P. Artificial neural networks as a novel approach to integrated pharmacokinetic-pharmacodynamic analysis. *J. Pharm. Sci.* **85**, 505–510 (1996).
41. Chow, H.H., Tolle, K.M., Roe, D.J., Elsberry, V. & Chen, H. Application of neural networks to population pharmacokinetic data analysis. *J. Pharm. Sci.* **86**, 840–845 (1997).
42. Mager, D.E., Shirey, J.D., Cox, D., Fitzgerald, D.J. & Abernethy, D.R. Mapping the dose-effect relationship of orbofiban from sparse data with an artificial neural network. *J. Pharm. Sci.* **94**, 2475–2486 (2005).
43. Kang, S.H. *et al*. Population pharmacokinetic and pharmacodynamic models of remifentanil in healthy volunteers using artificial neural network analysis. *Br. J. Clin. Pharmacol.* **64**, 3–13 (2007).
44. Bräm, D.S., Parrott, N., Hutchinson, L. & Steiert, B. Introduction of an artificial neural network-based method for concentration-time predictions. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 745–754 (2022).
45. Rackauckas, C. *et al*. *Universal differential equations for scientific machine learning* <https://arxiv.org/abs/2001.04385> (2020). Accessed May 30, 2023.
46. Poels, K. *A Machine Learning Based Approach for Toxicity Predictions in Immuno-Oncology. Presented at IQ Machine*

Intelligence for Quantitative Modeling in Drug Discovery & Development Applications Workshop <https://iqconsortium.org/images/uploads/Session_1_-_A_Machine_Learning_Based_Approach_for_Toxicity_Predictions_in_Immuno-Oncology_–_Kamrine_Poels_Pfizer.pdf> (September 2022). Accessed May 30, 2023.
47. Gong, X., Hu, M., Basu, M. & Zhao, L. Heterogeneous treatment effect analysis based on machine-learning methodology. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 1433–1443 (2021).
48. Hu, M. *BE ASSESSMENT MATE (BEAM) - A Data Analytics Tool to Enhance Efficiency, Quality, and Consistency of Bioequivalence Assessment. Presented at 2021 FDA Science Forum* <https://www.fda.gov/science-research/fda-science-forum/2021-fda-science-forum-agenda> (2021). Accessed May 31, 2023.
49. Shi, Y., Ren, P., Zhang, Y., Gong, X., Hu, M. & Liang, H. Information extraction from FDA drug labeling to enhance product-specific guidance assessment using natural language processing. *Front. Res. Metr. Anal.* **6**, 670006 (2021).
50. Gong, X., Hu, M. & Zhao, L. Big data toolsets to Pharmacometrics: application of machine learning for time-to-event analysis. *Clin. Transl. Sci.* **11**, 305–311 (2018).
51. Hu, M. *et al*. Predictive analysis of first abbreviated new drug application submission for new chemical entities based on machine learning methodology. *Clin. Pharmacol. Ther.* **106**, 174–181 (2019).
52. Hu, M. *et al*. Equivalence testing of complex particle size distribution profiles based on earth mover's distance. *AAPS J.* **20**, 62 (2018).
53. Rogstad, S. *et al*. Modern analytics for synthetically derived complex drug substances: NMR, AFFF-MALS, and MS tests for glatiramer acetate. *Anal. Bioanal. Chem.* **407**, 8647–8659 (2015).
54. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: a review of machine learning interpretability methods. *Entropy (Basel)* **23**, 18 (2021). https://doi.org/10.3390/e23010018
55. Song, C. & Zhang, H. Comments on fifty years of classification and regression trees. *Int. Stat. Rev.* **82**, 359–361 (2014).
56. Ribeiro, M.T., Singh, S. & Guestrin, C. "Why should I trust you?": explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 1135–1144.
57. Lundberg, S.M. *et al*. From local explanations to global understanding with explainable AI for trees. *Nat. Machine Intell.* **2**, 56–67 (2020).
58. Barrett, J.S., Nicholas, T., Azer, K. & Corrigan, B.W. Role of disease progression models in drug development. *Pharm. Res.* **39**, 1803–1815 (2022).
59. Goteti, K. *et al*. Opportunities and challenges of disease progression modeling in drug development – an IQ perspective. *Clin. Pharmacol. Ther.* **114**, 266–274 (2023).
60. 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).
61. Hermann, R., Karlsson, M.O., Novakovic, A.M., Terranova, N., Fluck, M. & Munafo, A. The clinical pharmacology of Cladribine tablets for the treatment of relapsing multiple sclerosis. *Clin. Pharmacokinet.* **58**, 283–297 (2019).
62. Lu, J. *et al*. *Deep Learning for PK/PD and Disease Progression Time-Course Predictions using Neural-ODE* <https://www.page-meeting.org/default.asp?abstract=9683> (2021).
63. Laurie, M.L. J *Neural Ordinary Differential Equations for Tumor Dynamics Modeling and Overall Survival Predictions* <https://www.page-meeting.org/default.asp?abstract=9992> (2022).
64. Chen, R.Q., Rubanova, Y., Bettencourt, J. & Duvenaud, D. *Neural Ordinary Differential Equations* <https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf> (2018). Accessed May 31, 2023.
65. Markus, A.F., Kors, J.A. & Rijnbeek, P.R. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J. Biomed. Inform.* **113**, 103655 (2021).
66. Páez, A. The pragmatic turn in explainable artificial intelligence (XAI). *Minds Machines* **29**, 441–459 (2019).
67. Ramadhan, A.M. *et al*. *Capturing missing physics in climate model parameterizations using neural differential equations*. <https://doi.org/10.48550/arXiv.2010.12559> (2020, revised 2023). Accessed May 30, 2023.
68. Huang, A.A. & Huang, S.Y. Increasing transparency in machine learning through bootstrap simulation and shapely additive explanations. *PLoS One* **18**, e0281922 (2023).
69. Ploug, T. & Holm, S. The four dimensions of contestable AI diagnostics – a patient-centric approach to explainable AI. *Artif. Intell. Med.* **107**, 101901 (2020).
70. Devlin, J.C., Chang, M.-W., Lee, K. & Toutanova, K. *BERT: pre-training of deep bidirectional transformers for language understanding*. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* <https://aclanthology.org/N19-1423> (June 2019).
71. Gu, Y. *et al*. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans. Comput. Healthcare* **3**, 2 (2021).
72. Lee, J. *et al*. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **36**, 1234–1240 (2019).
73. Rasmy, L., Xiang, Y., Xie, Z., Tao, C. & Zhi, D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ Digital Medicine* **4**, 86 (2021).
74. Huang, K., Altosaar, J. & Ranganath, R. Clinicalbert: modeling clinical notes and predicting hospital readmission <https://arxiv.org/abs/1904.05342> (2019). Accessed May 31, 2023.
75. Luo, R. *et al*. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Brief. Bioinform.* **23** (2022).
76. Su, P. & Vijay-Shanker, K. Investigation of improving the pre-training and fine-tuning of BERT model for biomedical relation extraction. *BMC Bioinformatics* **23**, 120 (2022).
77. Yoon, W., Jackson, R., Lagerberg, A. & Kang, J. Sequence tagging for biomedical extractive question answering. *Bioinformatics* **38**, 3794–3801 (2022).
78. Chen, Q., Du, J., Allot, A. & Lu, Z. LitMC-BERT: transformer-based multi-label classification of biomedical literature with an application on COVID-19 literature curation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **19**, 2584–2595 (2022).
79. Monarch, R.M. *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI* (Manning, Shelter Island, New York, USA, 2021).
80. Page, M.J. *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).
81. Ding, J.C., Cao, Y., Hayes, S., Bryman, G. & Yee, K. *Using Natural Language Processing (NLP) to Streamline Literature Selection for Meta-Analysis (MA). Presented at IQ Machine Intelligence for Quantitative Modeling in Drug Discovery & Development Applications Workshop* <https://iqconsortium.org/images/uploads/Using_Natural_Language_Processing_(NLP)_to_Streamline_Literature_Selection_for_Meta-Analysis_(MA).pdf> (September 2022).
82. Kloprogge, F. *An automated approach to extract pharmacokinetic parameters from scientific publications. Presented at IQ Machine Intelligence for Quantitative Modeling in Drug Discovery & Development Applications Workshop* <https://iqconsortium.org/images/uploads/An_automated_approach_to_extract_pharmacokinetic_parameters_from_scientific_publications.pdf> (September 2022).
83. Gonzalez Hernandez, F. *et al*. An automated approach to identify scientific publications reporting pharmacokinetic parameters. *Wellcome Open Res.* **6**, 88 (2021).
84. Marshall, I.J. & Wallace, B.C. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Syst. Rev.* **8**, 163 (2019).
85. Chen, L. *et al*. Extracting medications and associated adverse drug events using a natural language processing system

combining knowledge base and deep learning. *J. Am. Med. Inform. Assoc.* **27**, 56–64 (2020).
86. Feller, D.J., Zucker, J., Yin, M.T., Gordon, P. & Elhadad, N. Using clinical notes and natural language processing for automated HIV risk assessment. *J. Acquir. Immune Defic. Syndr.* **77**, 160–166 (2018).
87. Santos, A. *et al*. A knowledge graph to interpret clinical proteomics data. *Nat. Biotechnol.* **40**, 692–702 (2022).
88. Erdengasileng, A. *et al*. Pre-trained models, data augmentation, and ensemble learning for biomedical information extraction and document classification. *Database (Oxford)* **2022** (2022).
89. Zhang, S., Shao, J., Yu, D., Qiu, X. & Zhang, J. MatchMixeR: a cross-platform normalization method for gene expression data integration. *Bioinformatics* **36**, 2486–2491 (2020).
90. Chung, M., Zhou, J., Pang, X., Tao, Y. & Zhang, J. BioKDE: a deep learning powered search engine and biomedical knowledge discovery platform *BioCreative VII Challenge Evaluation Workshop, Virtual Workshop*, 254–259 (2021).
91. Tian, S. *et al*. Transformer-based named entity recognition for parsing clinical trial eligibility criteria. *ACM BCB* **2021** (2021).
92. Yu, K., Cui, Z., Sui, X., Qiu, X. & Zhang, J. Biological network inference with GRASP: a Bayesian network structure learning method using adaptive sequential Monte Carlo. *Front. Genet.* **12**, 764020 (2021).
93. Lombardo, F. *et al*. In Silico absorption, distribution, metabolism, excretion, and pharmacokinetics (ADME-PK): utility and best practices. An industry perspective from the International Consortium for Innovation through Quality in Pharmaceutical Development. *J Med Chem* **60**, 9097–9113 (2017).
94. Di Lascio, E., Gerebtzoff, G. & Rodríguez-Pérez, R. Systematic evaluation of local and global machine learning models for the prediction of ADME properties. *Mol. Pharm.* **20**, 1758–1767 (2023).
95. Miljković, F. *et al*. Machine learning models for human In vivo pharmacokinetic parameters with in-house validation. *Mol. Pharm.* **18**, 4520–4530 (2021).
96. Ruffolo, J.A., Guerra, C., Mahajan, S.P., Sulam, J. & Gray, J.J. Geometric potentials from deep learning improve prediction of CDR H3 loop structures. *Bioinformatics* **36**, i268–i275 (2020).
97. Gainza, P. *et al*. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **17**, 184–192 (2020).
98. Lim, Y.W., Adler, A.S. & Johnson, D.S. Predicting antibody binders and generating synthetic antibodies using deep learning. *MAbs* **14**, 2069075 (2022).
99. Gligorijević, V. *et al*. Function-guided protein design by deep manifold sampling. *bioRxiv*, 2021.12.22.473759 (2021).
100. Kuemmel, C. *et al*. Consideration of a credibility assessment framework in model-informed drug development: potential application to physiologically-based pharmacokinetic modeling and simulation. *CPT Pharmacometrics Syst. Pharmacol.* **9**, 21–28 (2020).
101. Harun, R., Yang, E., Kassir, N., Zhang, W. & Lu, J. Machine learning for exposure-response analysis: methodological considerations and confirmation of their importance via computational experimentations. *Pharmaceutics* **15** (2023).
102. Rubin, D.B. Inference and missing data. *Biometrika* **63**, 581–592 (1976).
103. Bronstein, M.B., Bruna, J., Cohen, T. & Veličković, P. *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges* (2021) https://doi.org/10.48550/arXiv.2104.13478 Accessed May 30, 2023.
104. Raissi, M., Perdikaris, P. & Karniadakis, G.E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
105. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. *Proceedings of the 33rd International Conference on International Conference on Machine Learning* **48**, 1050–1059 (2016).
106. Hern'an, M.A. & Robins, J. *Causal Inference: What if* (Chapman & Hall/CRC, Boca Raton, 2020).
107. Rogers, J.A., Maas, H. & Pitarch, A.P. An introduction to causal inference for pharmacometricians. *CPT Pharmacometrics Syst. Pharmacol.* **12**, 27–40 (2023).