

REVIEW ARTICLE

Role of AI in drug development: Current status, challenges, opportunities, and future promise

Sheikh Usman Iqbal^{1*}, Dmitriy Podolskiy², Mona G. Flores³, Victoria L. Chiou⁴, Jennifer Sheng⁵, Roberto Araujo⁶, Naveed Afzal⁷, David A. Hall⁸, Ingrid Vasiliu-Feltes⁹, Dipu Patel¹⁰, and Manolis Kellis^{11,12}

¹Julz Pharma, Boston, United States of America

²Astellas Pharma, Westborough, Massachusetts, United States of America

³Independent researcher, Portland, Oregon, United States of America

⁴Tempus AI, Inc., Chicago, Illinois, United States of America

⁵College of Pharmacy, University of Michigan, Ann Arbor, Michigan, United States of America

⁶Sanofi, Cambridge, Massachusetts, United States of America

⁷Takeda Pharmaceutical Company Limited, Cambridge, Massachusetts, United States of America

⁸AKT Health Inc., Cambridge, Massachusetts, United States of America

⁹Department of Health Management and Policy, Herbert Business School, University of Miami, Coral Gables, Florida, United States of America

¹⁰Department of PA Studies, School of Health and Rehabilitation Sciences, MGH Institute of Health Professions, Boston, Massachusetts, United States of America

¹¹Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

¹²Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

***Corresponding author:**

Sheikh Usman Iqbal
(usman@bu.edu)

Citation: Iqbal SU, Podolskiy D, Flores MG, *et al.* Role of AI in drug development: Current status, challenges, opportunities, and future promise. *Artif Intell Health*. doi: 10.36922/AIH025470103

Received: November 18, 2025

Revised: January 9, 2026

Accepted: January 15, 2026

Published online: March 13, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Artificial intelligence (AI) heralds a transformative shift in drug development, with speed, precision, and predictive power as its core features. Advances in systems-level biology platforms, coupled with substantial investments in generative AI-centric pharma integration, have fostered healthy optimism among stakeholders about identifying new cures through renewed approaches and improved productivity. However, navigating epistemological, ethical, patient safety, and ontological dimensions within research and development (R&D) presents challenges that AI must address to enhance its mainstream adoption and practical utility. Here, multidisciplinary experts discuss key applications of AI across the full continuum of drug development, examine the challenges encountered, and propose solution frameworks. Drug development remains fraught with unknown biology, patient heterogeneity, and perplexing therapeutic risks. Stringent regulatory and compliance guidelines further necessitate that conventional pharma processes, practices, and strategies remain paramount in R&D execution, while guiding the integration of AI in a “value-for-effort,” evidence-based, yet Promethean fashion.

Keywords: Generative artificial intelligence; Machine learning; Drug discovery; Synthetic control arms; Digital twins; AI-driven clinical research

1. Introduction

The development of new medicines is a long, complex, and multilayered process in which research, technology, and data science must interact to advance the quest for faster and more effective treatments. Artificial intelligence (AI) is increasingly believed to play a key role in this domain. However, industry, academic, and regulatory stakeholders are still pondering how best to implement effective AI-enabled innovation within drug development. Although drug development involves multiple integrated disciplines, the application of AI remains concentrated in specific areas, with relatively limited interaction across different functions. While many discovery scientists are embracing AI at the systems level, they often lack clear guidance on translating its relevant outputs to the clinical development arena, including population enrichment, trial design, and endpoint optimizations. Conversely, many clinical development professionals are not only unfamiliar with the role of AI in drug discovery but also with the application and integration of AI into clinical trials. First, there is a need to cross-fertilize the end-to-end knowledge in drug development in an interdisciplinary and integrated fashion to fully harness the benefits of AI, while understanding its limitations across the full scope of research and development (R&D) activities. Second, effective applications of AI in drug development can encounter challenges that emerge in opaque or unexpected ways. As with any transformative technology, it is critical to assess the disparity between AI's greatest projected capabilities and its practical, real-world applications. The objectives of this paper are as follows:

- (i) To provide a narrative review of the end-to-end landscape and potential scope of AI in key areas of drug development.
- (ii) To offer a balanced perspective on current opportunities and inherent limitations of AI.
- (iii) To propose practical frameworks and checklist-based criteria for more effective integration of AI into drug discovery, preclinical development, and clinical development.

2. Methods

2.1. Source materials

Anchored in real-life experiences and applied frameworks in drug development, this review draws upon a year-long series of seminars held in Cambridge's Kendall-MIT Square, known as the Pharma Leaders Decentralized Clinical Trials (DCT) Monthly Series.¹ From March to December 2024, the DCT series assembled a host of leaders from life sciences, pharmaceuticals, biotechnology,

and healthcare sectors to explore how different emerging technologies, including AI, digital health, and real-world evidence (RWE), can reshape drug development. The meetings focused on patient-centric trial paradigms, strategic shifts toward validated system-level AI, DCT systems designed from the ground up, and the importance of collaborative networks and domain-specific leadership. Insights and converging themes from this series informed the transition to a dedicated AI workstream and ultimately shaped the scope and direction of this paper.

2.2. Application, vetting, and organization of materials

Following the broader DCT series, a targeted panel comprising 11 multidisciplinary experts with extensive experience in AI integration across drug development was convened to author this paper. The panel possessed diverse expertise spanning AI and machine learning (ML), drug discovery, pharmacology, clinical development, medical evidence and RWE, data science and governance, and regulatory science. This heterogeneity facilitated both domain-specific representation and generalized expertise-oriented content development for the manuscript. The authors used both generalized and modular Delphi panels to analyze the role of multifaceted AI in drug development. The initial goal of the Delphi process was to map the current status, unmet needs, and critical knowledge gaps in the application of AI across discovery, preclinical, and clinical development stages. This was followed by a consensus-based approach to propose practical solution frameworks that could optimally influence the AI landscape in drug development. Our team first anonymized and aggregated responses from DCT speakers and panelists to a set of carefully designed question prompts. All feedback was then synthesized, categorized, and prioritized per thematic modules aligned with the paper's main sections. The respective modules were recirculated back to the Delphi panel for two additional rounds of discussion. Finally, viewpoints were refined, and insights converged while preserving the diversity of opinion. Consensus was measured using preset thresholds. The proposed solution frameworks and AI application checklists were also reviewed by select experts from the DCT think tank for applied input and were compared with existing literature to confirm contextual alignment.

The role and framework of AI across the R&D value chain are described in the following sequential sections, which cover key insights, application models, challenges, and solution strategies. An overview is presented in [Figure 1](#), which emphasizes the use of Kirkpatrick model of learning and evaluation.²

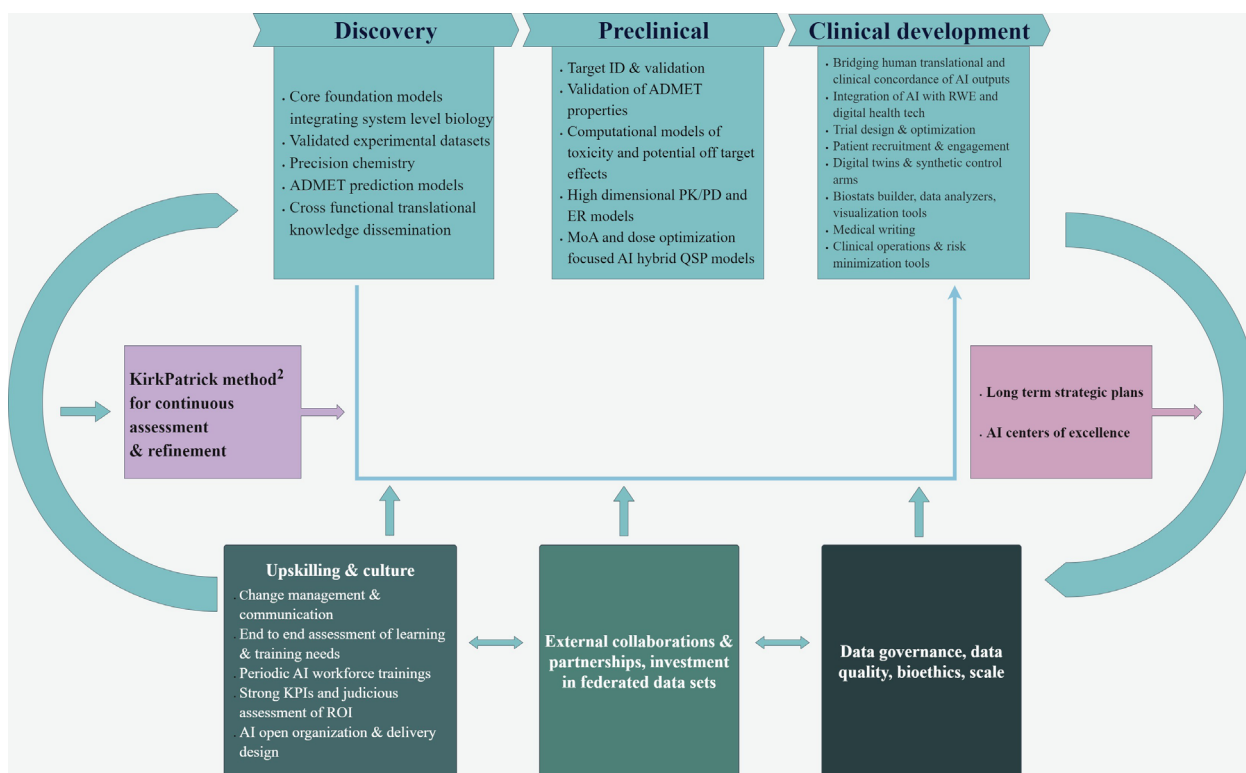


Figure 1. Pharma AI enterprise: “end-to-end” and “end-in-mind” integration of AI in drug development. Image created by the authors.

Abbreviations: ADMET: Absorption, distribution, metabolism, excretion, and toxicity; AI: Artificial Intelligence; ID: Identification; KPIs: Key performance indicators; QSP: Quantitative systems pharmacology; PK/PD: Pharmacokinetic/pharmacodynamic; ROI: Return on investment; RWE: Real-world evidence.

3. Results

3.1. Drug discovery

Over the last two decades, small-molecule drugs have remained—and are expected to continue to remain—the backbone of modern therapeutics, addressing a broad spectrum of disease mechanisms. Traditional drug discovery has involved phenotypic screening of large compound libraries for hit identification, followed by systematic optimization to improve compound efficacy, selectivity, and pharmacokinetic (PK) properties.^{3–7} ML extends this paradigm through condensed virtual screening approaches that avoid explicit biophysical simulations, instead relying on representation learning and latent embeddings.^{8–11} Generative AI (Gen AI) may redefine this process by enabling *in silico* compound development with targeted PK properties.^{12–16}

At its core, Gen AI for small-molecule discovery leverages deep learning models such as variational autoencoders (VAEs),^{17–20} generative adversarial networks (GANs),^{21–23} and reinforcement learning (RL)^{24–27} approaches to produce novel chemical structures with desired properties. These models are trained on large

datasets encoding known compound structures alongside biological or physicochemical attributes, enabling Gen AI to learn underlying rules of drug-likeness and bioactivity. Each of these tools (VAEs, GANs, and RLs) has strengths in drug discovery. First, VAEs can generate structurally diverse molecules and handle sparse datasets effectively, making them well-suited for chemical space exploration and lead optimization.²⁸ Second, GANs are useful for creating realistic, targeted molecular designs and optimizing specific properties, although they can be challenging to train and stabilize.²⁹ Third, RL excels at multi-objective, sequential tasks but needs significant computational resources and careful model setup.³⁰ Beyond VAEs, GANs, and RL, recent advances in Gen AI-driven drug discovery have been propelled by diffusion models, transformer-based architectures, graph-based models, equivalent neural networks, and large language models (LLMs). Representative examples of these models are shown in Table 1.

Overall, a major strength of generative models is their ability to explore chemical spaces beyond conventional libraries, proposing scaffolds that medical chemists may not yet have considered, including those that violate

Table 1. Examples of models representing current artificial intelligence methods for drug design

Method	Examples	Function in drug discovery	Limitations	References
Diffusion models	GeoDiff	<ul style="list-style-type: none"> (i) Learn the probabilistic evolution of molecular structures, enabling stable and high-fidelity three-dimensional (3D) molecule generation. (ii) Particularly powerful for modeling conformational diversity and binding poses. 	<ul style="list-style-type: none"> (i) High computational cost (ii) Scalability issues with large molecules (iii) Susceptibility to error accumulation in multi-stage processes (iv) Performance is also heavily dependent on the quality of training data and integrating external constraints 	15,31,32
Transformer-based approaches	SMILES transformer	<ul style="list-style-type: none"> (i) Leverage pretraining on SMILES to generate novel compounds and optimize for desired properties (e.g., bioactivity, absorption, distribution, metabolism, excretion, and toxicity) (ii) Effective for sequence-to-property translation and low-data scenarios 	<ul style="list-style-type: none"> (i) Reliance on large, time-consuming datasets (ii) Limited generalizability to diverse molecular structures (iii) The need for costly experimental validation of predicted compounds (iv) Limited adaptability to emerging targets and novel discovery 	33,34
Graph-based generative models	Graph-convolutional policy network (GCPN), GraphAF	<ul style="list-style-type: none"> (i) Model molecular graphs directly, preserving atomic connectivity and ensuring chemical validity (ii) GCPNs use reinforcement learning for property optimization; GraphAF uses autoregressive flows for efficient sampling 	<ul style="list-style-type: none"> (i) Over-smoothing with deep architectures, limitations in representing complex or dynamic molecular structures (ii) Inefficiencies in graph-based data storage and searchability (iii) Certain chemical structures, like coordination compounds or molecules with delocalized bonds, are poorly represented by static graphs 	16,35,36
Equivariant neural networks	E(n)-graph neural networks, equivariant diffusion	<ul style="list-style-type: none"> (i) Integrate physical symmetries (e.g., rotation, translation) for 3D-aware molecular modeling (ii) Crucial for tasks like protein-ligand docking, pose generation, and conformer sampling 	<ul style="list-style-type: none"> (i) High computational costs (ii) Limited availability of high-quality 3D molecular data (iii) Potential performance issues when the assumed symmetries do not match real-world data (iv) Redundant latent representations (v) Reduced stability in molecular generation (vi) Difficulties in interpretability and structure reconstruction 	37–39
Large language models	MolT5, ChemCrow	<ul style="list-style-type: none"> (i) Trained on vast chemical and textual corpora, enabling them to reason over chemical reactions, suggest synthetic routes, and propose novel compounds (ii) Useful for ideation and chemistry co-pilot tools 	<ul style="list-style-type: none"> (i) Tendency to generate answers regardless of accuracy, which can lead to errors or hallucinations (ii) Unintended biases when models are trained on a large dataset collected from the internet 	15,40–42

Lipinski's Rule of Five.^{43–48}

When coupled with target-specific scoring functions such as docking scores, predicted absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties, or ML-based activity predictions, these models can prioritize candidates, thereby streamlining hit identification and lead generation timelines. Although real-world validation outcomes for different Gen AI models remain limited, multiple benchmarking frameworks have been developed to compare subsets of VAE-, long short-term memory (LSTM)-, genetic algorithm (GA)-, and graph-Monte Carlo tree search (MCTS)-based models. One widely used example is the GuacaMol framework,⁴⁹ which defines both distribution-learning and goal-directed benchmarks (e.g., rediscovery, similarity, and multi-objective property optimization) and reports quantitative results for VAEs, ORGAN (VAE+RL), and SMILES LSTMs, making it a key reference for comparing VAE- and GAN/RL-style models.⁴⁹ Another example is DrugPose,⁵⁰ a more recent three-dimensional (3D) generative benchmark for early-stage drug discovery. This platform evaluates structure-based design methods by assessing how well generated ligands reproduce pose interaction patterns within binding sites and benchmarks diffusion models for 3D molecular structures and poses.⁵¹

Head-to-head quantitative comparisons within the same architecture class (e.g., two-dimensional 2D SMILES VAE versus 2D SMILES diffusion versus GAN versus RL) on GuacaMol-style tasks remain relatively sparse, as most diffusion-based studies target different benchmark classes. Consequently, cross-paper comparisons require careful interpretation.

As a practical takeaway to guide readers in selecting the most appropriate methods for their task, we note that VAEs and language-model based approaches perform strongly for distribution learning and library generation using 2D representations and are well characterized within the GuacaMol framework, making them reasonable starting points when reproducible benchmarks are required. For goal-directed optimization, RL-augmented models—such as ORGANstyle GANs, advantage actor-critic-based (A2C) generators, transformer based GAN + RL hybrids—and nonneural search methods (e.g., GA and MCTS) often achieve leading performance on constrained property optimization benchmarks. Finally, for structure-based and 3D molecular design, diffusion models and related 3D generative approaches currently show the greatest promise, with emerging benchmarks such as DrugPose explicitly designed to evaluate these models in early-stage drug discovery.

3.2. Proteomics

Concurrently, AI has begun to contribute significantly to the design of novel biologics, including antibodies, synthetic enzymes, cytokines, growth factors, hormones, peptides, fusion proteins, and nucleic acid-based therapeutics. AI-driven protein engineering platforms, such as Google's AlphaFold^{52–54} and the University of Washington's RoseTTAFold^{55–58}, have fundamentally reshaped *de novo* protein design by enabling high-accuracy prediction of 3D primary and secondary protein structures and their chemical properties at unprecedented scale and speed. Tasks such as simulating protein-protein interactions, predicting mutation impacts, and exploring novel motifs are no longer bottlenecked by time-intensive crystallography or cryo-electron microscopy, opening new frontiers in computational biology.

3.2.1. Challenges

In the context of proteomics, accurate protein structure prediction represents only the first step. Translating structural data into clinically viable therapeutics requires integrated knowledge of experimental biology, functional assays, cellular context, and PKs. Protein function depends on dynamic environmental factors—such as tissue-specific expression, post-translational modifications, and interactional networks—which static models cannot capture. Thus, validating the biological relevance and therapeutic potential of predicted structures remains a major challenge. Furthermore, the biological basis of disease extends beyond individual proteins to the interplay of the complex molecular systems, encompassing genomics (DNA), transcriptomics (mRNA), metabolomics, spatial biology, and immune microenvironments. Mapping and integrating these interdependencies are essential for understanding disease pathways, heterogeneity in treatment response, and off-target effects.

3.2.2. Responses

The next frontier in AI-driven biology lies in building multidimensional data sources and multiomics models that can integrate across biological scales and modalities. In response, pharmaceutical companies have spent the last decade investing in proprietary data banks and AI platforms tailored for drug discovery.^{59–62} These internal systems are supported by expansive structural, functional, and pharmacological data drawn from both in-house experiments and public sources. Leveraging these platforms has enabled companies to harness the combination of ML with high-throughput biology, such as automated imaging, Clustered Regularly Interspersed Short Palindromic Repeats screening, and single-cell sequencing, to build predictive models of disease mechanisms, drug-target

interactions, and phenotypic outcomes.⁶³ These models aim not only to predict molecular binding to a target but also to produce the desired effect within a specific biological context.

Heterogeneous datasets and multiomics models are still evolving and are being advanced in collaboration with a new generation of TechBIO companies.⁶⁴ Key efforts focus on capturing unknown biological hierarchies and simulating cellular systems to create predictive, generalized models of disease and therapeutic interventions. Currently, tangible impacts of system-level AI platforms on drug discovery are summarized in Table 2. Their integration into established discovery pipelines reflects a growing emphasis

on operational deployment rather than standalone proof-of-concept models.

As promising molecules are identified and advance to the lead optimization phase, key AI and ML applications relevant in such settings are shown in Table 3.

The success and sophistication of AI platforms and integrative models depend on the availability and integration of high-quality experimental datasets for training and validation, as well as rigorous grounding in real-world biology to produce biologically plausible and clinically relevant outputs. This ambitious vision, however, remains nascent. A few AI-derived drug candidates have entered early-phase clinical trials, but to date, no

Table 2. System-level AI platforms making tangible impacts in drug discovery

Application area	AI use cases	Required data types	Current achievements	Representative tools/companies
Target identification	Omics data mining, gene expression analysis, network biology, phenotypic screening	Genomic, transcriptomic, proteomic, pathway interaction data	Validated targets in oncology and rare diseases; target discovery platforms launched	Benevolent AI, Insitro, Atomwise, Insilico Medicine
Target validation	Multi-model integration of CRISPR screens, high-content imaging, multiomics to confirm target function	Functional genomics (CRISPR), imaging, knockdown assays, perturbation screens	Automated target validation pipelines; integration into early-stage discovery	Recursion, Verge Genomics, Deep Genomics
Small molecule generation	Generative models (e.g., GANs, VAEs) to create novel chemical scaffolds and drug-like molecules	Chemical libraries, activity data, molecular descriptors, 3D structures	Several AI-designated molecules entered Phase I trials (e.g. DSP-1181)	Insilico Medicine, Exscientia Iktos
Biologics design	Diffusion-based <i>de novo</i> protein engineering, <i>de novo</i> antibody and enzyme design using structure-function learning	Protein sequences, 3D structures, epitope data, immunogenicity assays	Proof-of-concept protein scaffolds and antibodies designed; some are in the preclinical stage	AbSci, Generate Biomedicines, BioNeMo (NVIDIA), Insilico Medicine
Lead optimization	QSAR modeling, reinforcement learning for activity/selectivity tuning	SAR data, molecular docking outputs, bioassay results	Improved hit-to-lead timelines, AI-driven compound libraries adopted in pharma	Cyclica, Reverie Labs, Relay Therapeutics, Insilico Medicine, Astellas
ADMET prediction and optimization	Product absorption, distribution, metabolism, excretion, and toxicity using ML algorithms; optimize molecules with generative AI	<i>In vitro/in vivo</i> pharmacokinetic data, toxicology reports, metabolic profiles	<i>In silico</i> ADMET is widely used in early screening; enhanced compound prioritization	Deep Chem, Aitia, Insilico Medicine, Astellas
Biomarker discovery	Multiomic feature selection, pathway enrichment models, and deep learning for novel biomarker identification	Clinical omics data, disease phenotypes, longitudinal patient data	AI-discovered biomarkers used in early diagnosis and response prediction	Owkin, Tempus AI, Freenome
Patient stratification	Clustering algorithms and ML-based risk scores for enrolling appropriate patient subgroups	EHRs, clinical trial data, omics profiles, treatment outcomes	Used in trial design and patient cohort selection, improving responder rate predictions	Unlearn, AI inference, GNS Healthcare

Abbreviations: 3D: Three-dimensional; ADMET: Absorption, distribution, metabolism, excretion, and toxicity; AI: Artificial intelligence; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; EHRs: Electronic health records; ML: Machine learning; QSAR: Quantitative structure–activity relationship; SAR: Structure–activity relationship.

Table 3. Key artificial intelligence and machine learning applications in lead optimization

Application	References
QSAR modeling to predict biological activity from molecular descriptors.	65–67
ADMET prediction to assess absorption, distribution, metabolism, excretion, and toxicity risks before synthesis.	68–71
Retrosynthesis planning using models such as the molecular transformer to guide the synthetic feasibility of proposed analogs.	72–75
Automatic compound prioritization, integration of predictions from multiple models into a composite score for decision-making.	76–78

Abbreviations: ADMET: Absorption, distribution, metabolism, excretion, and toxicity; QSAR: Quantitative structure–activity relationship.

AI-designed molecules have secured regulatory approval.⁷⁹ High-profile collaborations between pharmaceutical giants and AI-native biotech firms remain ongoing, representing an important test case for evaluating the performance of new-age AI-driven platforms.⁸⁰

Drug development is inherently iterative and interdisciplinary. Although discovery-based AI systems and applications have made significant progress, there remains a critical need for a collaborative interdisciplinary approach. The successful translation of these advancements into clinically differentiated, approved therapies will rely heavily on seamless integration with clinical development activities.

3.3. Clinical development

To realize their full potential as clinically differentiated approved drugs, AI-generated candidates must be integrated and synchronized with clinical development efforts to enable proactive knowledge translation and improve the likelihood of clinical success. Key requirements include validating biological rationale, target engagement, chemical profiling, and ADMET properties of AI-generated drug candidates through interdisciplinary collaboration between discovery and clinical teams. Furthermore, AI's predictive, analytical, and personalized medicine serve as important drivers across multiple independent clinical development domains, enabling the exploration of novel use cases. The following sections summarize key areas of AI-enabled clinical development.

3.3.1. Overview: Clinical use cases

AI-discovered drugs must undergo all standard drug development phases in compliance with the United States (U.S.) Food and Drug Administration (FDA) regulations to ensure patient safety, efficacy, and an appropriate benefit-to-risk assessment.⁸¹ Accordingly, recent FDA guidance on AI in clinical development emphasizes the need for validated algorithms, representative training datasets, well-documented workflows, and a clear scientific rationale for any AI-driven regulatory submissions across

all phases of clinical development.⁸² The key areas of AI impact in clinical development include, but are not limited to (i) pharmacology, (ii) molecular profiling for precision medicine, (iii) patient recruitment, (iv) trial design optimization, (v) clinical operations, and (vi) medical writing. These AI spheres are largely influenced and interdependent with other disciplines, where AI serves as a complementary tool rather than a central pillar.

3.3.2. Pharmacology

Artificial intelligence is redefining both preclinical and clinical pharmacology by enabling efficient animal testing, reducing the translational gaps between preclinical species and patients, and accelerating pipeline progression. In preclinical development, AI has used chemical structure descriptors and *in vitro* data to improve efficiency and/or accuracy in predicting human PKs and ADMET properties.^{83–85} A recent example applies a hybrid ML-physiologically-based PK approach to predict human PK, achieving robust accuracies. Among 106 small molecules tested, prediction accuracies reached 80–90% for 5-fold error and 40–60% for 2-fold error margin for both area under the curve and maximum (peak) plasma concentration.⁸⁶ These findings demonstrate the potential to minimize the use of animal models, thereby reducing costs and accelerating the development timeline. Additionally, the FDA SafeAI initiative and the AnimalGan platform have demonstrated a growing commitment to deploying AI to simulate organ-specific and human toxicities, with the potential to reduce or even waive requirements for animal toxicology testing.⁸⁷ These efforts are imperatively aligned with the FDA's recent release of a roadmap to reduce animal testing in preclinical safety studies,⁸⁸ which outlines the phased reduction of animal testing requirements for monoclonal antibodies and other drugs. In combination with *in vitro* assays, organoid chips, and computational modeling, including validated AI/ML methods, these approaches hold promise for more accurate human safety prediction, reduction, refinement, and replacement of animal use, and accelerated delivery of

new therapies to patients.⁸⁹

In clinical pharmacology, AI has demonstrated efficiency in covariate screening, as well as in the development of population PKs and exposure–response (ER) model structures, compared with traditional manual, labor-intensive processes, thereby accelerating the development of the final model.^{90–92} Additionally, the precision and accuracy of PK/pharmacodynamic (PD) and ER models have improved relative to conventional ER models, particularly in handling high-dimensional data and non-linear ER relationships.⁹³ Furthermore, AI-enabled digital twins (DTs), used in conjunction with quantitative systems pharmacology models, also show early promise in complementing clinical trial investigations, deepening understanding of mechanism of action, and supporting dose optimization strategies.^{94–96}

3.3.3. Molecular profiling and precision medicine

The growing landscape of emerging technologies, biomarker-directed therapies, complex drug mechanism classes, and targeted combination therapies has created multiple opportunities for AI clinical development use cases. AI-based neural network frameworks have been used to identify predictive—beyond purely prognostic—biomarkers from clinicogenomic datasets in both oncology and non-oncology indications.⁹⁷ LLMs trained on extensive molecular profiling data also show promise in accurately predicting alterations (e.g., oncogenic, mutational, overall mutational burden), which can enhance diagnostic accuracy and improve patient stratification and selection in clinical trials.^{98,99}

The identification and detection of actionable mutations, with subsequent matched therapeutic intervention, have been linked to improved patient outcomes and extended survival across multiple tumor types,¹⁰⁰ including in the setting of therapeutic resistance.¹⁰¹ For example, in colorectal cancer, combination therapies involving small-molecule targeted agents include cetuximab,¹⁰² which was initially approved for patients with *KRAS* wild-type, *EGFR*-expressing colorectal cancer and has more recently been approved in combination with encorafenib for BRAFV600E.¹⁰³ Beyond treatment matching as described above, AI integration remains an active and rapidly evolving area in oncology precision medicine. Traditional manual database searches could incorporate AI-based screening to enable broader and more efficient clinical literature reviews, thereby informing R&D decision-making.^{104,105} AI algorithms are expected to enhance their ability to predict patient outcomes and monitor treatment responses across clinical data input and model types, supporting the refined development of targeted therapies in oncology and other

therapeutic areas.

3.3.4. Patient recruitment and retention

Patient recruitment and identification of eligible patients are crucial steps in the successful design, execution, and completion of a clinical trial. Key efforts to facilitate trial recruitment include eligibility assessments through patient interviews, physical examinations, and analysis of real-world data (RWD). Furthermore, enrollment success can be enhanced by leveraging AI across four main modalities: (i) development and application of GenAI-based matching algorithms to align appropriate patients with trials' inclusion criteria;¹⁰⁶ (ii) analysis of patient data in unstructured electronic medical record (EMR) medical notes through natural language processing (NLP) using rule-based or ML approaches;¹⁰⁷ (iii) integration of medical knowledge into trained models to enhance contextual accuracy and reasoning of model outputs;¹⁰⁸ and (iv) adoption of advanced learning techniques—particularly neuronal networks—to understand, interpret, and generate natural language in LLMs.^{109,110} Each of these AI-based modalities comes with challenges and interdisciplinary dependencies to ensure feasible and cost-effective execution.

Large language models have shown superior effectiveness in extracting classified data from unstructured medical notes compared to conventional NLP- or ML-driven algorithms.¹¹¹ However, LLMs require substantial computing power and face privacy issues. Therefore, their widespread adoption remains uncertain given the current limitations in resources, knowledge, and the regulatory landscape in most institutions.

With respect to NLP-based approaches, their feasible and successful execution in real-world practice is often constrained by limitations in source data. EMRs used for NLP-driven patient selection are not deemed comprehensive data sources, as they are primarily designed for billing purposes and therefore provide an incomplete representation of patient care. The lack of longitudinal data and the unstructured nature of medical notes often lead to substantial missing data and inconsistencies, resulting in AI model drift, algorithmic biases, and inadequate patient profiling.^{112,113}

To optimize patient recruitment capabilities, contemporary EMRs increasingly need to be integrated and linked with multimodal data sources spanning insurance claims, hospital data, specialty labs, and omics data to ensure comprehensive yet targeted patient profiling and trial-specific outputs. Over the past 15 years, pharmaceutical companies, despite making significant advances in RWE, have continued partnering with contract research organizations, RWE providers, and

health systems to develop proprietary, comprehensive “fit-for-purpose” RWE data platforms capable of optimally profiling study subjects and supporting successful patient recruitment.^{114,115} Until high-quality multimodal RWE datasets are widely available, inconsistent and fragmented RWDs will continue to challenge the development of reliable, unbiased AI models for patient recruitment.

Beyond recruitment, patient retention represents a key patient-centric area where AI is increasingly being used in clinical trials. By analyzing LLM-based patient behavior characteristics across different disease cohorts and leveraging data from both trial and real-world sources, AI-based predictive algorithms can now identify patients at risk of dropping out versus those who are likely to remain study-compliant. Several AI companies focused on such patient-centered research are now working with pharmaceutical firms to develop study-specific, predictive behavior assessments and strategies that boost patient engagement and retention in clinical trials.^{116–118}

3.3.5. Trial design optimization

Currently, AI is increasingly recognized as a key enabler for optimizing trial design. Generative AI-enabled platforms can inform several key areas of clinical development, including protocol feasibility, patient enrichment, value-based endpoints, and the competitive positioning of the clinical development strategy.^{118,119} At the same time, such AI modalities are often used in conjunction with RWE and digital technologies to deliver practical and cost-effective enhancements to clinical trials. The open-source AI tool Trial Pathfinder,¹¹¹ for example, used EMRs to simulate data from completed non-small cell lung cancer trials, and demonstrated that inclusion criteria could be broadened without compromising statistical power. Deep learning algorithms have also been applied to numerous aspects of trial design, including patient flow modeling, effect size estimates, and endpoint optimization.¹²⁰ However, most efforts are still limited to pilot studies or proof-of-concept stages. Mainstream adoption of AI should follow a needs-based assessment, tested alongside conventional technologies to evaluate added value and operational feasibility for pharmaceutical stakeholders.

Effective AI-driven clinical trial optimization relies critically on access to high-quality, multimodal RWD for model training and validation. Pharmaceutical companies with robust RWE capabilities and diverse RWD platforms can better understand disease characteristics and patient experiences, enabling them to generate reliable insights for Gen AI-based trial design simulations.^{114,115,121} Conversely, single-sourced, non-representative, or low-quality RWE-based AI models can lead to inherent biases and skewed

predictions in trial design simulations. Hence, clinical development teams must prioritize RWE capabilities and consider AI within the broader RWD/RWE ecosystem, rather than relying solely on AI's black-box capabilities and computational inertias.

3.3.6. Digital twins in clinical trials

The widely publicized AI concept of DTs in clinical development involves an AI-generated virtual representation of a patient that can be used to predict prognosis, disease progression, and treatment outcomes. AI models in this area are primarily biological- and clinical-data-centric, leveraging multimodal inputs from large-scale bioclinical datasets to predict clinicopathological profiles, patient outcomes, and disease trajectories that closely approximate real-world progression patterns. A longitudinal progression study by Fisher *et al.* demonstrated that an unsupervised ML model could generate Alzheimer's disease trajectories aligned with those observed in real patients.¹²¹ More recently, Sanofi evaluated the efficacy of an asthma drug using DTs that integrated extensive data on disease biology and known pharmacology, with “blinded” model predictions closely matching outcomes from a Phase 1b study.¹²²

Digital twins must integrate comprehensive datasets encompassing patient-centered disease biology, pathophysiology, and known pharmacology data (when applicable) into a single computational framework to avoid bias, inaccurate projections, and uncertainty in model parameters. Furthermore, DT-driven AI models require highly validated, representative, up-to-date, and biomedically specific data sets as inputs. Given these stringent prerequisites, current DT approaches are likely to perform optimally in disease areas with well-characterized biology and a relatively simple or single etiological source, where the biological basis is well understood and fit-for-purpose, multidimensional data are available. In practice, few diseases exhibit such clearly defined etiologies or benefit from access to high-quality, augmented datasets. A recent collaboration among the U.S. National Science Foundation, National Institutes of Health, and FDA aims to develop DT frameworks and methodologies to promote biomedical innovation and clinical research.^{123,124} Similarly, the European Medicines Agency's (EMA) five-year AI Action Plan, launched in 2023, includes initiatives to advance DT technology through the development of specific tools, techniques, and integration of RWE.^{125,126} Given current technical limitations, data constraints, and incomplete understanding of disease biology, DTs are likely to remain applicable to specific disease areas and use cases until significant advances are achieved in this field.

3.3.7. External or synthetic control arms and regulatory drug approvals

Synthetic or external control arms (SCAs) simulations typically leverage Bayesian deep learning and GAN models and are most commonly used in clinical trials for research advancement purposes such as cost-effectiveness modeling, meta-analyses, and indirect treatment comparisons. These approaches are used both internally, for competitor analysis, and externally, to show the medico-economic value of therapies to Health Technology Authorities and payer organizations.¹²⁷

Artificial intelligence-driven SCAs have been increasingly publicized in the AI world as a promising application in clinical trials to support regulatory drug approvals, with some proponents suggesting their potential to partially replace randomized controlled trials.¹²⁸ However, the regulatory positioning of SCAs remains highly stringent and typically requires a custom historical control or RWD set that can accurately simulate the placebo arm according to key trial conditions and protocol. The limited availability of such tailored datasets, together with the current RWE landscape—characterized by disparate datasets and heterogeneous patient trajectories—hinders rigorous assessment and bias-free computation of control arms across multiple dimensions. As a result, regulators generally discourage reliance on AI-generated external or synthetic control data for primary evidence in drug approval decisions, thereby limiting their widespread adoption for regulatory purposes. Between 2000 and 2019, the FDA included external control data in only 45 drug approvals for benefit–risk assessments, primarily in rare disease or oncology settings or in cases where ethical concerns precluded the use of a placebo or no-treatment arm.¹²⁹ Similarly, the EMA accepted external control data in only 18 (17%) of 103 approved oncology submissions between 2016 and 2021, with 37% of submissions rejected due to study design issues and methodological inadequacies.¹³⁰ Consequently, any proposed use of SCAs as part of a regulatory strategy or drug approval should be guided by FDA recommendations on external control arms against the backdrop of a rigorous data-oriented feasibility assessment before implementation.¹³¹

Given the strong interdependence of RWE and AI in clinical development, [Figure 2](#) shows the integration of AI applications across the clinical development lifecycle against the backdrop of a multimodal RWE framework. The figure highlights key dataset types that constitute a multimodal RWE platform and provides a comprehensive framework for understanding how AI and RWE can be

optimally combined to create a robust, AI-based clinical development platform for designing patient-centric trials and facilitating regulatory-grade data.

3.3.8. Clinical operations and medical writing

Artificial intelligence-powered co-pilots combined with digital health demonstrate a promising role in improving efficiency and peak performance in trial management. Pharmaceutical companies are increasingly using AI tools with NLP and LLM capabilities for auto-drafting medical and trial documents, such as informed consent forms, site initiation visit materials, clinical study reports, and regulatory submissions.^{132–134} Furthermore, these tools are employed to predict trial operations and finances using risk-monitoring platforms,^{135–137} as well as to deploy AI co-pilots in digital patient portals for patient monitoring and engagement.¹³⁸

Robotic process automation and ML-based AI tools are enhancing pharmacovigilance (PV) by programming and presetting PV systems for data processing, auto-coding, auto-narrative generation, literature searches, and regulatory reporting, thereby freeing PV professionals for more analytical, medically oriented, and strategic tasks.^{139,140} Digital remote data capture with AI interfaces also has the potential to boost patient accessibility, enhance patient engagement and patient centricity, and improve data quality and timelines in real time.¹⁴¹

Thousands of tables, listings, and figures (TLFs) summarizing and presenting data lie at the core of Phase 2 and Phase 3 clinical trials. AI tools are now automating the creation and validation of TLFs, generating biostatistical builders, and facilitating clinical trial integrations.^{142–145} Data visualization and AI-enabled analytical tools, such as Spotfire Copilot™,^{146–148} are becoming increasingly mainstream in clinical trials to efficiently visualize complex data and support multimodal data analytics.

In summary, AI tools that enable smart data management, analytics, and digital integration can reduce administrative burdens, expedite study execution, and augment clinical intelligence, thereby potentially increasing the likelihood of clinical success.

In the evolving R&D landscape, clinical development, data science, governance, and bioethics are inextricably linked, each playing a crucial role within the biopharma ecosystem to ensure the safe, efficient, and responsible development of new treatments. Hence, technological efforts to integrate AI into drug development must interface with these disciplines in a cogwheel-like manner to foster strategic growth as well as a culture of learning and regulatory compliance.

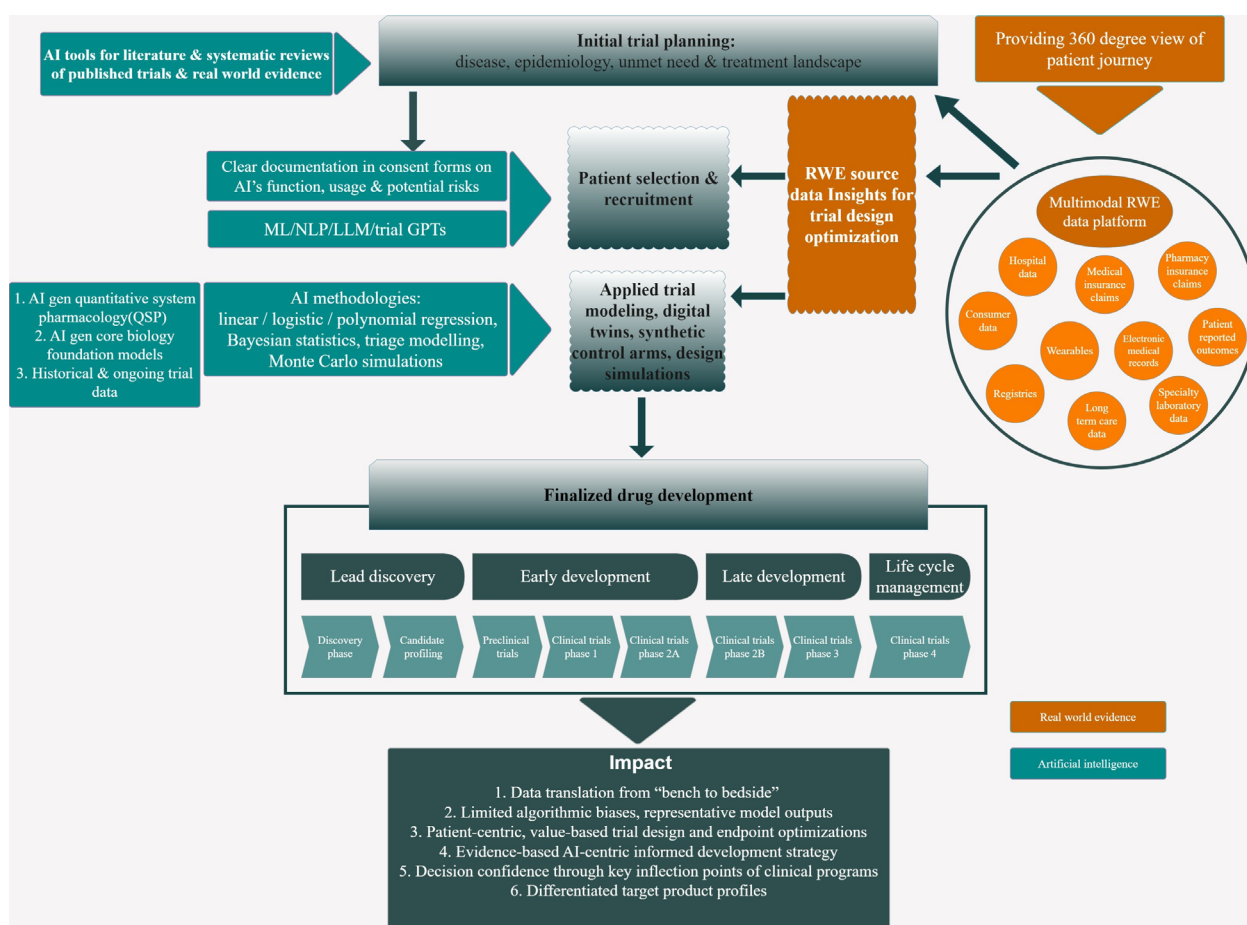


Figure 2. Integration of AI and RWE to create a robust, AI-based clinical development platform for regulatory use. Image created by the authors. Abbreviations: AI: Artificial intelligence; LLM: Large language model; ML: Machine learning; NLP: Natural language processing; RWE: real-world evidence; Trial GPTs: Trial-specific generative pretrained transformers.

3.4. Data, science, governance & ethics

Artificial intelligence integration is most effective when a dynamic R&D ecosystem encourages platform-based innovation through connected components, promoting flexible learning and skill development. Achieving this requires coordination of AI data strategies with regulatory compliance, domain expertise, scalability, and ethical standards. Key aspects of robust AI governance include:

- (i) Thorough data stewardship, ranging from data provenance to data integrity, validity, quality, privacy, and security, aligned with the latest standards from the National Institute of Standards and Technology (NIST), International Organization for Standardization (ISO), and General Data Protection Regulation (GDPR).^{149,150}
- (ii) Artificial intelligence risk reduction across the entire AI deployment lifecycle, with heightened focus on potential risks to patient security, safety, and clinical interests that can occur from

incorrect AI model interpretations. Such risks may occur during compound screening, target validation, biomarker discovery, and early-stage trial activities.

- (iii) Proactively addressing regulatory challenges by embedding regulatory-first thinking into AI workflows to ensure models are fit for clinical use and suitable for supporting drug approvals. This includes incorporating regulatory and compliance checkpoints early in development, supported by detailed documentation on testing and monitoring.
- (iv) Adopting the Consolidated Standards of Reporting Trials–Artificial Intelligence (CONSORT-AI) guidelines for reporting clinical trials involving AI interventions. These evidence-based reporting standards advocate for the comprehensive documentation of AI-enabled trials, outlining 14 essential aspects, including integration settings,

input-output mapping, human-AI interaction, and error analysis, to promote rigorous and transparent evaluation of AI applications.¹⁵¹

Ethical drug discovery should be inclusive and equitable, adhering to principles that place patient welfare at the forefront. AI must serve as a supportive tool rather than a substitute for human expertise, with human-in-the-loop oversight integrated at critical decision points. Additionally, comprehensive monitoring of algorithms across multiple sources and robust stakeholder accountability are required to maintain ethical standards.¹⁵²

Artificial intelligence integration in R&D with its associated governance, operational, and bioethics framework is presented in Figure 3.

4. Discussion

Recent advancements in AI-driven drug discovery and clinical operations have shown promising quantitative improvements in speed and efficiency. Across the industry, AI is projected to reduce early-stage drug discovery timelines by 30–50% and lower costs by up to 25% over the next five years.¹⁵³

While technological advances can improve agility and efficiency, these benefits alone do not guarantee clinical or regulatory success, which depends on a plethora of

additional factors. Indeed, AI-enabled pharmaceutical development over the past decade has been marked by multiple unsuccessful trials, resulting in the shelving and discontinuation of several drug candidates.⁷⁹ To date, no AI-enabled drugs have secured regulatory drug approval.^{154,155}

Most chronic diseases involve complex causal relationships and significant transcriptional changes that affect the expression of many proteins and disease pathways. Decoding disease progression, simulating complex biological systems, and designing precise therapies, therefore, remain persistent challenges in R&D. AI systems require high-quality, relevant, and diverse datasets to address overfitting biases, learn effectively, and perform reliably. There is a systemic need for comprehensive, multimodal AI platforms capable of breaking data silos and integrating multidimensional biology, disease-related targets, precision chemistry, and clinical translation in an end-to-end manner. Such integration is essential for advancing AI-enabled drug development. Furthermore, these multimodal bio-omics-clinical data platforms must be harmonized with RWE and digital health systems to connect early discovery insights with therapeutic relevance and promote real-world optimization of development programs.

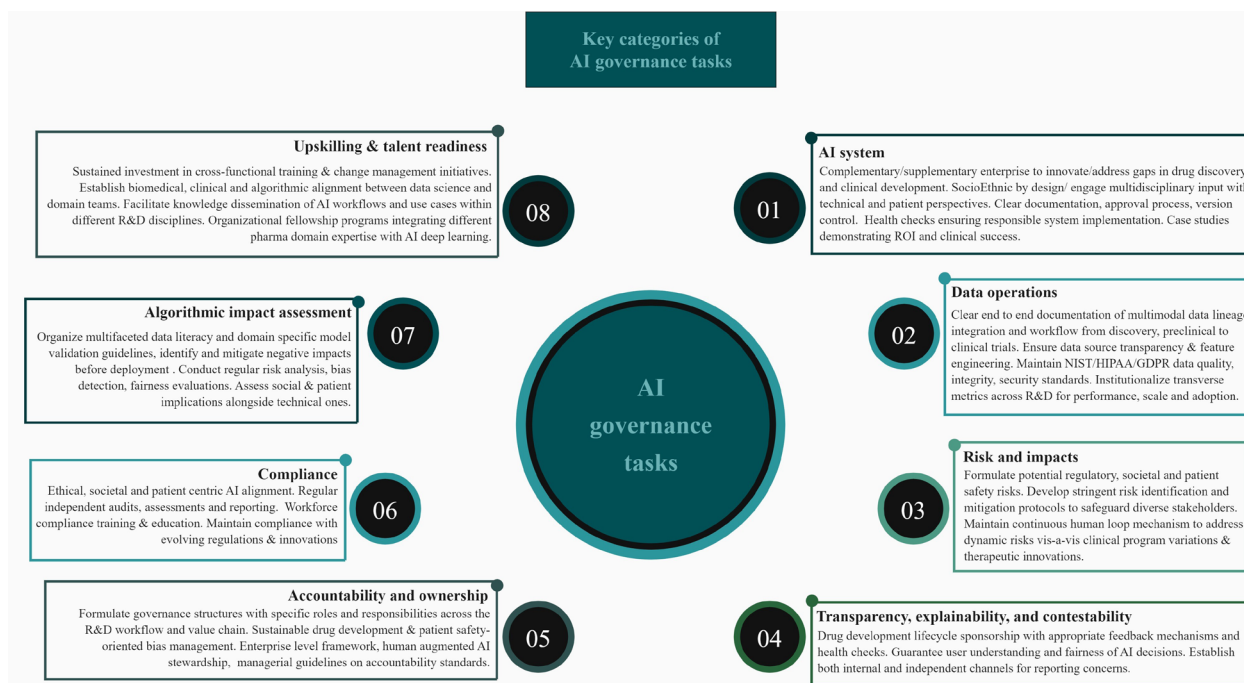


Figure 3. AI–R&D data science integration and governance framework. Image created by the authors.

Abbreviations: AI: Artificial intelligence; GDPR: General Data Protection Regulation; HIPAA: Health Insurance Portability and Accountability Act; NIST: National Institute of Standards and Technology; R&D: Research and development.

A foundational requirement for robust and generalizable AI in drug discovery and development is the establishment of a modern data integration architecture that unifies multimodal discovery, clinical, and real-world datasets. Interoperability frameworks such as the OMOP Common Data Model and HL7 FHIR—now bridged through the HL7 Vulcan FHIR-to-OMOP initiative—enable consistent data lineage, traceability, and scalable analytics across heterogeneous data sources.^{156,157} Federated learning frameworks such as MELLODDY and the AI Structural Biology network enable secure cross-company

model training without exposing proprietary data, providing an industrial-scale pathway for collaborative AI development.^{158,159} Collectively, these frameworks, detailed in Table 4, form a regulator-aligned, privacy-preserving data backbone that empowers AI systems to learn from richer, higher-quality evidence and supports improved decisionmaking across the R&D continuum.

In the next three to five years, investments in and adoption of Gen AI across foundational biology, target identification, and lead generation are expected to grow

Table 4. Existing frameworks and standards addressing data accessibility and integration in artificial intelligence (AI)-enabled drug discovery

Frameworks	Applications	References
OMOP Common Data Model and HL7FHIR bridged through the HL7 Vulcan FHIR→OMOP initiative	Ensure consistent data lineage, traceability, and scalable analytics across heterogeneous data sources.	156,157
CDISC–FHIR Joint Mapping Implementation Guide and TransCelerate’s Digital Data Flow (USDM)	Enable protocol digitization, automated study-asset generation, and seamless eSource-to-SDTM transformations across clinical systems.	160,161
FAIR data stewardship, operationalized through the ELIXIR Interoperability Platform and its associated resources (RDMkit, FAIRsharing, FAIR Cookbook)	Strengthen semantic harmonization and reproducibility for multiomics and real-world data integration.	162
GA4GH standards such as Phenopackets for structured phenotype representation and cloud-portable workflow APIs (e.g., WES/DRS)	Support consistent, reproducible computation across ecosystems.	163,164
BioCompute Objects (IEEE 2791–2020), supported by multiple FDA centers	Standardize computational workflow documentation to enhance transparency and regulatory review of AI-derived analyses.	165,166
Federated learning frameworks such as MELLODDY and the AI Structural Biology network	Enable secure, cross-company model training without exposing proprietary data, providing an industrial-scale path for collaborative AI development.	158,159

exponentially. At the same time, AI will continue to be steadily implemented and integrated into clinical trials, although outcomes may vary due to limited availability of AI-proficient talent in biopharma, stringent regulatory requirements, and challenges related to integration within existing R&D workflows. Nevertheless, advances in system-level biology, AI-enabled chemistry, high computational inertia, and substantial meteoric investments in Gen AI-centric R&D integration are collectively demonstrating that it is possible to identify new cures at a faster pace and greater scale than previously achievable—albeit with cautious optimism. For example, a fully end-to-end AI-designed antifibrotic inhibitor for idiopathic pulmonary fibrosis achieved positive Phase 2a clinical trial results, representing the first clinical evaluation of a fully AI-enabled drug with convincing proof of concept.¹⁶⁷

To bridge the gap between promising early findings and the ultimate goal of delivering safe and effective therapies to patients, several courses of action are warranted:

- (i) Continuous-loop learning and prudent knowledge dissemination across discovery, preclinical, and clinical development domains should be implemented.
- (ii) Adoption of a regulatory-first thinking and a phase-wise approach with clearly defined objectives for AI implementation is essential. Targeted use cases should be identified that correspond with the highest levels of decision confidence, therapeutic expertise, and organizational understanding of disease biology.
- (iii) Clinical trials should integrate AI-enabled drug discovery with early-stage feedback and optimization mechanisms.¹⁶⁸
- (iv) Enterprise data strategies in clinical development ought to evaluate the return on investment (ROI) of AI in conjunction with progress in RWE initiatives and digital health platforms, which often deliver significant value independently.

This allows judicious assessment of operational feasibility, computational costs, and data retrieval effectiveness for each application.

- (v) Enhanced collaboration among industry, academia, and regulators is needed to accelerate enterprise-level AI advancement through federated learning initiatives and the development of best practices for AI implementation in drug development.¹⁵⁸

5. Conclusion

Artificial intelligence in pharmaceutical drug development presents both significant opportunities and intrinsic challenges that require careful consideration to ensure regulatory compliance and uphold scientific integrity. Comprehensive integration of AI across drug development remains an evolving process rather than a revolutionary shift. AI can serve as an important pillar in the rapidly changing landscape of drug development; however, its successful integration into R&D depends more on the development of multimodal data platforms, effective integration with existing pharma processes, and use-case-driven demonstrations of ROI than a blanket spread of computational inertia, hopes, and expectations. It is imperative that AI technologies are applied ethically and responsibly, with stringent human oversight throughout the drug discovery and development lifecycle, and with a heightened focus on fairness, transparency, and patient welfare.

Acknowledgements

We gratefully acknowledge the contributions of the speakers and panelists from our 2024 monthly Pharma Leaders DCT series, whose insights informed the development of this paper. We also express our gratitude to Dr. Anthony Rosner for his medical writing assistance during the submission phase. Additionally, we thank Fion Liao, Aditya Tallapragada, and Rashmi Gurnani from AKT Health for their project management.

Funding

None.

Conflict of interest

Sheikh Usman Iqbal is a CMO consultant for Julz Pharma. Dmitriy Podolskiy is an employee of Astellas Pharma. Mona G. Flores is self-employed. Victoria L. Chiou is an employee of Tempus AI, Inc. Roberto Araujo is an employee of Sanofi. Naveed Afzal is an employee of Takeda Pharmaceutical Company Limited. David A. Hall is an employee of AKT Health Inc. The authors declare no other competing interests beyond the relationships stated above.

Author contributions

Conceptualization: All authors

Visualization: Naveed Afzal, David A. Hall, Ingrid Vasiliu-Feltes, Dipu Patel, Sheikh Usman Iqbal

Writing-original draft: All authors

Writing-review & editing: All authors

The manuscript was written and reviewed by different authors according to their respective sections. Dmitriy Podolskiy, Mona G. Flores, and Manolis Kellis were responsible for: Drug discovery section; Sheikh Usman Iqbal, Victoria L. Chiou, Jennifer Sheng, and Roberto Araujo were responsible for: Clinical development section; Naveed Afzal, David A. Hall, Ingrid Vasiliu-Feltes, and Dipu Patel were responsible for: Data, science, management & ethics section. All authors reviewed and contributed to the discussion and conclusion sections.

Ethics approval and consent to participate

Not applicable.

Consent for publication

No applicable.

Availability of data

No datasets were generated or analyzed during the current study.

References

- Hall DA. *DCT Event Series: Fostering Innovation and Collaboration in Decentralized Clinical Trials*. LinkedIn post. 2024. Available from: https://www.linkedin.com/posts/novative1_akthealth-pharmaleaders-dct-activity-7211452184587071489-uvRR [Last accessed on 2025 Dec 28].
- Kirkpatrick Partners. *The Kirkpatrick Model*. 2024. Available from: <https://www.kirkpatrickpartners.com/the-kirkpatrick-model/> [Last accessed on 2025 Dec 11].
- Vincent F, Nueda A, Lee J, *et al*. Phenotypic drug discovery: Recent successes, lessons learned and new directions. *Nat Rev Drug Discov*. 2022;21(12):899-914. doi: 10.1038/s41573-022-00472-w
- Moffat JG, Vincent F, Lee JA, *et al*. Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nat Rev Drug Discov*. 2017;16(8):531-543. doi: 10.1038/nrd.2017.111
- Wassermann AM, Camargo LM, Auld DS. Composition and applications of focus libraries to phenotypic assays. *Front Pharmacol*. 2014;5:164. doi: 10.3389/fphar.2014.00164

6. Hughes JP, Rees S, Kalindjian SB, *et al.* Principles of early drug discovery. *Br J Pharmacol.* 2011;162(6):1239–1249.
doi: 10.1111/j.1476-5381.2010.01127.x
7. Zhu T, Cao S, Su PC, *et al.* Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis. *J Med Chem.* 2013;56(17):6560–6572.
doi: 10.1021/jm301916b
8. Gao B, Qiang B, Tan H, *et al.* DrugCLIP: Contrastive protein-molecule representation learning for virtual screening. *arXiv.* Preprint online 2023.
doi: 10.48550/arXiv.2310.06367
9. Mbou Sob UA, Li Q, Arbesú M, *et al.* Generative Model for Small Molecules with Latent Space RL Fine-Tuning to Protein Targets. *arXiv.* Preprint online 2024.
doi: 10.48550/arXiv.2407.13780
10. Bian Y, Kwon JJ, Liu C, *et al.* Target-driven machine learning-enabled virtual screening (TAME-VS) platform for early-stage hit identification. *Front Mol Biosci.* 2023;10:1163536.
doi: 10.3389/fmolb.2023.1163536
11. Shen L, Feng H, Qiu Y, *et al.* SVSBI: Sequence-based virtual screening of biomolecular interactions. *Commun Biol.* 2023;6:536.
doi: 10.1038/s42003-023-04866-3
12. Ghayoor A, Kohan HG. Revolutionizing pharmacokinetics: The dawn of AI-powered analysis. *J Pharm Pharm Sci.* 2024;27:12671.
doi: 10.3389/jpps.2024.12671
13. Gangwal A, Lavecchia A. Unleashing the power of generative AI in drug discovery. *Drug Discov Today.* 2024;29(6):103992.
doi: 10.1016/j.drudis.2024.103992
14. Atz K, Cotos L, Isert C, *et al.* Prospective de novo drug design with deep interactome learning. *Nat Commun.* 2024;15(1):3408.
doi: 10.1038/s41467-024-47613-w
15. Tang X, Dai H, Knight E, *et al.* A survey of generative AI for de novo drug design: New frontiers in molecule and protein generation. *Brief Bioinform.* 2024;25(4):bbae338.
doi: 10.1093/bib/bbae338
16. Gangwal A, Ansari A, Ahmad I, *et al.* Generative artificial intelligence in drug discovery: Basic framework, recent advances, challenges, and opportunities. *Front Pharmacol.* 2024;15:1331062.
doi: 10.3389/fphar.2024.1331062
17. Zeng X, Wang F, Luo Y, *et al.* Deep generative molecular design reshapes drug discovery. *Cell Rep Med.* 2022;3(12):100794.
doi: 10.1016/j.xcrm.2022.100794
18. Ahmad M, Teli TA. De novo molecular generation augmentation for drug discovery using deep learning approaches: A comparative study of variational autoencoders. *J Angiother.* 2024;8(10):1–13.
doi: 10.25163/angiotherapy.8109996
19. Sousa T, Correia J, Pereira V, *et al.* Generative deep learning for targeted compound design. *J Chem Inf Model.* 2021;61(11):5343–5361.
doi: 10.1021/acs.jcim.0c01496
20. Samanta S, O'Hagan S, Swainston N, *et al.* VAE-Sim: A novel molecular similarity measure based on a variational autoencoder. *Molecules.* 2020;25(15):3446.
doi: 10.3390/molecules25153446
21. Macedo B, Ribeiro Vaz I, Taveira Gomes T. MedGAN: Optimized generative adversarial network with graph convolutional networks for novel molecule design. *Sci Rep.* 2024;14(1):1212.
doi: 10.1038/s41598-023-50834-6
22. Tripathi S, Augustin AI, Dunlop A, *et al.* Recent advances and application of generative adversarial networks in drug discovery, development, and targeting. *Artif Intell Life Sci.* 2022;2:100045.
doi: 10.1016/j.ailsci.2022.100045
23. Kao PY, Yang YC, Chiang WY, *et al.* Exploring the advantages of quantum generative adversarial networks in generative chemistry. *J Chem Inf Model.* 2023;63(11):3307–3318.
doi: 10.1021/acs.jcim.3c00562
24. Haddad R, Litsa EE, Liu Z, *et al.* Targeted molecular generation with latent reinforcement learning. *Sci Rep.* 2025;15(1):15202.
doi: 10.1038/s41598-025-99785-0
25. Xiong Y, Wang Y, Wang Y, *et al.* Improving drug discovery with a hybrid deep generative model using reinforcement learning trained on a Bayesian docking approximation. *J Comput Aided Mol Des.* 2023;37(11):507–517.
doi: 10.1007/s10822-023-00523-3
26. Wang Q, Wei Z, Hu X, *et al.* Molecular generation strategy and optimization based on A2C reinforcement learning in de novo drug design. *Bioinformatics.* 2023;39(11):btad693.
doi: 10.1093/bioinformatics/btad693
27. Bou A, Thomas M, Dittert S, *et al.* ACEGEN: Reinforcement learning of generative chemical agents for drug discovery. *J Chem Inf Model.* 2024;64(15):5900–5911.
doi: 10.1021/acs.jcim.4c00895
28. Gómez-Bombarelli R, Wei JN, Duvenaud D, *et al.* Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci.* 2018;4(2):268–276.

- doi: 10.1021/acscentsci.7b00572
29. De Cao N, Kipf T. MolGAN: An implicit generative model for small molecular graphs. *arXiv*. Preprint online 2018.
doi: 10.48550/arXiv.1805.11973
30. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv*. 2018;4(7):eaap7885.
doi: 10.1126/sciadv.aap7885
31. Xu M, Wang H, Hu Y, *et al*. GeoDiff: A geometric diffusion model for molecular conformation generation. *arXiv*. Preprint online 2022.
doi: 10.48550/arXiv.2203.02923
32. Xu M, Powers AS, Dror RO, *et al*. Geometric latent diffusion models for 3D molecule generation. *arXiv*. Preprint online 2023.
doi: 10.48550/arXiv.2305.01140
33. Honda S, Shi S, Ueda HR, *et al*. SMILES Transformer: Pre-trained molecular fingerprint for low data drug discovery. *arXiv*. Preprint online 2019.
doi: 10.48550/arXiv.1911.04738
34. Jiang J, Chen L, Ke L, *et al*. A review of transformer models in drug discovery and beyond. *J Pharm Anal*. 2025;15(6):101081.
doi: 10.1016/j.jpha.2024.101081
35. You J, Liu B, Ying R, *et al*. Graph convolutional policy network for goal-directed molecular graph generation. *arXiv*. Preprint online 2018.
doi: 10.48550/arXiv.1806.02473
36. Shi C, Xu M, Guo H, *et al*. GraphAF: A flow-based autoregressive model for molecular graph generation. *arXiv*. Preprint online 2020.
doi: 10.48550/arXiv.2001.09382
37. Satorras VG, Hoogeboom E, Welling M. E(n) equivariant graph neural networks. *arXiv*. Preprint online 2021.
doi: 10.48550/arXiv.2102.09844
38. Hoogeboom E, Satorras VG, Castañeda AG, *et al*. Equivariant diffusion for molecule generation in 3D. *arXiv*. Preprint online 2022.
doi: 10.48550/arXiv.2203.17003
39. Schneuing A, Harris C, Du Y, *et al*. Structure-based drug design with equivariant diffusion models. *Nat Comput Sci*. 2024;4(12):899–909.
doi: 10.1038/s43588-024-00737-x
40. Edwards C, Lai T, Ros K, *et al*. Translation between molecules and natural language. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP); Association for Computational Linguistics; 2022:375–413.
doi: 10.18653/v1/2022.emnlp-main.26
41. Rost B, Wang AY, Schwaller P. ChemCrow: A chemistry LLM agent powered by tools. *arXiv*. Preprint online 2023.
doi: 10.48550/arXiv.2304.05376
42. Bran AM, Cox S, Schilter O, Baldassari C, White AD, Schwaller P. Augmenting large language models with chemistry tools. *Nat Mach Intell*. 2024;6:525–535.
doi: 10.1038/s42256-024-00832-8
43. Vogt M. Exploring chemical space—Generative models and their evaluation. *Artif Intell Life Sci*. 2023;3:100064.
doi: 10.1016/j.aailsci.2023.100064
44. Vogt M. Using deep neural networks to explore chemical space. *Expert Opin Drug Discov*. 2022;17(3):297–304.
doi: 10.1080/17460441.2022.2019704
45. Bilodeau C, Jin W, Jaakkola T, *et al*. Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdiscip Rev Comput Mol Sci*. 2022;12(2):e1608.
doi: 10.1002/wcms.1608
46. Wang J, Mao J, Wang M, *et al*. Explore drug-like space with deep generative models. *Methods*. 2023;210:52–59.
doi: 10.1016/j.ymeth.2023.01.004
47. Anstine DM, Isayev O. Generative models as an emerging paradigm in the chemical sciences. *J Am Chem Soc*. 2023;145(16):8736–8750.
doi: 10.1021/jacs.2c13467
48. Gao W, Luo S, Coley CW. Generative artificial intelligence for navigating synthesizable chemical space. *arXiv*. Preprint online 2024.
doi: 10.48550/arXiv.2410.03494
49. Brown N, Fiscato M, Segler MH, *et al*. GuacaMol: Benchmarking models for de novo molecular design. *J Chem Inf Model*. 2019;59(3):1096–1108.
doi: 10.1021/acs.jcim.8b00839
50. Jocys Z, Grundy J, Farrahi K. DrugPose: Benchmarking 3D generative methods for early stage drug discovery. *Digit Discov*. 2024;3:1308–1318.
doi: 10.1039/D4DD00076E
51. Khater, T., Alkhatib, S.A., AlShehhi, A. *et al*. Generative artificial intelligence based models optimization towards molecule design enhancement. *J Cheminform* 2025;17:116.
doi: 10.1186/s13321-025-01059-4
52. Tunyasuvunakool K, Adler J, Wu Z, *et al*. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021;596(7873):590–596.
doi: 10.1038/s41586-021-03828-1
53. Marcu ȘB, Tăbircă S, Tangney M. An overview of AlphaFold's

- breakthrough. *Front Artif Intell.* 2022;5:875587.
doi: 10.3389/frai.2022.875587
54. Abramson J, Adler J, Dunger J, *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630(8016):493–500.
doi: 10.1038/s41586-024-07487-w
55. Baek M, DiMaio F, Anishchenko I, *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021;373(6557):871–876.
doi: 10.1126/science.abj8754
56. Baek M, McHugh R, Anishchenko I, *et al.* Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nat Methods.* 2024;21(1):117–121.
doi: 10.1038/s41592-023-02086-5
57. Krishna R, Wang J, Ahern W, *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science.* 2024;384(6693):ead12528.
doi: 10.1126/science.adl2528
58. Lisanza SL, Gershon JM, Tipps SWK, *et al.* Multistate and functional protein design using RoseTTAFold sequence space diffusion. *Nat Biotechnol.* 2025;43:1288–1298.
doi: 10.1038/s41587-024-02395-w
59. Profitt A. *AbbVie's R&D Convergence Hub.* *Bio-IT World.* 2023. Available from: <https://www.bio-itworld.com/news/2023/08/24/abbvie-s-r-d-convergence-hub> [Last accessed on 2025 Nov 03].
60. Amgen. *Generative Biology: Designing Biologic Medicines with Greater Speed and Success.* 2022. Available from: <https://www.amgen.com/stories/2022/06/generative-biology--designing-biologics-with-greater-speed-and-success> [Last accessed on 2025 Nov 04].
61. Abbvie. *Meet the ARCH: A Time-Saving Tool for Researchers Focused on Finding Cures.* 2023. Available from: <https://www.abbvie.com/who-we-are/our-stories/meet-the-arch-a-time-saving-tool-for-researchers-focused-on-finding-cures.html> [Last accessed on 2025 Nov 04].
62. *Decoding the Microcosm of Life, Advancing Human Health for All.* *Westlake Omics.* 2023. Available from: <https://www.westlakeomics.com/en/> [Last accessed on 2025 Nov 15].
63. Vamathevan J, Clark D, Czodrowski P, *et al.* Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* 2019;18(6):463–477.
doi: 10.1038/s41573-019-0024-5
64. Krummich, M. *TechBio Market Map: Navigating the Landscape of Next-Generation Biotech.* B2venture. 2024. Available from: <https://www.b2venture.vc/stories/techbio-market-map> [Last accessed on 2025 Nov 04].
65. Kolides A, Nawaz A, Rathor A, *et al.* Artificial intelligence foundation and pre-trained models: Fundamentals, applications, opportunities, and social impacts. *Simul Model Pract Theory.* 2023;126:102754.
doi: 10.1016/j.simpat.2023.102754
66. Cherkasov A, Muratov EN, Fourches D, *et al.* QSAR modeling: Where have you been? Where are you going to? *J Med Chem.* 2014;57(12):4977–5010.
doi: 10.1021/jm4004285
67. Lewis RA. A general method for exploiting QSAR models in lead optimization. *J Med Chem.* 2005;48(5):1638–1648.
doi: 10.1021/jm049228d
68. Swanson K, Walther P, Leitz J, *et al.* ADMET-AI: A machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics.* 2024;40:btac416.
doi: 10.1093/bioinformatics/btac416
69. Sohlenius-Sternbeck AK, Terelius Y. Evaluation of ADMET Predictor in early discovery drug metabolism and pharmacokinetics project work. *Drug Metab Dispos.* 2022;50(2):95–104.
doi: 10.1124/dmd.121.000552
70. Han H, Shaker B, Lee JH, *et al.* Employing automated machine learning (AutoML) methods to facilitate the *in silico* ADMET properties prediction. *J Chem Inf Model.* 2025;65(7):3215–3225.
doi: 10.1021/acs.jcim.4c02122
71. Kumar A, Kini SG, Rath E. A recent appraisal of artificial intelligence and *in silico* ADMET prediction in the early stages of drug discovery. *Mini Rev Med Chem.* 2021;21(18):2788–2800.
doi: 10.2174/1389557521666210401091147
72. Schwaller P, Petraglia R, Zullo V, *et al.* Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem Sci.* 2020;11:3316–3325.
doi: 10.1039/C9SC05704H
73. Wang Y, Pang C, Wang Y, *et al.* Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks. *Nat Commun.* 2023;14(1):6155.
doi: 10.1038/s41467-023-41698-5
74. Nakamura S, Yasuo N, Sekijima M. Molecular optimization using a conditional transformer for reaction-aware compound exploration with reinforcement learning. *Commun Chem.* 2025;8(1):40.
doi: 10.1038/s42004-025-01437-x
75. Bu Y, Gao R, Zhang B, *et al.* CoGT: Ensemble machine learning method and its application on JAK inhibitor discovery. *ACS Omega.* 2023;8(14):13232–13242.
doi: 10.1021/acsomega.3c00160

76. Bleicher LS, van Daelen T, Honeycutt JD, *et al.* Enhanced utility of AI/ML methods during lead optimization by inclusion of 3D ligand information. *Front Drug Discov.* 2022;2:1074797.
doi: 10.3389/fddsv.2022.1074797
77. Yonchev D, Bajorath J. Integrating computational lead optimization diagnostics with analog design and candidate selection. *Future Sci OA.* 2020;6(3):FSO451.
doi: 10.2144/fsoa-2019-0131
78. Tosh C, Tec M, White JB, *et al.* A Bayesian active learning platform for scalable combination drug screens. *Nat Commun.* 2025;16(1):156.
doi: 10.1038/s41467-024-55287-7
79. Mak K-K, Pichika MR. Artificial intelligence in drug development: Present status and future prospects. *Drug Discov Today.* 2019;24(3):773–780.
doi: 10.1016/j.drudis.2018.11.014
80. Kint S, Dolfsma W, Robinson D. Strategic partnerships for AI-driven drug discovery: The role of relational dynamics. *Drug Discov Today.* 2024;29(12):104242.
doi: 10.1016/j.drudis.2024.104242
81. U.S. Food and Drug Administration. *Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: Draft guidance for industry and other interested parties.* 2025. FDA-2024-D-4689. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological> [Last accessed on 2025 Nov 06].
82. Kim J, Mikson C, Pollard V, *et al.* Key takeaways from FDA's draft guidance on use of AI in drug and biological life cycle. DLA Piper. 2025. Available from: <https://www.dlapiper.com/en-us/insights/publications/2025/01/fda-releases-draft-guidance-on-use-of-ai> [Last accessed on 2025 Jan 29].
83. Sakiyama Y. The use of machine learning and nonlinear statistical tools for ADME prediction. *Expert Opin Drug Metab Toxicol.* 2009;5(2):149–169.
doi: 10.1517/17425250902753261
84. Tiwari SK, Singh DK, Ladumor MK, *et al.* Study of degradation behaviour of montelukast sodium and its marketed formulation in oxidative and accelerated test conditions and prediction of physicochemical and ADMET properties of its degradation products using ADMET Predictor™. *J Pharm Biomed Anal.* 2018;158:106–118.
doi: 10.1016/j.jpba.2018.05.040
85. Siramshetty VB, Xu X, Shah P. Artificial intelligence in ADME property prediction. *Methods Mol Biol.* 2024;2714:307–327.
doi: 10.1007/978-1-0716-3441-7_17
86. Zhang J, McDonald MA, Koscher BA, *et al.* Application of machine learning and mechanistic modeling to predict intravenous pharmacokinetic profiles in humans. *J Med Chem.* 2025;68(11):7737–7750.
doi: 10.1021/acs.jmedchem.5c00340
87. U.S. Food and Drug Administration. *SafetAI Initiative.* 2025. Available from: <https://www.fda.gov/about-fda/nctr-research-focus-areas/safetai-initiative> [Last accessed on 2025 Nov 19].
88. U.S. Food and Drug Administration. *Roadmap to Reducing Animal Testing in Preclinical Safety Studies.* Available from: <https://www.fda.gov/media/186092/download> [Last accessed on 2025 Nov 19].
89. U.S. Food and Drug Administration. *FDA announces plan to phase out animal testing requirement for monoclonal antibodies and other drugs.* 2025. Available from: <https://www.fda.gov/news-events/press-announcements/fda-announces-plan-phase-out-animal-testing-requirement-monoclonal-antibodies-and-other-drugs> [Last accessed on 2025 Nov 19].
90. Li X, Sale M, Nieforth K, *et al.* pyDarwin: A machine learning-enhanced automated nonlinear mixed effect model selection toolbox. *Clin Pharmacol Ther.* 2024;115(4):758–773.
doi: 10.1002/cpt.3114
91. Certara. *Machine learning model selection with Darwin in Pirana.* 2023. Available from: <https://www.certara.com/on-demand-webinar/machine-learning-model-selection-with-darwin-in-pirana/> [Last accessed on 2025 Nov 5].
92. Rebai I, Duval V, Akil A, *et al.* *mlcov: New machine learning based R package for covariate selection.* In: Proceedings of the Population Approach Group Europe (PAGE) Meeting; June 2024; Verona, Italy (or virtual). Available from: <https://www.page-meeting.org/Abstracts/mlcov-new-machine-learning-based-r-package-for-covariate-selection/> [Last accessed on 2025 Nov 05].
93. Raheem E, Lu Z, Zheng F, *et al.* Machine learning in predicting longitudinal platelet counts: Applications in dose optimization. *Blood.* 2024;144:4985.
doi: 10.1182/blood-2024-199117
94. Kaddi C, Tao M, Bergeler S, *et al.* Quantitative systems pharmacology-based digital twins approach supplements clinical trial data for enzyme replacement therapies in Pompe disease. *Clin Pharmacol Ther.* 2025;117(2):579–588.
doi: 10.1002/cpt.3498
95. Joslyn LR, Huang W, Miles D, *et al.* Digital twins elucidate critical role of Tscm in clinical persistence of TCR-engineered cell therapy. *NPJ Syst Biol Appl.* 2024;10(1):11.
doi: 10.1038/s41540-024-00335-7
96. Susilo ME, Li CC, Gadkar K, *et al.* Systems-based digital

- p twins to help characterize clinical dose-response and propose predictive biomarkers in a Phase I study of bispecific antibody, mosunetuzumab, in NHL.
- Clin Transl Sci.*
- 2023;16(7):1134–1148.
-
- doi: 10.1111/cts.13501
97. Alum, EU. AI-driven biomarker discovery: Enhancing precision in cancer diagnosis and prognosis. *Discov Onc* 2025;16:313.
doi: 10.1007/s12672-025-02064-7
98. Lotter W; Hassett MJ; Schultz N; Kehl KL; Van Allen EM; Cerami E. Artificial Intelligence in Oncology: Current Landscape, Challenges, and Future Directions. *Cancer Discov.* 2024;14(5):711–726.
doi: 10.1158/2159-8290.CD-23-1199
99. Wallenta-Law J; Bapat B; Sweetnam C, *et al.* Real-World Impact of Comprehensive Genomic Profiling on Biomarker Detection, Receipt of Therapy, and Clinical Outcomes in Advanced Non–Small Cell Lung Cancer. *JCO Precis Oncol.* 2024;8:e2400075.
doi: 10.1200/PO.24.00075
100. Benary M, Wang XD, Schmidt M, *et al.* Leveraging Large Language Models for Decision Support in Personalized Oncology. *JAMA Netw Open.* 2023;6(11):e2343689.
doi: 10.1001/jamanetworkopen.2023.43689
101. Bidard F-C, Mayer EL, Park YH, *et al.* First-line camizestrant for emerging ESR1-mutated advanced breast cancer. *N Engl J Med.* 2025;393(6):569–580.
doi: 10.1056/NEJMoa2502929
102. Van Cutsem E, Köhne CH, Hitre E, *et al.* Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med.* 2009;360(14):1408–1417.
doi: 10.1056/NEJMoa0805019
103. Kopetz S, Yoshino T, Van Cutsem E, *et al.* Encorafenib, cetuximab and chemotherapy in BRAF-mutant colorectal cancer: A randomized phase 3 trial. *Nat Med.* 2025;31(3):901–908.
doi: 10.1038/s41591-024-03443-3
104. Fischer A; Pallavajjala A, Jiang L, *et al.* Artificial intelligence-assisted serial analysis of clinical cancer genomics data identifies changing treatment recommendations and therapeutic targets. *Clin Cancer Res.* 2022;28(11):2361–2372.
doi: 10.1158/1078-0432.CCR-21-4061
105. Jin Q, Wang Z, Floudas CS, *et al.* Matching patients to clinical trials with large language models. *Nat Commun.* 2024;15:9074.
doi: 10.1038/s41467-024-53081-z
106. Wornow M, Lozano A, Dash D, *et al.* Zero-shot clinical trial patient matching with LLMs. *NEJM AI.* 2025;2(1):Alcs2400360.
doi: 10.1056/Alcs2400360
107. Li I, Pan J, Goldwasser J, *et al.* Neural natural language processing for unstructured data in electronic health records: A review. *Comput Sci Rev.* 2022;46:100511.
doi: 10.1016/j.cosrev.2022.100511
108. Beattie J, Neufeld S, Yang D, *et al.* Utilizing large language models for enhanced clinical trial matching: A study on automation in patient screening. *Cureus.* 2024;16(5):e60044.
doi: 10.7759/cureus.60044
109. Kurnaz S, Loaiza-Bonilla A, Carvallo Castañeda D, *et al.* Effect of a novel artificial intelligence (AI)-enabled multi-trial matching system on patient matching using real-world data. *J Clin Oncol.* 2024;42(16 Suppl):e13501.
doi: 10.1200/JCO.2024.42.16_suppl.e13501
110. Anuyah S, Singh MK, Nyavor H. Advancing clinical trial outcomes using deep learning and predictive modelling: Bridging precision medicine and patient-centered care. *World J Adv Res Rev.* 2024;24(3):001–025.
doi: 10.30574/wjarr.2024.24.3.3671
111. Liu R, Rizzo S, Whipple S, *et al.* Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature.* 2021;592(7855):629–633.
doi: 10.1038/s41586-021-03430-5
112. Savitz ST, Savitz LA, Fleming NS, *et al.* How much can we trust electronic health record data? *Healthcare.* 2020;8(3):100444.
doi: 10.1016/j.hjdsi.2020.100444
113. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med.* 2018;178(11):1544–1547.
doi: 10.1001/jamainternmed.2018.3763
114. Borfitz D. AstraZeneca's pragmatic approach to clinical research. Clinical Research News. 2021. Available from: <https://www.clinicalresearchnewsonline.com/news/2021/03/22/astrazeneca-s-pragmatic-approach-to-clinical-research> [Last accessed on 2025 Nov 03].
115. Kallich JD. OSCER story, Act III. LinkedIn. 2015. Available from: <https://www.linkedin.com/pulse/oscer-story-act-iii-joel-kallich/> [Last accessed on 2025 Nov 15].
116. BEKhealth. Preventing patient dropouts with AI-based outreach: How AI enhances patient retention in clinical trials. 2025. Available from: <https://www.bekhealth.com/blog/preventing-patient-dropouts-with-ai-based-outreach-how-ai-enhances-patient-retention-in-clinical-trials/> [Last accessed on 2026 Jan 05].
117. Sambursky V. Clinical trials & patient retention: How

- projective AI goes beyond the boundaries of predictive AI*. Endominance. 2022. Available from: <https://www.endominance.com/blog/2022/07/20/discover-how-projective-ai-maximizes-patient-retention-in-clinical-trials/> [Last accessed on 2026 Jan 05].
118. Harrer S, Shah P, Antony B, *et al*. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci*. 2019;40(8):577–591.
doi: 10.1016/j.tips.2019.05.005
119. BioPharmaTrend. *The Pivotal Role of AI in Clinical Trials: From Digital Twins to Synthetic Control Arms*. Blog Article. 2025. Accessed on November 5th, 2025. Available from: <https://www.biopharmatrend.com/artificial-intelligence/the-pivotal-role-of-ai-in-clinical-trials-from-digital-twins-to-synthetic-control-arms-176/> [Last accessed on 2025 Nov 05].
120. Fisher CK, Smith AM, Walsh JR, *et al*. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Sci Rep*. 2019;9:13622.
doi: 10.1038/s41598-019-49656-2
121. Gatto NM, Campbell UB. Hope is not a strategy: Using robust real-world evidence to make better clinical development decisions. *Ther Innov Regul Sci*. 2025;59(6):1288–1293.
doi: 10.1007/s43441-025-00822-x
122. Klabunde T. *Digital “Twinning”: Clinical trials powered by AI*. *Sanofi Magazine*. 2024. Available from: <https://www.sanofi.com/en/magazine/our-science/digital-twinning-clinical-trials-ai> [Last accessed on 2025 Nov 05].
123. U.S. National Science Foundation. *Foundations for Digital Twins as Catalysts of Biomedical Technological Innovation (FDT-BioTech)*. Posted 2024. Available from: <https://www.nsf.gov/funding/opportunities/fdt-biotech-foundations-digital-twins-catalyzers-biomedical> [Last accessed on 2025 Nov 07].
124. U.S. National Science Foundation. *NSF, NIH and FDA support research in digital twin technology for biomedical applications*. Posted August 20, 2025. Available from: <https://www.nsf.gov/news/nsf-nih-fda-support-research-digital-twin-technology> [Last accessed on 2025 Nov 07].
125. European Medicines Agency, Heads of Medicines Agencies. *Multi-annual artificial intelligence workplan 2023–2028: HMA-EMA joint Big Data Steering Group*. 2023. Available from: https://www.ema.europa.eu/en/documents/work-programme/multi-annual-artificial-intelligence-workplan-2023-2028-hma-ema-joint-big-data-steering-group_en.pdf [Last accessed on 2025 Nov 15].
126. European Medicines Agency. *Real-world evidence framework to support EU regulatory decision-making: Report on experience gained with regulator-led studies September 2021–February 2023*. 2023. Available from: https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-report-experience-gained-regulator-led-studies-september-2021-february-2023_en.pdf [Last accessed on 2025 Nov 03].
127. Patel D, Grimson F, Mihaylova E, *et al*. Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health*. 2021;24(8):1118–1125.
doi: 10.1016/j.jval.2021.01.015
128. Servier. *Will synthetic control arms revolutionize clinical trials?* 2025. Available from: <https://servier.com/en/newsroom/synthetic-control-arms-revolutionize-clinical-trials/> [Last accessed on 2025 Nov 03].
129. Jahanshahi M, Gregg K, Davis G, *et al*. The use of external controls in FDA regulatory decision making. *Ther Innov Regul Sci*. 2021;55(5):1019–1035.
doi: 10.1007/s43441-021-00302-y
130. Wang X, Dormont F, Lorenzato C, *et al*. Current perspectives for external control arms in oncology clinical trials: Analysis of EMA approvals 2016–2021. *J Cancer Policy*. 2023;35:100403.
doi: 10.1016/j.jcpo.2023.100403
131. U.S. Food and Drug Administration. *Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products*. 2023. FDA-2022-D-2983. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products> [Last accessed on 2025 Nov 19].
132. Waters M. AI meets informed consent: A new era for clinical trial communication. *JNCI Cancer Spectr*. 2025;9(2):pkaf028.
doi: 10.1093/jncics/pkaf028
133. Clinion. *CSR Automation: Clinical Study Report Solution*. 2024. Available from: <https://www.clinion.com/csr-automation/> [Last accessed on 2025 Nov 03].
134. Atria. *30%–50% Faster Clinical Study Reports with Generative Artificial Intelligence-Powered Automation*. 2026. Available from: <https://www.atria.com/white-papers/transforming-clinical-study-reports-with-gen-ai-intelligence-powered-automation> [Last accessed on 2025 Nov 19].
135. Fast Data Science. *Transforming clinical trials with fast clinical AI*. Available from: <https://clinicaltrialrisk.org/clinical-trial-protocol-software/transforming-clinical-trials-with-fast-clinical-ai/> [Last accessed on 2025 Nov 03].
136. Fast Data Science. *AI-powered clinical trial analysis dashboard*. Available from: <https://clinical.fastdatascience.com/dashboard> [Last accessed on 2025 Nov 05].
137. CluePoints. *AI-Powered Clinical Trial Management & Documentation Solutions*. 2026. Available from: <https://cluepoints.com/what-we-do/risk-based-quality->

- management-rbqm/documentation/ [Last accessed on 2025 Nov 05].
138. Lampreia F, Madeira C, Dores H. Digital health technologies and artificial intelligence in cardiovascular clinical trials: A landscape of the European space. *Digit Health*. 2024;10. doi: 10.1177/20552076241277703
139. Desai MK. Artificial intelligence in pharmacovigilance—opportunities and challenges. *Perspect Clin Res*. 2024;15(3):116–121. doi: 10.4103/picr.picr_290_23
140. Painter JL, Kassekert R, Bate A. An industry perspective on the use of machine learning in drug and vaccine safety. *Front Drug Saf Regul*. 2023;3. doi: 10.3389/fdsfr.2023.1110498
141. Goldberg JM, Amin NP, Zachariah KA, *et al*. The introduction of AI into decentralized clinical trials. *JACC Adv*. 2024;3(8):101094. doi: 10.1016/j.jacadv.2024.101094
142. Yang Y, Krusche P, Pantoja K, *et al*. Using large language models to generate clinical trial tables and figures. *arXiv*. Preprint online 2024. doi: 10.48550/arXiv.2409.12046
143. Certara. *Pinnacle 21[®] Enterprise Software for Clinical Data Standardization*. Available from: <https://www.certara.com/pinnacle-21-enterprise-software/> [Last accessed on 2025 Nov 03].
144. Ross S, Carmeli I. Optimizing clinical research: Using AI for automated validation of output tables against ADaM. In: Proceedings of the PHUSE US Connect 2024; 5–8 May 2024; Bethesda, MD, USA. Available from: www.lexjansen.com/phuse-us/2024/et/PAP_ET06.pdf [Last accessed on 2025 Nov 19].
145. Shukla R, Bahl A. Transforming clinical trials with AI driven protocol optimization and next gen statistical programming. In: Proceedings of the PHUSE US Connect 2025; 16–19 March 2025. Orlando, FL, USA. Available from: https://www.lexjansen.com/phuse-us/2025/et/PAP_ET15.pdf [Last accessed on 2025 Nov 05].
146. Revvity Signals. *Clinical Data Like You've Never Seen It Before: Why Spotfire is the Leading Tool for Clinical Analytics*. 2023. Available from: https://revvitysignals.com/sites/default/files/2023-08/RS_White%20Paper%20_Clinical%20Data%20Like%20You%E2%80%99ve%20Never%20Seen%20It%20Before_v4_08102023_FINAL.pdf [Last accessed on 2025 Nov 15].
147. Spotfire Community. *Spotfire Copilot™*. 2025. Available from: <https://community.spotfire.com/articles/spotfire/spotfire-copilot/> [Last accessed on 2026 Jan 03].
148. Charman R. *Unveiling Spotfire Copilot™ 2.0: Discover the latest transformative features!* Spotfire Blog. 2025. Available from: <https://www.spotfire.com/blog/2025/03/12/unveiling-spotfire-copilot-2-0-discover-the-latest-transformative-features/> [Last accessed on 2026 Jan 03].
149. ISO/IEC 42001:2023. Information technology—Artificial intelligence—Management system. International Organization for Standardization. 2023. Available from: www.iso.org/standard/42001 [Last accessed on 2025 Nov 03].
150. European Parliamentary Research Service. *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. 2020. Available from: [www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf) [Last accessed on 2025 Nov 05].
151. Liu X, Rivera SC, Moher D, *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Lancet Digit Health*. 2020;2(10):e537–e548. doi: 10.1016/S2589-7500(20)30218-1
152. World Health Organization. *Ethics and governance of artificial intelligence for health: WHO guidance*. 2021. Available from: <https://www.who.int/publications/i/item/9789240029200> [Last accessed on 2025 Nov 05].
153. McKinsey & Company. *Unlocking peak operational performance in clinical development with artificial intelligence*. 2025. Available from: www.mckinsey.com/industries/life-sciences/our-insights/unlocking-peak-operational-performance-in-clinical-development-with-artificial-intelligence [Last accessed on 2025 Nov 05].
154. Williams, D. *The mother of invention: From steam engines to AI-designed drugs*. Drug Target Review. 2025. Available from: <https://www.drugtargetreview.com/article/190739/steam-to-ai-drug-innovation/> [Last accessed on 2026 Jan 03].
155. Wilczok D, Zhavoronkov A. Progress, Pitfalls, and Impact of AI-Driven Clinical Trials. *Clin Pharmacol Ther*. 2024;117(4):887–890. doi: 10.1002/cpt.3542
156. HL7 International, OHDSI. *FHIR to OMOP: HL7 Vulcan Accelerator project*. 2025. Available from: <https://hl7vulcan.org/projects/fhir-to-omop/> [Last accessed on 2026 Jan 03].
157. Jayathissa P, Subbiah S, Seneviratne M, *et al*. OMOP-on-FHIR: Integrating the Clinical Data Through FHIR Bundle to OMOP CDM. *Stud Health Technol Inform*. 2025;327:667–671. doi: 10.3233/shti250432
158. Apheris. *AI Structural Biology (AISB) Network*. Available from: <https://www.apheris.com/join-a-network/aisb> [Last accessed on 2025 Nov 03].
159. Heyndrickx W, Mervin L, Morawietz T, *et al*. MELLODDY: Cross-pharma Federated Learning at Unprecedented Scale Unlocks Benefits in QSAR without Compromising

- Proprietary Information. *J Chem Inf Model*. 2024;64(7):2331-2344.
doi: 10.1021/acs.jcim.3c00799
160. *FHIR to CDISC Joint Mapping Implementation Guide, Release 1.0.0—STU 1*. HL7 International. 2021. Available from: <https://hl7.org/fhir/uv/cdisc-mapping/STU1/> [Last accessed on 2026 Jan 03].
161. TransCelerate BioPharma Inc. *Digital Data Flow*. Available from: <https://www.transceleratebiopharmainc.com/initiatives/digital-data-flow/> [Last accessed on 2026 Jan 03].
162. RDMkit. *The Research Data Management toolkit for Life Sciences*. Available from: <https://rdmkit.elixir-europe.org/> [Last accessed on 2026 Jan 03].
163. Global Alliance for Genomics and Health. *Phenopackets*. Available from: <https://www.ga4gh.org/product/phenopackets/> [Last accessed on 2026 Jan 03].
164. Global Alliance for Genomics and Health. *Workflow Execution Service (WES)*. Available from: <https://www.ga4gh.org/product/workflow-execution-service-wes/> [Last accessed on 2026 Jan 03].
165. About Us. BioCompute Object Portal. Available from: <https://www.biocomputeobject.org/about/> [Last accessed on 2026 Jan 04].
166. Keeney J, King CH, Wang T, *et al*. *Updates to the BioCompute tools and guidelines*. Available from: <https://www.fda.gov/media/171316/download> [Last accessed on 2026 Jan 04].
167. Xu Z, Ren F, Wang P, *et al*. A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: A randomized phase 2a trial. *Nat Med*. 2025;31(8):2602-2610.
doi: 10.1038/s41591-025-03743-2
168. Austin D, Biswas K, Pollock K, *et al*. Interdisciplinary analysis of drugs: Structural features and clinical data. *J Clin Transl Sci*. 2022;6(1):e43.
doi: 10.1017/cts.2022.375