AI-Driven Drug Discovery: A Critical Evaluation of Model Reliability, Data Bias, and Clinical Translation

Anjali Yadav, Siddhesh Gavate, Naresh Chaudari, Ankit Boraste

School of Pharmaceutical Sciences, Sandip University Nashik

Corresponding author Email: anjaliysops@gmail.com
Doi: https://doi.org/10.5281/zenodo.20665482
Received: 17 May 2026
Accepted: 24 May 2026

Abstract

Artificial Intelligence (AI) is rapidly transforming the landscape of drug discovery by enabling faster and more efficient identification of potential therapeutic candidates. Advanced computational techniques such as machine learning (ML), deep learning (DL), and natural language processing (NLP) have demonstrated significant potential in analyzing complex biological datasets, predicting drug-target interactions, and optimizing lead compounds.
Despite these advancements, several critical challenges hinder the widespread adoption of AI in pharmaceutical research. Model reliability remains a major concern due to issues such as overfitting, limited generalizability, and lack of interpretability. Additionally, biases present in training datasets can result in skewed predictions, thereby affecting the safety and efficacy of AI-generated drug candidates. Furthermore, the translation of AI-based discoveries from computational models to clinical applications continues to face substantial barriers, including regulatory constraints and validation complexities. This review critically examines these challenges, emphasizing the need for robust validation strategies, high-quality datasets, and interdisciplinary collaboration. Addressing these limitations is essential for realizing the full potential of AI in developing safe, effective, and accessible therapeutics.

Keywords

Artificial Intelligence, Drug Discovery, Machine Learning, Deep Learning, Model Reliability, Data Bias, Clinical Translation, Computational Biology, Pharmacoinformatics

Introduction

The conventional drug discovery process is a lengthy, costly, and high-risk endeavor, often requiring over a decade of research and substantial financial investment to bring a single drug to market [1]. High attrition rates during various stages of development further complicate this process, emphasizing the need for innovative approaches to improve efficiency and success rates. In recent years, Artificial Intelligence (AI) has emerged as a transformative tool in pharmaceutical research, offering the ability to analyze vast and complex datasets with unprecedented speed and accuracy [2].

AI-driven methodologies, particularly machine learning (ML) and deep learning (DL), have enabled significant advancements in key stages of drug discovery, including target identification, virtual screening, lead optimization, and toxicity prediction [3]. These techniques utilize large-scale biological and chemical data to identify patterns and relationships that may not be apparent through traditional experimental methods.

Additionally, natural language processing (NLP) facilitates the extraction of valuable insights from scientific literature, clinical reports, and biomedical databases, further accelerating research progress [4].

One of the most promising applications of AI is the use of generative models to design novel molecular entities

with desired pharmacological properties. These models can predict drug-target interactions and optimize molecular structures, thereby reducing the time required for experimental validation [5]. However, despite these advantages, the reliability of AI models remains a critical issue. Many models are trained on limited datasets that may not adequately represent biological diversity, leading to overfitting and reduced predictive performance in real-world scenarios [6].

Another significant challenge is data bias, which can arise from imbalanced datasets, incomplete representation of populations, or selective reporting of experimental results. Such biases can compromise the accuracy and fairness of AI predictions, potentially leading to ineffective or unsafe drug candidates [7]. Moreover, the "black-box" nature of many deep learning models limits interpretability, making it difficult for researchers and regulatory authorities to understand and trust the decision-making process [8].

Translating AI-generated discoveries into clinical applications presents additional hurdles. Although AI models may show promising results in silico, many drug candidates fail during clinical trials due to issues related to safety, efficacy, or regulatory compliance [9]. This gap between computational predictions and clinical outcomes underscores the need for rigorous validation frameworks and closer collaboration between computational scientists, pharmacologists, and clinicians.

This review aims to critically evaluate the current challenges in AI-driven drug discovery, focusing on model reliability, data bias, and clinical translation. By addressing these limitations, the pharmaceutical industry can better leverage AI technologies to enhance drug development processes and improve patient outcomes.

3.1 Evolution of AI in Pharmaceutical R&D

The integration of Artificial Intelligence (AI) into pharmaceutical research has progressed over several decades, driven by improvements in computational infrastructure, algorithm development, and data availability. Early AI applications in the 1970s and 1980s relied on rule-based expert systems designed to assist in molecular structure elucidation and clinical reasoning. While these systems demonstrated the conceptual feasibility of AI, their utility was limited by insufficient datasets and computational power [12].

During the 1990s, the field shifted toward quantitative modeling approaches, particularly quantitative structure–activity relationship (QSAR) models. These models applied statistical and early machine learning techniques to predict biological activity based on molecular descriptors, marking a transition toward data-driven drug discovery [13]. With the expansion of cheminformatics databases and high-throughput screening technologies in the early 2000s, machine learning algorithms such as support vector machines and random forests gained prominence [14].

The last decade has witnessed a major transformation with the advent of deep learning and big data analytics. Advanced neural network architectures, including graph neural networks and convolutional models, have enabled the analysis of complex biological and chemical systems at scale [15]. Notably, breakthroughs in protein structure prediction using deep learning have demonstrated the profound impact of AI in understanding molecular biology [16]. Additionally, generative AI models have enabled the design of novel chemical entities, shifting the paradigm from predictive modeling to creative molecular design [17].

Currently, AI technologies are integrated across multiple stages of drug discovery, including target identification, lead optimization, and clinical trial design. Despite rapid progress, ongoing challenges related to interpretability, reliability, and translational success continue to shape the evolution of this field [18].

3.2 Hype vs Actual Clinical Success Rates

The rapid advancement of AI in drug discovery has generated considerable enthusiasm, often accompanied by exaggerated expectations regarding its ability to revolutionize pharmaceutical development. Industry reports and media narratives frequently highlight AI as a solution capable of significantly reducing development timelines and costs [19].

However, real-world clinical success rates of AI-driven drug discovery remain relatively modest. Despite substantial investments, only a limited number of AI-designed drug candidates have progressed to advanced clinical trial stages, and very few have achieved regulatory approval to date [20].

This disparity between expectations and outcomes can be attributed to several factors. Drug discovery involves highly complex biological systems that are difficult to model accurately using current AI techniques. Additionally, many AI models are trained on limited or biased datasets, which can compromise their predictive accuracy in real-world scenarios [21].
Furthermore, the transition from computational predictions to clinical validation introduces uncertainties related to pharmacokinetics, toxicity, and inter-patient variability [22].

Nonetheless, AI has contributed to incremental improvements, particularly in early-stage drug discovery processes such as virtual screening and target identification. Rather than replacing traditional approaches, AI currently functions as a complementary tool that enhances efficiency and decision-making [23].

3.3 Scope: Reliability, Bias, and Translational Bottlenecks

The application of AI in drug discovery must be critically assessed in the context of three major challenges: model reliability, data bias, and translational bottlenecks.

Model reliability refers to the ability of AI systems to produce accurate and consistent predictions across diverse datasets and experimental conditions. In drug discovery, unreliable predictions can lead to costly downstream failures, emphasizing the importance of robustmodel validation [24].
Data bias is another critical limitation. Biomedical datasets often exhibit imbalances, incomplete representation, and experimental inconsistencies. These biases can influence model training and result in skewed or non-generalizable predictions [25].

Translational bottlenecks arise when AI-generated findings fail to translate into clinically viable outcomes. While AI models may demonstrate strong performance in computational settings, they often fail to capture the full complexity of biological systems, leading to discrepancies during experimental or clinical validation [26].

Collectively, these challenges highlight the need for a balanced and critical approach to AI adoption in drug discovery.

4. AI in Drug Discovery: Current Capabilities

4.1 Target Identification and Validation

AI has significantly enhanced target identification by enabling the integration and analysis of large-scale omics datasets, including genomics, proteomics, and transcriptomics. Machine learning models can identify disease-associated genes and predict target druggability based on biological networks and interaction patterns [27].

Additionally, AI-driven approaches facilitate hypothesis generation by uncovering hidden relationships within complex biological systems. Network-based models can reveal novel disease pathways, expanding therapeutic possibilities [28]. However, experimental validation remains essential to confirm the biological relevance of AI-predicted targets.

4.2 Virtual Screening and De Novo Design

Virtual screening is a well-established application of AI, enabling the rapid evaluation of large chemical libraries to identify potential drug candidates. Machine learning models can predict binding affinities and prioritize compounds more efficiently than traditional docking approaches [29].

AI also enables de novo drug design through generative models such as variational autoencoders and generative adversarial networks. These models can generate novel molecular structures with desired pharmacological properties, expanding the chemical search space [30]. However, challenges related to synthetic feasibility and biological validation persist.

4.3 Lead Optimization and ADMET Prediction

AI plays a critical role in lead optimization by predicting pharmacokinetic and toxicological properties (ADMET).

Machine learning models can evaluate how structural modifications influence drug behavior, enabling the selection of compounds with optimal efficacy and safety profiles [31].

AI-based toxicity prediction is particularly valuable for reducing late-stage failures.
Additionally, multi-objective optimization techniques allow simultaneous improvement of multiple drug properties. Despite these advances, generalization limitations remain a challenge due to complex biological interactions [32].

4.4 AI in Clinical Trial Design

AI is increasingly applied in clinical trial design to improve efficiency and success rates.
Machine learning models enable patient stratification, identifying individuals most likely to respond to a treatment [33]. AI also supports patient recruitment through analysis of electronic health records and optimization of dosing strategies.

Adaptive clinical trial designs, powered by AI, allow real-time modification of trial parameters based on incoming data. However, regulatory, ethical, and data privacy concerns continue to limit widespread adoption [34].

5. Model Reliability: The Core

5.1 What Defines Reliability in AI-Driven Drug Discovery

Reliability in AI-driven drug discovery refers to the consistency, robustness, and generalizability of model predictions across different datasets and conditions. Reliable models demonstrate stable performance and minimal sensitivity to data variability [35].

5.2 Overfitting, Data Leakage, and Poor Generalization

Overfitting occurs when models learn dataset-specific patterns rather than generalizable features, leading to poor real-world performance. Data leakage further compromises model evaluation by introducing information from test datasets into training processes [36].

5.3 Reproducibility Crisis in AI Models

Reproducibility remains a major concern in AI research. Variability in data preprocessing, model parameters, and computational environments can lead to inconsistent results. Greater transparency and standardized reporting are required to address this issue [37].

5.4 External Validation vs Benchmark Datasets

Benchmark datasets are commonly used for model evaluation but may not reflect real-world conditions. External validation using independent datasets provides a more realistic assessment of model performance and generalizability [38].

5.5 Lack of Standardization Across Studies

The absence of standardized methodologies and evaluation metrics makes it difficult to compare results across studies. Establishing unified frameworks is essential for improving reproducibility and facilitating AI adoption in drug discovery [39].

6. Data Bias: The Hidden Risk

Artificial intelligence has significantly accelerated drug discovery; however, its effectiveness depends heavily on the quality and representativeness of underlying datasets. One of the most critical yet often underestimated challenges in AI-driven drug discovery is data bias. Bias does not simply reduce model accuracy; it can systematically distort predictions, reinforce existing scientific gaps, and ultimately hinder clinical translation [40].

Bias in AI systems can be broadly categorized into data-driven, algorithmic, and human-induced biases, all

of which interact throughout the drug discovery pipeline [41]. In pharmaceutical research, where outcomes directly impact patient health, such biases may lead to inequitable therapeutic solutions, failed clinical trials, and overlooked drug candidates.

6.1 Sources of Bias in Biomedical Datasets

Biomedical datasets are inherently complex and often reflect historical, technical, and societal limitations. Bias can emerge during data collection, curation, and annotation processes.

Sampling bias is a major concern, where datasets disproportionately represent certain populations. For example, many genomic datasets are dominated by individuals of European ancestry, limiting model generalizability across diverse populations and increasing the risk of inaccurate predictions in global healthcare applications [42].

Measurement bias arises from systematic variations in experimental techniques, laboratory conditions, and instrumentation, which can introduce inconsistencies that models may incorrectly interpret as meaningful biological signals [43].

Publication bias further exacerbates the issue, as positive findings are more likely to be reported than negative or inconclusive results. Consequently, AI models trained on literature-based datasets may overestimate drug efficacy and fail to account for unsuccessful outcomes [44].

Labeling bias occurs when human annotations are influenced by subjective judgment or incomplete knowledge. In drug discovery, classification of compounds as “active” or “inactive” often varies depending on experimental thresholds, leading to inconsistencies across datasets [45].

These issues are compounded by the retrospective and fragmented nature of many biomedical datasets, which are often not designed for machine learning applications and may fail to capture biological complexity [46].

6.2 Chemical Space Bias and Dataset Imbalance

Chemical space encompasses an enormous number of possible molecular structures; however, available datasets cover only a limited and highly selective subset.

Most datasets are biased toward previously synthesized compounds, molecules with known biological activity, and chemically “drug-like” structures. This results in chemical space bias, where AI models struggle to generalize beyond familiar molecular scaffolds and fail to identify truly novel compounds [47].

Dataset imbalance is another significant issue. In many screening datasets, inactive compounds significantly outnumber active ones (or vice versa), leading to skewed model predictions. Models trained on imbalanced datasets may achieve high statistical accuracy while lacking practical utility in identifying new drug candidates [48].

Experimental bias, such as preferential testing of specific compound classes, further distorts the dataset. Additionally, improper dataset splitting strategies can introduce hidden biases, artificially inflating model performance during validation [49].

Together, these factors limit innovation and create a misleading perception of model effectiveness.

6.3 Algorithmic Bias and Skewed Predictions

Algorithmic bias arises from the design and optimization of machine learning models. Most models are optimized for overall accuracy rather than fairness or representativeness, resulting in better performance on dominant classes and poor performance on rare or minority cases [50].

Feature representation also plays a critical role. The choice of molecular descriptors or embeddings can influence how models interpret chemical structures. Inadequate representation of biochemical properties can lead to systematically biased predictions [51].

Black-box models, such as deep neural networks, often lack interpretability, making it difficult to detect

hidden biases. This lack of transparency reduces trust in AI-generated predictions and complicates regulatory acceptance [52].

Moreover, models trained on retrospective benchmarks often fail to perform effectively in prospective settings, highlighting a gap between evaluation metrics and real-world applicability [53].

Algorithmic bias often reinforces underlying data bias, creating a feedback loop that amplifies prediction errors across the drug discovery pipeline.

6.4 Case Examples of Biased Outcomes

Real-world examples demonstrate the significant impact of bias in AI systems. In healthcare,
AI models trained on non-diverse datasets have shown reduced diagnostic accuracy for underrepresented populations, contributing to healthcare disparities [54].

In chemical research, biased datasets have led to overfitting to known reaction pathways, limiting the ability of models to predict novel chemical transformations and restricting innovation [55].

Toxicity prediction models have also been affected by biased training data, resulting in incorrect classification of compounds as safe or harmful. Such errors can lead to failures in later-stage clinical trials [56].

Beyond drug discovery, algorithmic bias has been widely documented across various domains, where systems inadvertently favor certain groups due to biased data, reinforcing existing inequalities [57].

These examples highlight that bias is not merely theoretical but has tangible scientific and ethical consequences.

6.5 Strategies for Bias Mitigation (Curation, Augmentation, Federated Learning)

Addressing bias in AI-driven drug discovery requires a comprehensive approach combining data-centric and algorithmic strategies.

7.1. Data Curation

Rigorous data curation involves removing errors, standardizing experimental conditions, and ensuring balanced representation across populations and chemical classes. High-quality
datasets are essential for improving model reliability [58].

2. Data Augmentation

Data augmentation techniques, including generative modeling, can expand dataset diversity by creating synthetic molecular structures and simulating biological interactions. This
improves generalization and reduces dataset imbalance [59].

3. Federated Learning

Federated learning enables collaborative model training across institutions without sharing sensitive data. This approach enhances dataset diversity, preserves privacy, and reduces
institutional bias [60].

4. Bias Detection and Auditing

Emerging auditing frameworks allow systematic identification and quantification of bias in datasets and models. These tools support continuous monitoring and targeted mitigation
strategies [61].

5. Explainable AI (XAI)

Improving model interpretability helps uncover hidden biases and ensures that predictions are biologically

meaningful. Explainable models enhance trust and facilitate regulatory acceptance [62].

Conclusion of Section

Data bias represents a critical and often hidden risk in AI-driven drug discovery. It arises from complex interactions between datasets, algorithms, and human decisions. If not addressed, bias can compromise model reliability, hinder innovation, and exacerbate healthcare disparities.

However, through rigorous data curation, advanced modeling approaches, and collaborative frameworks such as federated learning, these challenges can be mitigated. Ensuring fairness and representativeness in AI systems is both a scientific necessity and an ethical imperative for the advancement of precision medicine.

8. Bridging the Gap: Solutions and Emerging Approaches

8.1 Explainable AI (XAI) for Drug Discovery

The increasing reliance on deep learning in drug discovery has exposed a fundamental limitation: lack of interpretability. While these models achieve high predictive performance in tasks such as molecular property prediction and target identification, their "black-box" nature raises concerns regarding reproducibility, scientific validity, and regulatory acceptance.

Explainable AI (XAI) has therefore emerged as a critical enabler for trustworthy AI-driven drug discovery. XAI techniques such as SHAP and LIME allow the decomposition of model predictions into feature-level contributions, enabling researchers to identify which molecular substructures or physicochemical properties influence activity predictions [64,65]. This is particularly relevant in understanding structure–activity relationships (SAR), where mechanistic interpretability is essential. By aligning computational predictions with established chemical knowledge, XAI enhances both credibility and usability.

However, current approaches are largely post hoc and may not accurately reflect the true internal reasoning of models. This creates the risk of "illusory explanations," where interpretations appear plausible but are not causally grounded [66]. Consequently, there is a shift toward inherently interpretable architectures, including graph-based models with attention mechanisms that provide more transparent reasoning pathways. From a translational perspective, explainability is increasingly becoming a regulatory requirement. Without clear justification of predictions, AI-generated drug candidates face barriers in preclinical validation and approval processes. Thus, XAI should be viewed not as an optional feature but as a foundational component of robust AI systems in drug discovery.

8.2 Integration with Multi-Omics and Real-World Data

Drug discovery is inherently complex, involving interactions across multiple biological layers. Traditional AI models, often trained on isolated datasets, fail to capture this complexity. The integration of multi-omics data—genomics, transcriptomics, proteomics, and metabolomics—represents a significant advancement toward systems-level understanding of disease mechanisms [67].

By combining multi-omics datasets, AI models can uncover intricate biological networks and identify context-specific drug targets. For instance, integrating genomic mutations with transcriptomic profiles enables more accurate prediction of disease pathways and therapeutic responses, particularly in oncology where heterogeneity is a major challenge [68].

In parallel, real-world data (RWD) such as electronic health records and patient registries provide insights into actual clinical outcomes beyond controlled trial environments. AI models trained on RWD can better predict drug effectiveness, adverse events, and patient-specific responses [69]. This enhances clinical relevance and bridges the gap between experimental predictions and real-world performance.

Despite these advantages, significant challenges persist. Data heterogeneity, missing values, and batch effects complicate integration efforts. Moreover, high dimensionality combined with limited sample sizes increases the risk of overfitting. Advanced approaches such as graph neural networks and multi-modal deep learning are being developed to address these issues.

Overall, the integration of multi-omics and real-world data marks a transition from reductionist to holistic drug discovery, improving both predictive accuracy and translational potential.

8.3 Federated and Privacy-Preserving Learning

Data privacy concerns remain a major obstacle in leveraging large-scale healthcare datasets. Institutions are often unable to share sensitive data due to legal, ethical, and competitive constraints. Federated learning (FL) addresses this challenge by enabling decentralized model training without direct data sharing [70].

In federated learning, models are trained locally on institutional datasets, and only model parameters are shared and aggregated. This allows multiple stakeholders to collaboratively build robust models while preserving data privacy. Additional techniques such as differential privacy and secure multi-party computation further strengthen data protection [71].

However, federated learning introduces new complexities. Data across institutions is often non-identically distributed, which can lead to biased or unstable models. Communication overhead and synchronization challenges also affect scalability. Furthermore, privacy-preserving techniques may reduce model accuracy, highlighting a trade-off between performance and security [72].

Despite these limitations, privacy-preserving AI is essential for the future of drug discovery, particularly in the context of increasingly stringent data protection regulations. Federated learning represents a crucial step toward collaborative and ethical AI development.

8.4 Human–AI Collaboration Models

The narrative that AI will replace human expertise in drug discovery is increasingly being replaced by a more realistic perspective: collaboration. Human–AI collaboration models leverage the complementary strengths of computational systems and domain experts. AI excels at analyzing large datasets, identifying patterns, and generating hypotheses.

However, it lacks contextual understanding, ethical reasoning, and the ability to critically evaluate its own outputs. Human experts, particularly medicinal chemists and clinicians, play a crucial role in interpreting results, validating predictions, and designing experiments [73].

Collaborative workflows involve iterative interactions where AI-generated insights are refined through human expertise. Interactive decision-support systems allow researchers to explore model outputs, adjust parameters, and test alternative hypotheses, enhancing both transparency and control.

This paradigm of augmented intelligence not only improves efficiency but also fosters trust in AI systems. By integrating human judgment with machine intelligence, drug discovery becomes more robust, reliable, and clinically relevant.

8.5 Standardization and Benchmarking Frameworks

A major barrier to the adoption of AI in drug discovery is the lack of standardized evaluation frameworks. Variability in datasets, performance metrics, and validation strategies makes it difficult to compare models and assess their real-world applicability.

Benchmark datasets such as MoleculeNet provide a starting point for evaluation, but they often fail to capture the complexity and variability of real-world data [74]. Additionally, issues such as data leakage and overfitting can lead to inflated performance metrics, undermining model reliability [75].

To address these challenges, robust validation strategies, including time-split validation and external testing on independent datasets, are essential. Standardized reporting guidelines are also needed to ensure transparency and reproducibility.

Ultimately, the development of comprehensive benchmarking frameworks is critical for translating AI models from academic research to industrial and clinical applications.

9. Industry Landscape and Trends

9.1 Role of Pharma–AI Collaborations

Collaborations between pharmaceutical companies and AI firms have become a defining feature of the modern drug discovery ecosystem. These partnerships combine domain expertise with advanced computational capabilities, accelerating various stages of the drug development pipeline [76].

However, the success of such collaborations depends on effective integration of workflows, data quality, and alignment of strategic objectives. Without these factors, AI remains an isolated tool rather than a transformative force.

9.2 Startups and Innovation Ecosystem

AI-driven startups are playing a pivotal role in driving innovation within drug discovery. These companies focus on specialized domains such as protein structure prediction, generative chemistry, and predictive toxicology.

While startups offer agility and technological innovation, they face challenges related to scalability, data access, and regulatory validation. Many rely on partnerships with larger pharmaceutical companies to translate their innovations into clinical applications [77].

9.3 Investment Trends and Pipeline Analysis

Investment in AI-based drug discovery has grown significantly, reflecting strong industry confidence. However, there is a shift from hype-driven funding to outcome-based investment, where clinical validation and measurable impact are prioritized [78].

Despite a growing number of AI-generated candidates entering development pipelines, relatively few have reached late-stage clinical trials. This highlights the persistent gap between computational predictions and clinical success.

10. Future Outlook: Toward Trustworthy AI

10.1 Regulatory Evolution

Regulatory frameworks are gradually evolving to accommodate AI-driven approaches. There is increasing emphasis on transparency, validation, and risk assessment. Future regulations are expected to mandate explainability and continuous monitoring of AI systems [79]

10.2 Personalized and Precision Medicine

AI enables the integration of genetic, environmental, and lifestyle data, facilitating personalized treatment strategies. This represents a major shift toward precision medicine, improving therapeutic efficacy and patient outcomes [80].

10.3 From Hype to Clinical Reality

While AI has demonstrated significant potential, its long-term success depends on successful clinical translation. The focus is shifting from theoretical models to practical applications that deliver tangible benefits in real-world settings [81].

11. Conclusions (with citation numbering after 81)

The integration of artificial intelligence (AI) into drug discovery is often portrayed as a revolutionary advancement capable of overcoming the long-standing inefficiencies of pharmaceutical research. However, a critical evaluation suggests that, despite its impressive computational capabilities, AI is not yet clinically reliable. The gap between technological promise and clinical applicability remains substantial, and it is essential to address this gap with a realistic and evidence-based perspective.

One of the most fundamental challenges lies in model reliability. AI models, particularly those based on machine learning and deep learning, depend heavily on the quality and diversity of training datasets. While these models are highly efficient at identifying patterns within structured data, they often struggle to generalize when applied to real-world clinical scenarios. Biological systems are inherently complex and variable, and AI models trained on limited datasets may fail when exposed to new patient populations or disease conditions.
Studies have demonstrated that AI-driven predictions, although accurate in computational environments, frequently do not translate effectively into clinical success due to this variability [85]. This limitation highlights that AI systems are still far from achieving true biological understanding.

Closely related to this issue is data bias, which remains one of the most significant barriers in AI-driven drug discovery. Most AI models rely on historical datasets that may be incomplete, unbalanced, or biased toward specific populations. Such biases can lead to inaccurate predictions and may compromise both drug efficacy and patient safety. Additionally, inconsistencies in data collection, annotation, and reporting practices further reduce the reproducibility of AI models. The absence of standardized datasets and evaluation frameworks makes it difficult to validate results across different studies, thereby limiting trust in AI-based systems [83]. Without addressing these biases, AI risks reinforcing existing limitations rather than resolving them.

Another critical concern is the lack of robust validation frameworks. Although AI can rapidly generate hypotheses, predict drug-target interactions, and optimize lead compounds, these outputs are often validated only through computational metrics rather than experimental or clinical evidence. This creates a significant gap between theoretical predictions and practical applications. The integration of AI with model-informed drug development has emphasized the necessity of standardized validation protocols, including experimental verification and regulatory alignment, to ensure the reliability and reproducibility of AI-generated outcomes [85]. Until such frameworks are established, AI cannot be considered a fully dependable tool in pharmaceutical development.

The challenge of clinical translation further illustrates the limitations of AI. Despite notable progress in early-stage research, only a limited number of AI-generated drug candidates successfully progress to clinical approval. Drug development involves multiple complex stages, including pharmacokinetics, toxicity assessment, and large-scale clinical trials, which AI models cannot fully replicate. High attrition rates in clinical trials continue to demonstrate that computational predictions alone are insufficient for successful drug development [82].
Moreover, regulatory frameworks are still evolving to accommodate AI-based methodologies, creating additional barriers to clinical implementation.

The issue of interpretability, often referred to as the “black-box problem,” also restricts the widespread adoption of AI in healthcare. Many advanced AI models produce predictions without clear explanations, making it difficult for clinicians and regulatory authorities to trust and validate their outputs. In a field where transparency and accountability are critical, this lack of interpretability poses a significant challenge.

Despite these limitations, AI has undeniably enhanced efficiency in drug discovery by accelerating target identification, improving lead optimization, and enabling large-scale data analysis. However, these advantages should not be mistaken for complete reliability. AI should be regarded as a supportive and augmentative tool rather than a replacement for experimental and clinical expertise.

In conclusion, AI-driven drug discovery can be described as powerful but not yet clinically dependable. The most significant barriers—data bias and insufficient validation—continue to limit its real-world impact. Future progress will depend on improving data quality, developing transparent and interpretable models, and establishing standardized validation frameworks. Until these challenges are addressed, AI will remain an important yet imperfect tool, capable of accelerating research but not fully replacing traditional drug discovery methods.

12. **Declarations**

Funding

The authors declare that no specific funding was received for this study. Conflicts of Interest

The authors declare no conflicts of interest. Ethical Approval

This study is based on previously published literature and does not involve human or animal subjects.

Data Availability

All data used in this study are derived from publicly available sources [82–86]. Author

Contribution

All authors contributed equally to the conceptualizationn, analysis, and writing of this manuscript.

Acknowledgments

The authors acknowledge the contributions of previously published research that supported this study.

Refernces

1. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9(3):203–214.
2. Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Drug Discov Today. 2019;24(3):773–780.
3. 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.
4. Wang Y, Wang J, Cao Z, et al. Systematic review of natural language processing in biomedical research. Brief Bioinform. 2020;21(4):1231–1245.
5. Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37(9):1038–1040.
6. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15(141):20170387.
7. Mehrabi N, Morstatter F, Saxena N, et al. A survey on bias and fairness in machine learning. ACM Comput Surv. 2021;54(6):1–35.
8. Rudin C. Stop explaining black box machine learning models for high-stakes decisions. Nat Mach Intell. 2019;1(5):206–215.
9. Ekins S, Puhl AC, Zorn KM, et al. Exploiting machine learning for end-to-end drug discovery and development. Nat Mater. 2019;18(5):435–441.
10. Schneider P, Walters WP, Plowright AT, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353–364.
11. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
12. Luger GF. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 6th ed. Pearson; 2008.
13. Hansch C, Fujita T. p-σ-π analysis: a method for correlating biological activity and chemical structure. J Am Chem Soc. 1964;86:1616–1626.
14. Cherkasov A, et al. QSAR modeling: where have you been? Where are you going to? J Med Chem. 2014;57(12):4977–5010.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–444.
16. Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–589.
17. Elton DC, Boukouvalas Z, Fuge MD, Chung PW. Deep learning for molecular design. Mol Syst Des Eng. 2019;4:828–849.
18. Ekins S. The next generation of computational drug discovery. Pharm Res. 2016;33:2594–2603.

19. Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018;17:97–113.
20. Paul D, et al. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021;26:80–93.
21. Beam AL, Kohane IS. Big data and machine learning in healthcare. JAMA. 2018;319(13):1317–1318.
22. Scannell JW, et al. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012;11:191–200.
23. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects. Bioinformatics. 2018;34:i457–i466.
24. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12:56–68.
25. Stokes JM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180:688–702.
26. Brown N, et al. GuacaMol: benchmarking models for de novo molecular design. J Chem Inf Model. 2019;59:1096–1108.
27. Yang K, et al. ADMET prediction using machine learning. J Chem Inf Model. 2019;59:3370–3388.
28. Wu Z, et al. MoleculeNet: a benchmark for molecular machine learning. Chem Sci. 2018;9:513–530.
29. Miotto R, et al. Deep learning for healthcare: review. Brief Bioinform. 2018;19:1236–1246.
30. Harrer S, et al. Artificial intelligence for clinical trial design. Drug Discov Today. 2019;24:205–214.
31. Kaufman S, et al. Leakage in data mining: formulation, detection, and avoidance. ACM Trans Knowl Discov Data. 2012;6:1–21.
32. Hutson M. Artificial intelligence faces reproducibility crisis. Science. 2018;359:725–726.
33. Chicco D. Ten quick tips for machine learning in computational biology. BioData Min. 2017;10:7.
34. Walsh I, et al. DOME: recommendations for supervised machine learning validation. Nat Methods. 2021;18:1121–1127.
35. Norori N, et al. Addressing bias in big data and AI for healthcare. Patterns. 2021;2(4):100347.
36. Suresh H, Guttag JV. A framework for understanding sources of harm in machine learning. arXiv. 2021.
37. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538:161–164.
38. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2:e124.
39. Song F, et al. Publication bias in clinical research. Health Technol Assess. 2010;14:1–193
40. Northcutt CG, et al. Confident learning: estimating errors in datasets. J Artif Intell Res. 2021;70:1373–1411.
41. Reymond JL. The chemical space project. Acc Chem Res. 2015;48:722–730.
42. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21:1263–1284.
43. Wallach I, Heifets A. Most machine learning models in drug discovery fail due to bias. arXiv. 2018
44. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50:742–754.
45. Kapoor S, Narayanan A. Leakage and reproducibility in machine learning. arXiv. 2022.
46. Obermeyer Z, et al. Dissecting racial bias in healthcare algorithms. Science. 2019;366:447–453.
47. Baeza-Yates R. Bias on the web. Commun ACM. 2018;61:54–61
48. Wilkinson MD, et al. The FAIR guiding principles for scientific data management. Sci Data. 2016;3:160018.
49. Rieke N, et al. Federated learning for healthcare. Nat Mach Intell. 2020;2:705–712.
50. Samek W, et al. Explainable artificial intelligence: understanding machine learning decisions. Proc

IEEE. 2021;109:247–278
51. Cross H, et al. Bias in medical artificial intelligence systems. NPJ Digit Med. 2024;7:12.
52. Griffiths RR, et al. Dataset bias in chemical reaction prediction. arXiv. 2021
53. Ghassemi M, et al. A review of challenges and opportunities in machine learning for healthcare. JAMA. 2021;325:825–826
54. Elton DC, et al. Deep learning for molecular generation. Mol Syst Des Eng. 2019;4:828–849.
55. Norori N, et al. Bias in healthcare AI systems. Patterns. 2021;2:100347.
56. Suresh H, Guttag JV. Sources of bias in machine learning pipelines. arXiv. 2021.
57. Popejoy AB, Fullerton SM. Lack of diversity in genomics research. Nature. 2016;538:161–164.
58. Ioannidis JPA. Reproducibility crisis in research findings. PLoS Med. 2005;2:e124
59. Wilkinson MD, et al. FAIR data principles for scientific data. Sci Data. 2016;3:160018.
60. Rieke N, et al. Federated learning applications in healthcare. Nat Mach Intell. 2020;2:705–712
61. Ghislat G, et al. Data-centric AI challenges in drug discovery. arXiv. 2024.
62. Samek W, et al. Explainable AI: interpreting deep learning models. Proc IEEE. 2021;109:247–278.
64. Lundberg Sivı, Lee 1. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017.
65. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Drug Discov Today. 2020;25(7):1124-32.
66. Samek W, Wiegand T, Müller KR. Explainable artificial intelligence. ITU J. 2019
67. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;18:83.
68. Zitnik M, et al. Machine learning for integrating data in biology and medicine. Nat Rev Drug Discov. 2019;18:733-52.
69. Topol EJ. High-performance medicine. Nat Med. 2019;25:44-56.
70. McMahan B, et al. Communication-efficient learning of deep networks. AISTATS. 2017.
71. Rieke N, et al. The future of digital health with federated learning. Nat Mach Intell. 2020;2:305-11.
72. Li T, et al. Federated learning challenges and methods. IEEE Signal Process Mag. 2020;37:50-60
73. Holzinger A. Interactive machine learning for health informatics. Artif Intell Med. 2016;66:1-9.
74. Wu Z, et al. MoleculeNet benchmark. Chem Sci. 2018;9:513-30
75. Kapoor S, Narayanan A. Leakage in machine learning science. Nat Mach Intell. 2023.
76. Vamathevan J, et al. Machine learning in drug discovery. Nat Rev Drug Discov. 2019;18:463-77.
77. Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018
78. Paul D, et al. Artificial intelligence in drug discovery pipelines. Drug Discov Today. 2021.
79. European Medicines Agency. Al regulatory guidance. 2022.
80. Collins FS, Varmus H. Precision medicine initiative. N Engl J Med. 2015;372:793-5
81. Topol EJ. Deep Medicine. Basic Books; 2019.
82. Kandhare P, Kurlekar M, Deshpande T, Pawar A. A review on revolutionizing healthcare technologies with Al and ML applications in pharmaceutical sciences. Drugs Drug Candidates. 2025;4:9.
83. Khurshid B, Khurshid S, Rafique AM, Javaid M. Al for drug discovery: from algorithms to medicines. J Pharma Bio Med. 2025;3(2):171-184.