

The Role of Artificial Intelligence in Accelerating Drug Discovery and Development

Jeffrey Zhang^{a, d}, Licun Wu^{b, c, d}

Abstract

Artificial intelligence (AI) is rapidly reshaping the pharmaceutical industry through its impact in the way drugs are discovered, developed, and delivered. Traditionally, drug development has been a lengthy, expensive, and failure-prone process, often requiring over a decade and billions of dollars to bring a single therapy to market. AI has the potential to address some of these inefficiencies by supporting faster, more data-driven decision-making in certain areas across the research and development (R&D) pipeline. This review summarizes ten key domains where AI applications are emerging, with varying degrees of demonstrated impact: 1) target identification; 2) hit discovery; 3) lead optimization; 4) preclinical modeling; 5) clinical trial design and stratification; 6) post-marketing surveillance and real-world evidence generation; 7) biomarker discovery; 8) molecular synthesis automation; 9) cost and time reduction; and 10) regulatory decision support. AI techniques - including machine learning, natural language processing, deep learning, and generative models - have shown capability in accelerating *in silico* screening, predicting pharmacokinetic and toxicity profiles, simulating clinical trials, and optimizing molecular design. Additionally, AI is enabling dynamic clinical trial designs, synthetic control arms, and automated patient matching, improving trial success rates. Through automated synthesis planning and robotic chemistry, AI reduces the cycle time from hypothesis to compound validation. Post-market, AI enhances pharmacovigilance by mining electronic health records and social media to detect adverse drug events earlier than traditional systems. As regulatory agencies increasingly accept AI-

derived evidence, the pharmaceutical landscape is transitioning toward more efficient, scalable, and personalized drug development pathways. Despite its momentum, challenges remain, such as data bias, model transparency, and regulatory harmonization. This review underscores AI's potential role in shaping the future of therapeutic innovation and highlights the areas that must be addressed to fully realize its potential.

Keywords: Artificial intelligence; Drug discovery; Machine learning; Clinical trials; Biomarker discovery; Molecular synthesis; RWE; Pharmaceutical innovation

Introduction

The process of drug research and development (R&D) is traditionally time-consuming, expensive, and fraught with high failure rates. On average, bringing a new drug to market can take over a decade and cost upwards of \$2.6 billion, with an estimated failure rate of approximately 90% during clinical trials [1]. The emergence of artificial intelligence (AI) presents a transformative solution, with the potential to optimize various stages of the drug development pipeline - from target identification and molecular design to clinical trial optimization and the generation of real-world evidence (RWE).

AI techniques can process and extract insights from high-dimensional datasets, which are increasingly central to biomedical research. In early-stage drug discovery, deep learning has been applied to identify novel targets and mechanisms by mining high-dimensional omics data and scientific literature [2]. Graph convolutional networks and knowledge graph methods have also been employed to uncover complex relationships between drugs, targets, and side effects [3]. Notably, platforms like BenevolentAI and DeepMind's AlphaFold have enabled structure-based drug design by predicting protein structures with unprecedented accuracy, which significantly enhances target validation and rational drug design [4].

In hit identification and lead optimization, generative AI and molecular property predictors are used to design new chemical entities (NCEs) with favorable pharmacokinetics and safety profiles. Deep generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs), can explore vast chemical spaces and generate candidate compounds that meet predefined activity criteria [5]. Companies like Insilico Medicine and Atomwise utilize these approaches to shorten the hit-to-lead transition phase.

Manuscript submitted August 26, 2025, accepted October 2, 2025
Published online xx xx, 2025

^aDepartment of Immunology, Institute of Medical Science, University of Toronto, Toronto, ON M5S 3K3, Canada

^bLatner Thoracic Surgery Research Laboratories, Division of Thoracic Surgery, Toronto General Hospital, Toronto General Hospital Research Institute, University Health Network, University of Toronto, Toronto, ON M5G 1L7, Canada

^cPrincess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada

^dCorresponding Authors: Jeffrey Zhang, Department of Immunology, Institute of Medical Science, University of Toronto, Toronto, ON M5S 3K3, Canada. Email: jyzhang.887603@gmail.com; Licun Wu, Latner Thoracic Surgery Research Laboratories, Division of Thoracic Surgery, Toronto General Hospital, Toronto General Hospital Research Institute, University Health Network, University of Toronto, Toronto, ON M5G 1L7, Canada. Email: wulicun@hotmail.com

doi: <https://doi.org/10.14740/aicm8>

Table 1. AI Impacts on Drug R&D Identified by ChatGPT

| Stage | AI applications |
|-------------------------|---|
| Target identification | Analyzes omics data and literature to find and validate disease-relevant targets |
| | Speeds up discovery of novel and druggable targets |
| Hit discovery | Screens virtual compound libraries, predicts binding affinities, <i>de novo</i> design |
| | Rapid identification of candidate molecules |
| Lead optimization | Models structure-activity relationships (SAR), predicts ADMET properties, suggests synthetic pathways |
| | Improves efficacy, reduces toxicity, enhances developability |
| Preclinical modeling | Predicts <i>in vivo</i> outcomes, toxicity, and off-target effects |
| | Enhances animal model selection and safety profiling |
| Biomarker discovery | Extracts predictive and prognostic biomarkers from multi-modal data |
| | Enables precision medicine and better patient stratification |
| Clinical trial design | Selects optimal patient cohorts, predicts outcomes, designs adaptive trials |
| | Increases trial efficiency and success rates |
| Real-world evidence | Analyzes EHRs, registries, and wearable data |
| | Improves post-market surveillance and supports drug repurposing |
| Cost and time reduction | Automates processes, reduces lab and trial failures |
| | Shortens development timelines and lowers overall R&D costs |

AI: artificial intelligence; R&D: research and development; ADMET: absorption, distribution, metabolism, excretion, and toxicity; EHRs: electronic health records.

Preclinical testing also benefits from AI-powered *in silico* simulations. Predictive models of absorption, distribution, metabolism, excretion, and toxicity (ADMET) are used to assess compound viability before costly wet-lab experiments. For example, deep neural networks trained on historical toxicology data can anticipate hepatotoxicity or cardiotoxicity, reducing some of the reliance on animal models [6].

In clinical development, AI improves patient stratification and trial design by integrating real-world data (RWD), such as electronic health records (EHRs), medical imaging, and genomic profiles. AI algorithms have been proposed to stratify patients and predict therapeutic outcomes, which could support precision medicine and potentially improve clinical trial design [7]. Adaptive trial designs, guided by AI, allow real-time adjustments to protocols, heightening efficiency and ethical standards.

Post-approval, AI supports pharmacovigilance by detecting adverse drug reactions (ADRs) from EHRs and patient-reported data [8]. In parallel, AI-driven databases such as the Therapeutic Target Database facilitate drug repurposing by integrating drug-target-disease relationships [9]. Collectively, AI applications show promise in accelerating timelines, reducing costs, and supporting personalized approaches, although these benefits remain variably validated. As regulatory bodies like the Food and Drug Administration (FDA) and European Medicines Agency (EMA) begin embracing AI-based tools, the pharmaceutical industry is poised for a paradigm shift, transitioning from intuition-driven to data-driven development.

Using four AI platforms, we compared their functional capabilities in drug development. Each platform generated unique findings (Tables 1-4), and the most common categories were summarized in Table 5 and Figure 1. Six key impact areas and transformative applications in drug development were

described in detail: 1) target identification and drug candidate selection; 2) biomarker discovery and drug design; 3) lead optimization and preclinical testing; 4) clinical trial design and patient stratification; 5) RWE and pharmacovigilance; and 6) economic impact and future directions (Fig. 2).

AI in Target Identification and Drug Candidate Selection

Target identification is a critical and foundational stage in drug discovery, involving the selection of molecular entities - typically genes, proteins, or pathways - that are causally involved in disease pathology. Historically, this process relied on labor-intensive experimental studies and hypothesis-driven research. However, with the advent of high-throughput omics technologies and AI, target discovery has become significantly more data-driven and efficient.

AI enables the integration and interpretation of vast, heterogeneous datasets - including genomics, transcriptomics, proteomics, metabolomics, and EHRs - to identify and prioritize potential drug targets. Graph-based AI approaches - particularly those leveraging knowledge graph embeddings - have increasingly been applied to uncover disease-gene associations and predict therapeutic targets [2, 10]. Knowledge graph-based approaches further link biological pathways, molecular functions, and clinical phenotypes, allowing AI to uncover non-obvious relationships between diseases and potential targets [11].

In addition, AI has been instrumental in mining biomedical literature and clinical trial databases using natural language processing (NLP) techniques. Systems like AlphaFold by

Table 2. AI Impacts on Drug R&D Identified by DeepSeek

| Stage of drug R&D | AI applications |
|-----------------------|---|
| Target identification | Analyzes genomics/proteomics data |
| | Predicts protein structures (e.g., AlphaFold) |
| | Prioritizes high-value targets |
| Drug discovery | Generative AI designs novel molecules |
| | Virtual screening (e.g., Schrodinger) |
| | <i>De novo</i> drug design (e.g., Insilico Medicine) |
| Preclinical testing | Predicts ADMET (toxicity/safety) |
| | AI-driven lab automation (robotics) |
| | Reduces animal testing |
| Clinical trials | AI matches patients using EHRs |
| | Predicts trial success/failure |
| | Monitors adverse effects in real time |
| Drug repurposing | Identifies new uses for existing drugs (e.g., baricitinib for COVID-19) |
| | Saves time/resources vs. developing new drugs |
| Manufacturing | Optimizes production with predictive AI |
| | AI forecasts supply chain demands |
| | Reduces waste and ensures consistent drug supply |

AI: artificial intelligence; R&D: research and development; ADMET: absorption, distribution, metabolism, excretion, and toxicity; EHRs: electronic health records; COVID-19: coronavirus disease 2019.

DeepMind predict protein structures from amino acid sequences with near-experimental accuracy. This structural insight enables rational target selection and downstream structure-based drug design [4]. Ultimately, AI-driven target identification shortens discovery timelines, improves precision, and enhances the likelihood of clinical success by focusing on biologically validated and mechanistically relevant targets.

AI in Biomarker Discovery and Drug Design

Hit discovery - the process of identifying bioactive molecules

that interact with a therapeutic target - is a crucial step in drug discovery. Traditional methods such as high-throughput screening (HTS) are costly, time-consuming, and limited in chemical diversity. AI has emerged as a powerful alternative, enabling *in silico* screening of vast chemical libraries and the generation of novel compounds with desired biological activity [12].

AI models, particularly deep learning architectures like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, can predict molecular activity, physicochemical properties, and binding affinities. These models are trained on large databases such as ChEMBL, ZINC,

Table 3. AI Impacts on Drug R&D Identified by Grok

| Stage | AI applications |
|-----------------------------|--|
| Target identification | Analyzes genomics/proteomics data to identify novel drug targets with high precision |
| Drug design | Predicts molecular interactions to design compounds with optimal binding properties |
| Virtual screening | Screens millions of compounds <i>in silico</i> to identify candidates, reducing experimental costs |
| Predictive toxicology | Assesses toxicity and side effects early, minimizing late-stage clinical trial failures |
| Clinical trial optimization | Improves patient recruitment and predicts outcomes, streamlining trial design |
| Repurposing drugs | Identifies new uses for existing drugs, accelerating development for new indications |
| Personalized medicine | Tailors treatments using patient data, speeding up precision therapy development |
| Automation of data analysis | Processes complex datasets (imaging, chemical, clinical), reducing errors and manual effort |
| Synthesis prediction | Predicts optimal chemical synthesis routes, minimizing iterations and costs |
| Regulatory support | Streamlines documentation and predicts regulatory outcomes, speeding up submissions |

AI: artificial intelligence; R&D: research and development.

Table 4. AI Impacts on Drug R&D Identified by Perplexity

| Stage | AI application |
|--|---|
| Identification of drug candidates | Speeds up discovery of potential compounds |
| Prediction of drug properties and safety | Improves accuracy in forecasting efficacy, toxicity, and interactions |
| Optimization of preclinical and clinical trials | Streamlines trial design, patient recruitment, monitoring, and outcome prediction |
| Integration of experimental feedback (“lab in the loop”) | Enables iterative refinement of drug candidates using AI and lab data |
| Cost and time reduction | Significantly lowers R&D expenses and shortens development timelines |
| Advanced structural biology | AI tools like AlphaFold predict protein structures, aiding targeted drug design |
| Drug repurposing | Identifies new therapeutic uses for existing drugs |
| Personalized medicine | Analyzes genomics, proteomics, and clinical data to tailor treatments |
| Generative AI for molecule design | Designs new molecular structures with desired properties |
| Higher success rates in clinical development | Increases probability of drug candidates progressing through clinical phases |

AI: artificial intelligence; R&D: research and development.

and PubChem, facilitating the virtual screening of millions of compounds in drastically shorter durations [13, 14]. For example, AtomNet, a deep CNN developed by Atomwise, predicts the likelihood that small molecules will bind to protein targets based on three-dimensional (3D) structural data [13].

Generative AI approaches, including VAEs, GANs, and reinforcement learning models, are also being employed to design NCEs. These systems can optimize drug-likeness, synthetic accessibility, and pharmacokinetics simultaneously, significantly accelerating lead identification [5]. Notably, Insilico Medicine used AI to design and validate a fibrosis inhibitor candidate (INS018_055) in less than 18 months - compared to the 4 - 6 years typically required [15]. Furthermore, AI enhances hit-to-lead transition by integrating multi-objective optimization, predicting ADMET properties early in the pipeline, thereby reducing downstream failure rates [6].

Thus, AI-driven hit discovery accelerates the identifica-

tion of promising compounds, reduces experimental workload, and opens the door to previously unexplored areas of chemical space.

AI in Lead Optimization and Preclinical Testing

Lead optimization and preclinical testing represent pivotal phases where promising hits are refined to maximize efficacy, selectivity, and safety. AI accelerates this process through predictive modeling and multi-parameter optimization.

Quantitative structure-activity relationship (QSAR) models enhanced by machine learning (e.g., random forests, graph neural networks) predict how structural modifications affect potency and off-target interactions [16]. Deep learning platforms like DeepTox forecast toxicity endpoints (e.g., hepatotoxicity) with > 80% accuracy, reducing animal testing

Table 5. Summary of Four AI Platforms

| Category | ChatGPT (C) | DeepSeek (D) | Grok (G) | Perplexity (P) |
|-------------------------|-------------------------|-----------------------|-----------------------------|--|
| Target identification | Target identification | Target identification | Target identification | Identification of drug candidates |
| Hit discovery | Hit discovery | Drug discovery | Drug design | Prediction of drug properties and safety |
| Lead optimization | Lead optimization | Preclinical testing | Virtual screening | Optimization of preclinical and clinical trials |
| Preclinical modeling | Preclinical modeling | Clinical trials | Predictive toxicology | Integration of experimental feedback (“lab in the loop”) |
| Biomarker discovery | Biomarker discovery | Drug repurposing | Clinical trial optimization | Cost and time reduction |
| Clinical trial design | Clinical trial design | Manufacturing | Repurposing drugs | Advanced structural biology |
| Real-world evidence | Real-world evidence | - | Personalized medicine | Drug repurposing |
| Cost and time reduction | Cost and time reduction | - | Automation of data analysis | Personalized medicine |
| Molecular synthesis | - | - | Synthesis prediction | Generative AI for molecule design |
| Regulatory and success | - | - | Regulatory support | Higher success rates in clinical development |

The table summarizes the outputs from four AI platforms, intended as an illustrative comparison of emphasis areas rather than a systematic analysis. AI: artificial intelligence.

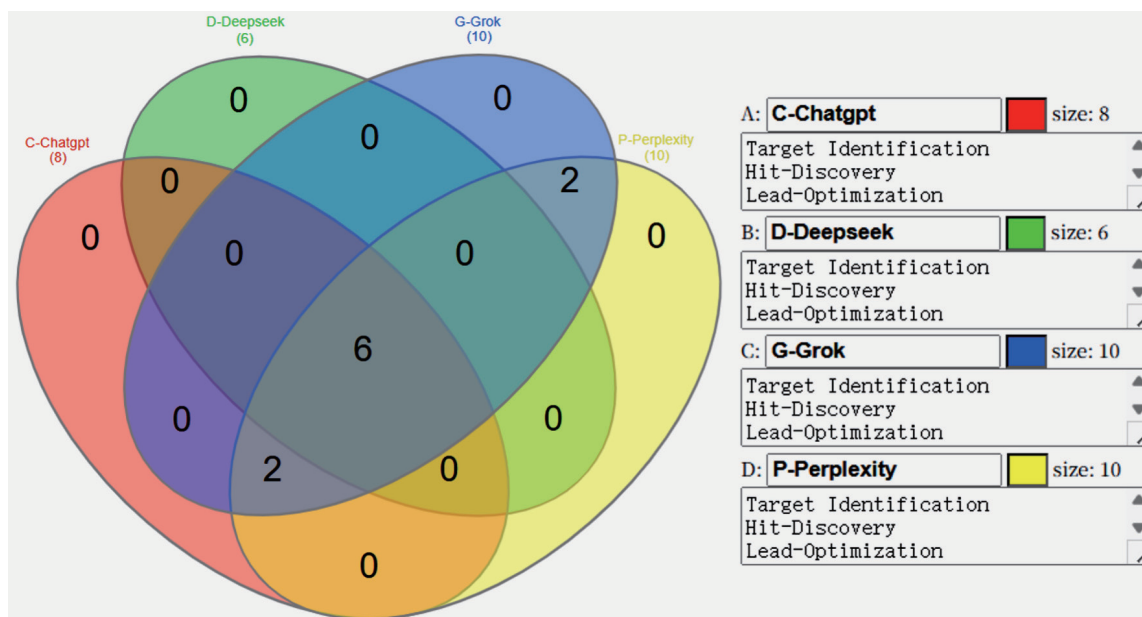


Figure 1. Comparison of the functional capabilities of four AI platforms. Venn diagram illustrating the overlap of identified items among four AI tools: C-ChatGPT (red, eight items), D-DeepSeek (green, six items), G-Grok (blue, 10 items), and P-Perplexity (yellow, 10 items). The central intersection (six items) represents elements common to all four tools. Two additional overlaps are observed between G-Grok and P-Perplexity (two items) and between C-ChatGPT and G-Grok (two items). All other pairwise or triple intersections contain no shared items (0). This visualization highlights the degree of consensus and uniqueness among the AI outputs, emphasizing the core set of commonly identified elements across platforms.

[6]. ADMET prediction tools (e.g., ADMETLab 2.0) integrate physicochemical properties and metabolic stability data to prioritize compounds [17]. Active learning systems iteratively refine synthesis priorities based on experimental feedback, creating closed-loop optimization cycles. For example, Bayesian optimization has reduced the number of required synthesis rounds by 30-50% in kinase inhibitor development [18].

Clinical Trial Design and Patient Stratification

Clinical trials represent one of the most resource-intensive phases of drug development, accounting for approximately 40% of total costs. A significant contributor to high failure rates in clinical trials is suboptimal patient selection and recruitment inefficiencies. AI offers powerful solutions to overcome these challenges by improving trial design, enhancing patient stratification, and optimizing recruitment processes.

One key application is predictive enrollment, where AI leverages NLP to extract eligibility criteria directly from EHRs and clinical notes. AI-driven systems using NLP to extract eligibility criteria from EHRs and unstructured clinical notes can match patients to trials in seconds - in contrast to the hours often required for manual screening - thereby greatly accelerating recruitment timelines [19]. Reports from IBM Watson's AI-powered trial matching system suggest up to an 80% improvement in recruitment efficiency, resulting in faster trial progression and reduced costs, though subsequent independent evaluations have questioned the robustness and generalizability of these findings [20].

Another innovative strategy is the use of synthetic control arms. AI algorithms generate virtual patient cohorts based on historical clinical data and RWE, allowing the reduction or elimination of placebo groups in trials. AI-generated synthetic control arms, such as those developed by Unlearn.AI, are being explored in neurodegenerative disease trials to replace or supplement placebo groups, thus enhancing patient retention and ethical compliance while maintaining statistical rigor [21].

Furthermore, AI-enabled adaptive trial designs - utilizing reinforcement learning and related techniques - can optimize dosing and treatment allocations in real time. Simulations suggest these approaches can significantly reduce adverse events beyond conventional methods [22]. Adaptive designs also improve statistical power and flexibility, increasing the likelihood of trial success.

Together, these AI-driven innovations in clinical trial design and patient stratification offer substantial improvements in efficiency, cost-effectiveness, and ethical standards, marking a significant advance in the future of clinical research.

AI in RWE and Pharmacovigilance

Post-marketing surveillance and pharmacovigilance are critical for ensuring drug safety once therapies reach broader patient populations. Traditional ADR reporting systems rely heavily on voluntary submissions, which often suffer from underreporting and delayed detection. AI has emerged as a transformative technology to enhance RWE generation and improve the detection of safety signals by mining diverse and

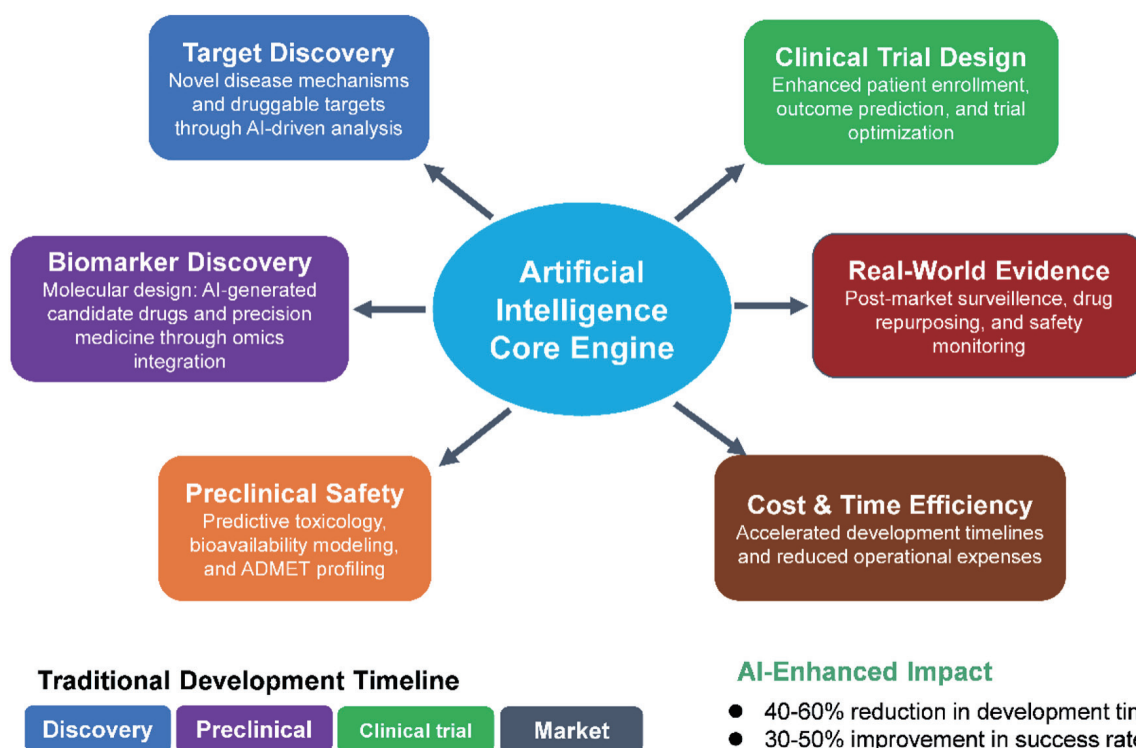


Figure 2. Schematic representation of the AI core engine's role across the drug development pipeline. In the discovery phase, AI enables target discovery by uncovering novel disease mechanisms and identifying druggable targets through advanced computational analysis. Biomarker discovery integrates multi-omics datasets to design molecular candidates and facilitate precision medicine approaches. In the preclinical stage, AI enhances preclinical safety via predictive toxicology, bioavailability modeling, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiling, reducing reliance on extensive *in vivo* testing. During clinical trials, AI improves clinical trial design by optimizing patient selection, predicting outcomes, and refining trial protocols. In the market phase, AI drives RWE generation for post-market surveillance, drug repurposing, and proactive safety monitoring. Across all stages, AI delivers cost and time efficiency through accelerated development timelines, improved decision-making, and reduced operational expenses. Color coding indicates phase categories: discovery (blue), preclinical (orange), clinical trial (green), and market (red/brown). AI: artificial intelligence; RWE: real-world evidence.

large-scale healthcare data sources.

One significant advancement is the use of AI-powered NLP and machine learning algorithms to extract ADRs from EHRs, social media, and patient forums. AI-driven NLP systems have significantly improved automatic ADR detection in text, achieving F-scores above 0.80 in some datasets - indicative of strong performance compared to earlier benchmarks [8]. This improved accuracy enables earlier recognition of safety concerns and more timely regulatory actions.

In addition to ADR detection, AI leverages knowledge graphs and network analysis to predict drug-drug interactions (DDIs), a major source of preventable adverse events. AI models identify high-risk drug combinations with three times greater sensitivity compared to manual expert review, enabling clinicians and regulators to proactively manage polypharmacy risks [2].

Moreover, AI accelerates drug repurposing efforts by rapidly integrating and analyzing heterogeneous datasets. AI-driven drug repurposing platforms, such as BenevolentAI, identified baricitinib as a potential treatment for coronavirus disease 2019 (COVID-19) by early February 2020 - leading to clinical trial initiation within weeks. This rapid timeline is

notably faster than the typical 12 - 18 months often required for traditional drug-repositioning approaches [23]. This rapid repurposing was crucial for timely therapeutic interventions during the health crisis.

Overall, AI-driven RWE generation and pharmacovigilance provide a more dynamic, accurate, and efficient framework for ongoing drug safety monitoring and therapeutic innovation, improving patient outcomes and regulatory responsiveness.

Economic Impact and Future Directions

The integration of AI into drug R&D is generating substantial economic benefits across all phases of the pharmaceutical pipeline. Quantifiable gains in time savings and cost reductions have been reported, though most evidence comes from early-phase studies or modeling rather than late-stage validation.

In the discovery phase, AI enhances target identification, hit discovery, and lead optimization by enabling more efficient analyses of omics data, chemical libraries, and protein-ligand

interactions. Deep learning and knowledge graph-based models have improved the prioritization of viable lead candidates, increasing throughput and reducing early attrition rates [2].

During preclinical testing, AI-powered predictive models support more accurate toxicity and pharmacokinetic profiling, reducing reliance on costly wet-lab assays and improving early decision-making. Studies show that machine learning models can significantly lower toxicity-related attrition, conserving resources and accelerating the transition to clinical trials [17].

In the clinical development phase, AI contributes to more efficient trial design, patient recruitment, and real-time monitoring. Adaptive trial frameworks and synthetic control arms, supported by AI methods, have demonstrated improved statistical power and increased success rates in phase II and III trials [20, 21].

Despite these gains, several challenges remain. AI systems can suffer from data biases, particularly regarding underrepresented populations, which may limit generalizability and equity in drug development outcomes. Addressing these biases is essential to ensure fair and effective AI applications [24]. Moreover, regulatory acceptance requires the development of explainable AI (XAI) frameworks that provide transparency and interpretability in decision-making processes [24].

Reflecting this, the US FDA released draft guidance in 2023 outlining best practices for AI and machine learning in drug development, signaling increasing regulatory support for these technologies [25]. Continued interdisciplinary collaboration, methodological innovation, and regulatory engagement will be pivotal in harnessing AI's full potential to reshape pharmaceutical R&D.

Critical Perspectives on AI in Drug Development

While AI has demonstrated remarkable promise across the drug discovery and development pipeline, its proven clinical impact remains limited. Much of the current literature focuses on early-stage advances and vendor-driven case reports rather than large-scale late-stage clinical validation. For instance, AlphaFold's protein structure predictions have been independently benchmarked and widely adopted in structural biology, representing a genuine breakthrough in target discovery and rational drug design [4]. By contrast, IBM Watson for Oncology - once heralded as a transformative AI platform - was criticized for generating inconsistent or clinically questionable recommendations [26-28]. This divergence illustrates that algorithmic sophistication alone does not guarantee meaningful clinical utility; success depends on robust training data, integration into practice, and independent validation.

When examining outcomes beyond early discovery, it becomes clear that AI's maturity is still emerging. As of 2023, only a few dozen compounds developed with AI assistance have entered clinical trials (mostly in phase I or II), and no AI-discovered molecule had yet publicly achieved full regulatory approval [29, 30]. Notable cases, such as Insilico Medicine's fibrosis inhibitor INS018_055 - which progressed from AI-driven target discovery to phase II clinical trials in approximately 18 months, exemplify how AI is increasingly compressing the

timeline from target identification to clinical entry [15]. Nevertheless, the regulatory impact of AI-enabled compounds is still prospective, underscoring the need to evaluate AI's role by concrete, measurable outcomes such as attrition rates, cost reductions, and trial success rates, rather than potential alone.

The risks of AI deployment in drug development are equally significant. Patient data privacy is a central concern under frameworks like HIPAA and GDPR, requiring strong governance and deidentification protocols. Moreover, training on non-representative datasets risks embedding or amplifying bias. For example, Rajkomar et al noted that models trained on narrow patient populations may underperform for under-represented groups, while Obermeyer et al demonstrated how a widely used clinical algorithm systematically disadvantaged Black patients [31, 32]. These examples underscore the risk of exacerbating healthcare inequities if AI is not carefully audited and diversified. Equally concerning is the "black-box" nature of many models, which can limit interpretability and clinician trust. To address these issues, XAI frameworks and regulatory oversight are critical, and the FDA's 2023 draft guidance on AI/machine learning in drug development provides an important step toward establishing such safeguards [25].

Across the drug development pipeline, the advantages of AI-enabled methods over traditional approaches remain early-stage or projected. In target identification, conventional hypothesis-driven experiments and manual literature mining are now complemented by AI-driven integration of omics, EHRs, and structural biology, with successes such as AlphaFold providing unprecedented accuracy in protein structure prediction [4]. In hit discovery and lead design, labor-intensive HTS is increasingly complemented - and partly supplanted - by virtual screening and generative AI models, which allow the rapid exploration of millions of compounds in days and expand access to novel chemical space [33, 34]. For preclinical testing, predictive toxicology systems such as DeepTox and ADMET-lab 2.0 reduce reliance on animal models and accelerate decision-making, though concerns remain regarding overfitting and generalizability [6, 18]. In clinical trial design, AI tools are increasingly applied to optimize patient matching, predict eligibility, and support adaptive protocols. Studies show these systems can reduce manual screening time and improve recruitment efficiency, particularly by automating eligibility matching and leveraging electronic health record data. While vendor claims of dramatic improvements exist, the peer-reviewed evidence to date only supports incremental but meaningful gains in trial speed and efficiency [34]. Post-market, AI systems are increasingly used in pharmacovigilance, where they can detect ADRs more rapidly than traditional manual reporting methods and reveal hidden drug-drug interactions through large-scale data mining and molecular network approaches [35]. These conclusions remain partly speculative, as very few AI-enabled compounds have progressed beyond early-stage clinical trials or received regulatory approval.

In summary, AI should be viewed as a complementary tool rather than a wholesale replacement for traditional biomedical approaches. Its adoption must be guided by balanced evaluation of successes and failures, transparent reporting of both achievements and limitations, and careful attention to equity and interpretability. Frameworks such as Rayyan can play an

important role in enabling rigorous comparative evaluation of AI versus traditional approaches and highlight them as valuable tools for future work. Only by grounding enthusiasm in empirical validation will AI's role in drug development move from promising potential to proven impact.

Summary

AI is fundamentally restructuring drug R&D from a linear, high-risk process into an iterative, data-driven continuum. Though quantitative projections vary and remain speculative, AI has the potential to significantly shorten timelines and reduce costs in drug development. Realizing this potential requires sustained investment in three key areas: 1) data infrastructure: - developing federated learning systems to access diverse datasets while preserving privacy [36]; 2) regulatory harmonization: - establishing global standards for validating AI-derived evidence [5]; 3) human-AI collaboration: creating hybrid workflows combining computational speed with medicinal chemistry expertise [37].

The pharmaceutical industry is at a stage where AI adoption may contribute to competitive advantages in drug development, though the extent of its long-term impact remains uncertain.

Conclusions

AI is transforming various stages of drug R&D, enabling data-driven, time-efficient, and cost-effective innovation. Through advanced algorithms, AI enhances target discovery, molecular design, and clinical trial execution while improving regulatory alignment and post-market safety. As model transparency improves and regulatory frameworks mature, AI can become essential in the acceleration of therapeutic advancements and precision medicine.

Acknowledgments

None to declare.

Financial Disclosure

None to declare.

Conflict of Interest

The authors have no conflicts of interest to disclose.

Author Contributions

Jeffrey Zhang conceived the research idea and designed the study, wrote and finalized the manuscript. Licun Wu proposed

the overall idea and supervised Jeffrey Zhang to complete the manuscript and helped him finalize all submitted files.

Data Availability

The author declares that data supporting the findings of this study are available within the article.

References

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ*. 2016;47:20-33. [doi pubmed](#)
2. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23(6):1241-1250. [doi pubmed](#)
3. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34(13):i457-i466. [doi pubmed](#)
4. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589. [doi pubmed](#)
5. Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, Terentiev VA, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol*. 2019;37(9):1038-1040. [doi pubmed](#)
6. Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: Toxicity prediction using deep learning. *Frontiers in Environmental Science*. 2016;3:80. [doi](#)
7. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nature Medicine*. 2019;25(1):24-29. [doi](#)
8. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *Journal of Biomedical Informatics*. 2015;53:196-207. [doi](#)
9. Zhou Y, Zhang Y, Lian X, Li F, Wang C, Zhu F, Qiu Y, et al. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res*. 2022;50(D1):D1398-D1407. [doi pubmed](#)
10. Zeng X, Tu X, Liu Y, Fu X, Su Y. Toward better drug discovery with knowledge graph. *Curr Opin Struct Biol*. 2022;72:114-126. [doi pubmed](#)
11. Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. *Sci Data*. 2023;10(1):67. [doi pubmed](#)
12. Walters WP, Murcko MA. Assessing the impact of generative AI on medicinal chemistry. *Nature Biotechnology*. 2020;38:143-145. [doi](#)
13. Wallach I, Dzamba M, Heifets A. AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint*. 2015. <https://arxiv.org/abs/1510.02855>
14. Kimber TB, Chen Y, Volkamer A. Deep learning in vir-

- tual screening: recent applications and developments. *Int J Mol Sci.* 2021;22(9). [doi pubmed](#)
15. Ren F, Aliper A, Chen J, et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nature Biotechnology.* 2025;43:63-75. [doi](#)
 16. Cherkasov A, et al. QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry.* 2014;57(12):4977-5010. [doi](#)
 17. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, Yin M, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 2021;49(W1):W5-W14. [doi pubmed](#)
 18. Shields BJ, Stevens J, Li J, Parasram M, Damani F, Alvarado JIM, Janey JM, et al. Bayesian reaction optimization as a tool for chemical synthesis. *Nature.* 2021;590(7844):89-96. [doi pubmed](#)
 19. Kaskovich S, Wyatt KD, Oliwa T, Graglia L, Furner B, Lee J, Mayampurath A, et al. Automated matching of patients to clinical trials: a patient-centric natural language processing approach for pediatric leukemia. *JCO Clin Cancer Inform.* 2023;7:e2300009. [doi pubmed](#)
 20. Liu R, Rizzo S, Whipple S, Pal N, Pineda AL, Lu M, Arnieri B, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature.* 2021;592(7855):629-633. [doi pubmed](#)
 21. Katsoulakis E, Wang Q, Wu H, Shahriyari L, Fletcher R, Liu J, Achenie L, et al. Digital twins for health: a scoping review. *NPJ Digit Med.* 2024;7(1):77. [doi pubmed](#)
 22. Maier C, Hartung N, Kloft C, Huisinga W, de Wiljes J. Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology. *CPT Pharmacometrics Syst Pharmacol.* 2021;10(3):241-254. [doi pubmed](#)
 23. Ge Y, et al. An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *Sig Transduct Target Ther.* 2021;6:165. [doi](#)
 24. Rajkomar A, et al. Machine learning in medicine. *NEJM.* 2019;380(14):1347-1358. [doi](#)
 25. FDA. Using artificial intelligence & machine learning in the development of drug & biological products. 2023. <https://www.fda.gov/media/167973/download>.
 26. Zhou N, Zhang CT, Lv HY, Hao CX, Li TJ, Zhu JJ, Zhu H, et al. Concordance study between IBM Watson for oncology and clinical practice for patients with cancer in China. *Oncologist.* 2019;24(6):812-819. [doi pubmed](#)
 27. Aikemu B, Xue P, Hong H, Jia H, Wang C, Li S, Huang L, et al. Artificial intelligence in decision-making for colorectal cancer treatment strategy: an observational study of implementing Watson for oncology in a 250-case cohort. *Front Oncol.* 2020;10:594182. [doi pubmed](#)
 28. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application. *Scientific Reports.* 2021;11:5792. [doi](#)
 29. Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. *Drug Discov Today.* 2019;24(3):773-780. [doi pubmed](#)
 30. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today.* 2021;26(1):80-93. [doi pubmed](#)
 31. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine.* 2018;169(12):866-872. [doi](#)
 32. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. [doi pubmed](#)
 33. Macalino SJ, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. *Arch Pharm Res.* 2015;38(9):1686-1701. [doi pubmed](#)
 34. Ocana A, Pandiella A, Privat C, Bravo I, Luengo-Oroz M, Amir E. Integrating artificial intelligence in drug discovery and early drug development: A transformative approach. *Biomarker Research.* 2025;13:45. [doi](#)
 35. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, Jung K, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf.* 2014;37(10):777-790. [doi pubmed](#)
 36. Rieke N, et al. The future of digital health with federated learning. *NPJ Digital Medicine.* 2020;3(1):1-7. [doi](#)
 37. Corrigan-Curay J, et al. RWE and real-world data for evaluating drug safety and effectiveness. *JAMA.* 2018;320(9):867-868. [doi](#)