Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 21, No. 2, 2026

AI Empowered Breakthroughs and Innovations in Drug Development

Luoyi Wang [1], Hongze Tan [2, *], Siqi Wang [3]

[1] College of Chemistry, Northeast Normal University, Changchun 130024, China
[2] College of Earth Sciences, Jilin University, Changchun 130061, China
[3] School of Mechanical and Electrical Engineering, Changchun University of Science and Technology, Changchun 130022, China
***Corresponding author:** 15844174531@163.com

Abstract: The development of novel drugs has long faced the severe challenge of the "Double Ten Law," with pharmaceutical R&D persistently encountering the "triple dilemma" of prolonged cycles, high costs, and low success rates, falling into the trap of "anti-Moore's Law." The rise of artificial intelligence (AI) technology offers a comprehensive solution to this predicament, achieving efficiency revolutions across all stages—from target discovery and molecular design to clinical trials. To outline the technological evolution of AI in drug discovery and design and analyze its role in driving innovative models for new drug discovery, this study conducts a systematic literature review and case analysis. It focuses on the core pain points of traditional drug development, elaborates on the application pathways and technological innovations of AI throughout the drug development process, presents practical achievements and potential enabled by AI based on real-world data, examines current challenges in data, technology, and talent, and outlines future prospects. The findings aim to provide insights for technological innovation and industrial upgrading in the field of drug development.

Keywords: Artificial intelligence; Drug discovery; Target identification; Molecular design; Clinical trials.

1. Introduction

Drug development is a core industry for safeguarding human health, with its innovation level directly impacting breakthroughs in major disease treatment strategies and the refinement of healthcare systems. However, traditional drug development models have long relied on trial-and-error experimentation and empirical accumulation, resulting in a entrenched dilemma characterized by "high investment, prolonged cycles, and low returns." This approach has become increasingly inadequate to meet the growing global health demands and complex challenges in disease prevention and treatment.

The rapid advancement of artificial intelligence (AI) technology is reshaping the paradigm of modern medical practice. With groundbreaking progress in deep learning algorithms and the maturation of big data infrastructure, AI has transitioned from laboratory settings to clinical applications, demonstrating revolutionary potential in critical areas such as disease diagnosis, treatment optimization, and drug development. As AI technology achieves breakthroughs in big data processing, pattern recognition, and predictive modeling, the deep integration of AI with drug development has become an inevitable trend driving industrial transformation [1].

While AI is reshaping the landscape of drug discovery and design at an unprecedented pace, it still faces multiple challenges in achieving full-scale implementation, including data, modeling, regulatory compliance, and ethical considerations. For instance, drug discovery often involves patient privacy and sensitive biological information. How to achieve sharing and utilization while complying with data protection regulations remains a critical issue that AI-driven drug discovery must address. Future efforts should focus on establishing unified standards in algorithm validation, data traceability, and model updates to ensure the legitimacy and controllability of drug discovery and clinical practice.

Against this backdrop, drawing on practical achievements in AI-driven drug development, this paper systematically reviews the latest research advancements of AI in drug discovery and design. It focuses on exploring AI empowerment throughout the entire process from target screening to clinical trials, systematically analyzing its technological applications, outcomes, and challenges. The study also outlines typical models and application cases at different stages, while summarizing the new paradigm of AI-powered drug design and future development trends. These insights aim to provide valuable references for research and industrial practices in related fields.

2. Core Challenges and Cause Analysis in Traditional Drug Development

2.1. Challenges and Current Status

2.1.1. Prolonged cycle duration: Efficiency bottleneck in cross-phase processes

Traditional drug development is a multi-stage, highly complex systems engineering process. The average time from target discovery to final drug approval spans 10-15 years, significantly exceeding innovation cycles in other technological fields. During the preclinical research phase, a series of experiments including target validation, molecular screening, and activity/toxicity testing must be completed, requiring repeated screening and optimization over 3-6 years, often resulting in stagnation due to "in vitro efficacy but in vivo inefficacy." The clinical trial phase is divided into Phase I, II, and III, involving patient recruitment, efficacy validation, and safety assessment. Phase III trials alone may take 3-5 years, during which disease treatment needs may have evolved, extending the cycle to 6-7 years. The prolonged development timeline not only delays drug approval relative to clinical demands but also exposes pharmaceutical companies to dual risks of intensified market competition and shortened patent protection periods, severely impacting investment returns and innovation incentives for research entities.

2.1.2. Exorbitant Costs: Inefficient and wasteful resource allocation

The high costs associated with traditional drug development are reflected in resource consumption throughout the entire process, with total R&D investment averaging over $2 billion. Phase III clinical trials account for a significant proportion of costs, reaching up to 40%, which poses a core challenge in cost control. Preclinical research relies on advanced laboratory equipment and specialized research teams, while clinical trials require funding for medical institutions, participant subsidies, and data analysis expenses. Given the preclinical molecular elimination rate of up to 90%, substantial human and material resources are invested in developing ineffective candidate molecules. Additionally, challenges such as patient recruitment difficulties and trial protocol adjustments during clinical trials further drive up R&D costs. Regulatory agencies worldwide (e.g., FDA, NMPA) impose stringent requirements on R&D processes and data authenticity, compelling companies to allocate significant resources to meet compliance standards. Statistics indicate that global resource wastage due to inefficient R&D exceeds $50 billion annually. Excessive cost barriers limit participation from small and medium-sized research entities, exacerbating resource concentration and innovation barriers in the pharmaceutical development sector.

2.1.3. Low success rate: Risk accumulation across multiple stages

The overall success rate of traditional drug development is merely 10%, with failure risks permeating every stage from target discovery to clinical trials. Target invalidation is the primary cause of R&D failure. Traditional target discovery relies on experimental screening, with 90% of candidate targets being eliminated due to low clinical translational value. Secondly, inaccurate toxicity prediction leads to the termination of numerous candidate molecules during preclinical or clinical trial phases, while insufficient efficacy and excessive adverse reactions in clinical trials further reduce development success rates. The cumulative risks across multiple stages result in a "severe imbalance between input and output" in traditional drug development, constraining the sustainable growth of the pharmaceutical industry.

The core cause of challenges in traditional drug development lies in its "experience-driven, blind screening" research model, which lacks precise understanding of biological system complexity and efficient control over the entire R&D process. With the increasing complexity of disease spectra and the growing demand for precision medicine, the traditional R&D model has become inadequate to meet industrial development needs, urgently requiring paradigm shifts driven by technological innovation.

2.2. Case Analysis

Alzheimer's disease (AD), as a representative of neurodegenerative disorders, has seen drug development as a "hotspot" for traditional models. The consecutive failures of two major projects—Lilly's Solanezumab and Roche's Gantenerumab—have collectively exposed challenges such as preclinical model distortion, limitations of single-target hypothesis, and uncontrolled costs during the R&D cycle [2].

Eli Lilly began targeting β-amyloid (Aβ) in Alzheimer's disease (AD) research in the 1990s. Solanezumab, a monoclonal antibody targeting soluble Aβ monomers, demonstrated significant reduction in Aβ deposition and improvement in memory behavior in preclinical transgenic mouse studies. Subsequently, Phase I safety validation and preliminary exploration of Phase II biomarker positivity were successfully completed. Based on cognitive benefit trends observed in mild patients, Eli Lilly initiated two global Phase III clinical trials involving over 2,000 participants, spanning nearly 20 years. However, results published in 2016 revealed that neither trial met the primary endpoints of cognitive improvement and functional enhancement in daily activities, with only minimal benefits observed in very mild cases, rendering the therapies clinically impractical. The projects were ultimately terminated, with cumulative investments exceeding $1 billion being lost. The core issue lies in the fact that traditional transgenic mouse models could only simulate single pathological features of Aβ deposition, failing to replicate the complex mechanisms underlying human AD, including neuroinflammation, Tau protein abnormalities, vascular damage, and aging processes. This led to clinical translation failures characterized by "animal efficacy but human ineffectiveness." Additionally, the single Aβ-target hypothesis overlooked the multifactorial pathogenesis of AD, fundamentally establishing the basis for research failures.

Following the failure of the Aβ-targeted therapies, the industry shifted focus to Tau protein targets. Roche's Gantenerumab, a monoclonal antibody targeting pathological Tau protein, followed a similar failed trajectory. The drug demonstrated successful reduction of neurofibrillary tangles and improvement in behavioral performance in preclinical Tau transgenic mice, with Phase I/II trials confirming decreased cerebrospinal fluid Tau markers and favorable tolerability. However, both global Phase III clinical trials involving early-stage AD patients (with multi-year follow-ups and large-scale enrollment) failed to achieve the primary endpoint of cognitive function improvement. In 2024, Roche announced the termination of all Tau antibody programs. This failure further highlights the trial-and-error nature of traditional R&D approaches, characterized by jumping from one erroneous target to another without systematic understanding of AD's complex pathological network. Additionally, challenges in early-stage patient screening and prolonged follow-up cycles in AD clinical trials have led to exponential increases in R&D costs and timelines, ultimately resulting in a "the more you invest, the greater the losses" dilemma.

3. AI-Enabled Full-Process Technical Pathways for Drug Development

3.1. Target Discovery: A Breakthrough from "Experimental Screening" to "Data-Driven Approaches"

3.1.1. Predictive Intelligence Technology

Target discovery serves as the starting point in drug development, with its accuracy directly determining the success or failure of research efforts. Traditional target discovery relies on biological experiments and literature mining, which is time-consuming (2-3 years) and exhibits low clinical translation efficiency.

With the deepening application of AI in life sciences, drug discovery has entered a computational innovation phase centered on predictive intelligence [3]. During this period, models primarily rely on large-scale biomedical data for feature learning and mechanism inference to achieve rapid predictions of molecular properties, target interactions, pharmacokinetic characteristics, and toxicity risks. Numerous algorithms that automatically analyze chemical structures, protein sequences, and multi-omics data have significantly improved the efficiency of tasks such as DTI prediction, ADMET modeling, virtual screening, and disease target identification [4], driving the transition of drug discovery from "experience-dependent" to "intelligent computation" [5-7], as shown in Table 1.

Table 1 Typical application directions of predictive AI in drug discovery

Direction of application	Primary mission	Common models	Core Contributions
DTI calculate	Prediction of small molecule-protein binding relationships	CNN, RNN, Long Short-Term Memory (LSTM), Transformer	Capture sequence and structure dependencies to enhance prediction accuracy of binding modes and affinity
Molecular Properties and ADMET Prediction	Toxicity, solubility, and metabolic stability modeling	GNN, CNN, RNN	Automatically extract molecular features to achieve multi-attribute joint prediction
Reaction Prediction and Synthesis Path Planning	Based on the reaction of reactants to form products or synthetic pathways	Transformer, Seq2Seq	Capture reaction selectivity to achieve high-precision product prediction

3.1.1.1 DTI prediction

DTI prediction is one of the core components in drug discovery, aiming to elucidate the interaction relationships between small-molecule drugs and protein targets. Predictive deep learning methods can automatically learn interaction patterns from large-scale datasets by simultaneously representing drug structures and protein sequence features.

Early sequence-based interaction prediction studies primarily utilized drug SMILES structures or molecular fingerprints combined with protein amino acid sequences, employing deep neural networks to learn interaction patterns. Convolutional neural networks (CNNs) extracted local structural features from molecular graphs and protein sequences, achieving higher prediction accuracy than traditional machine learning methods. With advancements in Graph Neural Networks (GNNs), models can directly perform convolutional operations on molecular graph structures, enabling end-to-end learning of drug-protein structural dependencies to more accurately simulate binding processes. Some studies further integrated Transformer architectures with self-attention mechanisms to capture long-range interaction signals, enhancing model generalization capabilities and interpretability in complex sequence pair modeling.

Structure-based interaction prediction has evolved alongside advancements in protein structure prediction technologies, with DTI (Differential Tandem Interaction) expanding from two-dimensional sequence models to structural hierarchies. The AlphaFold series models demonstrated breakthroughs in deep learning for protein folding prediction [8], enabling large-scale reliable three-dimensional protein structure determination. Building upon this foundation, numerous DTI prediction methods further utilize three-dimensional structural pockets, ligand orientations, and local chemical environments as input features to enhance prediction accuracy for binding modes and binding energies. Convolutional Neural Networks (CNNs) capture local geometric and physicochemical environmental characteristics through spatial convolution on protein-ligand 3D grids, facilitating prediction of stable ligand-binding conformations within protein pockets. Graph Neural Networks (GNNs) effectively model atomic-atomic interaction patterns by transmitting messages across atomic-level graph structures, demonstrating superior physical interpretability in binding energy predictions. Recent years have seen the introduction of novel architectures such as Transformer architectures and equivariant networks (E(3)-Equivariant Networks), which maintain spatial rotation and translation invariance while capturing more complex global dependencies.

3.1.1.2 Molecular Characteristics Other Prediction Methods

Molecular property prediction and ADMET (absorption, distribution, metabolism, excretion, toxicity) characterization are critical tasks in early drug screening. In molecular graph modeling, Graph Neural Networks (GNNs) have become the mainstream approach. By employing message-passing mechanisms between nodes and edges, GNNs progressively aggregate information about atoms and their local environments, enabling precise representation of chemical bond relationships and molecular topological structures. This facilitates effective prediction of molecular solubility, toxicity, pharmacodynamics, and physicochemical properties. For character-based molecular representations (e.g., SMILES sequences), Recurrent Neural Networks (RNNs) and Transformer models demonstrate superior performance by capturing contextual dependencies among atomic symbols within sequences, thereby learning chemical semantic patterns and enabling property prediction. Studies [9-10] indicate that both language models possess the capability to learn complex molecular distributions, though their application scenarios differ: RNNs are more suitable for datasets emphasizing local features, while Transformer models excel in modeling global dependencies or processing multi-distribution chemical spaces. Additionally, Convolutional Neural Networks (CNNs) based on two-dimensional molecular images have been employed to extract local structural features, providing complementary approaches for ADMET attribute modeling and molecular property prediction.

3.1.2. Multimodal Data Fusion

AI technology achieves precise and efficient target discovery through multimodal data fusion and intelligent algorithm modeling [11]. In terms of technical approaches, AI integrates genomics, proteomics, transcriptomics, and clinical data to construct disease mechanism association networks, overcoming the limitations of traditional single-indicator assessments. For instance, in tumor target discovery, algorithms not only analyze genetic mutation data but also incorporate multidimensional information such as protein interaction patterns and gene expression levels to comprehensively elucidate the association mechanisms between targets and diseases. Meanwhile, the application of incremental learning technology enables models to rapidly update parameters when new clinical case data are added, eliminating the need for full model retraining and perfectly aligning with the rapid accumulation characteristics of biomedical data.

In particular, the groundbreaking advancements of the AlphaFold series tools in protein structure prediction have not only enabled structure-based drug design but also significantly accelerated the evaluation of drugability for targets. During the target identification phase, protein function annotation serves as the foundation for understanding disease mechanisms and discovering potential intervention nodes. These methods can not only predict whether interactions will occur but also generate finer-grained structural biology information such as binding sites, binding conformations, or binding energies, providing more direct evidence for virtual screening and pharmacodynamic optimization. With continuous progress in predictive AI models, DTI has evolved from sequence-based coarse-grained prediction to a refined modeling system capable of capturing three-dimensional structures, pocket mechanisms, and physical interactions, substantially advancing the intelligent processes of target identification, drug repurposing, and virtual screening.

3.1.3. Virtual Screening

The known compound space is estimated to encompass 10^60 molecules, whereas conventional average random screening of 10,000–20,000 compounds typically yields only 1–2 potential lead compounds. Virtual screening (VS) rapidly identifies potentially active molecules from large-scale compound libraries through intelligent computational simulations. By predicting compound-target interactions and binding affinities, AI enables rapid screening of billions of molecular libraries via virtual screening [12].

3.2. Molecular Design: Innovation from "Blind Synthesis" to "Directed Generation"

3.2.1. Generative AI and Molecular Design

Candidate molecule design constitutes a pivotal step in drug development. Traditional approaches rely on empirical trial-and-error methods by researchers, enabling only the synthesis of dozens of candidate molecules within 24 hours, with 90% being eliminated due to insufficient activity, poor metabolic stability, or excessive toxicity. The emergence of generative models marks a shift in drug discovery from "predicting the optimal" to "creating the optimal." The core of this paradigm shift lies in models that no longer merely evaluate the potential of existing compounds but directly learn chemical space distributions to generate novel molecular structures conforming to target characteristics from the latent space. AI technologies leverage generative algorithms and multi-objective optimization models to achieve enhanced efficiency and precision in molecular design.

Breakthroughs in generative AI technology serve as the core driving force for molecular design innovation. Leveraging advanced algorithms such as diffusion models and generative adversarial networks (GANs), AI systems can generate 100,000 drug-like molecules per 24 hours, significantly enhancing the supply efficiency of candidate molecules. Simultaneously, by integrating physical laws with multi-omics data, AI constructs multi-objective equilibrium optimization models capable of simultaneously optimizing molecular bioactivity, metabolic stability, toxicity, and drug-likeness, addressing the traditional design challenge of "single-optimal metrics at the expense of overall performance." Data demonstrate that AI technology reduces preclinical toxicity elimination rates from 90% to 65%, markedly improving candidate molecule quality. This article systematically summarizes mainstream generative model types, as shown in Table 2.

Table 2. Characteristics and Application Directions of Mainstream Generation Models in Drug Molecule Design

Types of models	Core Idea	Superiority	Typical representative	Direction of application
VAE (variational autoencoder)	Learning latent spaces and reconstructing molecular structures	High controllability and easy optimization	Junction TreeVAE	Molecular Formation and Optimization
GAN (generative adversarial network)	Adversarial Training of Generators and Discriminators	Capable of generating diverse molecules	MolGAN	New molecular formation, enhanced pharmacological efficacy
Flow-Based Model (Flow Model)	Reversible mappings enable exact likelihood calculations	Stable training, structural continuity	GraphNVP	Molecular Continuous Space Generation
Diffusion Model	Gradual denoising and distribution recovery	High fidelity and strong generative diversity	EDM, MDM	3D Molecular Conformation Generation
LLM (Large Language Model)	Based on SMILES/SELFIES/text understanding chemical semantics	High versatility and support for multimodal applications	MolT5, Token-Mol, DrugChat, TCMChat	Text-to-molecule generation, compound comprehension

3.2.2. Typical Applications

In the field of modernization of traditional Chinese medicine (TCM), AI technology has achieved leapfrog development in the analysis of TCM components. Leveraging ultra-large-scale data, AI algorithms as the engine, and specialized knowledge bases as support, it transforms the previously inexhaustible space of TCM components into innovative resources that can be analyzed, screened, optimized, and converted. With tools such as the "Bencao Think Tank" and the "GNDC Component Database," AI has identified over 220 million components from the genomes of 1,037 medicinal species, increasing the efficiency of TCM metabolite analysis by 1,000-fold. This represents not only a technological breakthrough but also marks the true entry of TCM modernization into a new era driven by data, intelligent discovery, and precision research and development. It lays an unprecedented resource and technological foundation for improving TCM quality, secondary development of classic prescriptions, and the research and development of innovative drugs derived from natural products.

3.3. Clinical Trials: Transition from "Time-consuming and Costly" to "Precision and Efficient"

Clinical trials, as an indispensable core component in the new drug development pipeline, serve as a critical means to validate drug safety, efficacy, and applicability. Their research model and implementation efficiency directly determine the time-to-market cycle, R&D costs, and clinical translation success rates of novel drugs. For decades, traditional clinical trials have been trapped in the developmental dilemma of being "time-consuming, costly, inefficient, and high-risk," becoming a core bottleneck that constrains the acceleration of pharmaceutical innovation. Under the experience-driven R&D model, Phase II-III clinical trial cycles typically span 3-5 years, accounting for over 60% of the entire new drug development cycle, with clinical phase investments representing more than 60% of total R&D costs. A single clinical trial failure can result in losses ranging from hundreds of millions to billions of dollars. Concurrently, challenges such as patient recruitment difficulties, prolonged enrollment cycles, and recruitment delays caused by narrow indication definitions and complex exclusion criteria are widespread. Trial protocol design relies heavily on investigators' prior experience, with core elements like dose gradients, sample size estimation, and endpoint selection lacking scientific data support, often leading to ineffective trials and resource wastage. Multi-center clinical trials face significant challenges in integrating multi-source data (e.g., electronic medical records [EMR], imaging data, laboratory test results), frequently encountering issues like data gaps, logical inconsistencies, and outliers. Quality control efforts predominantly depend on post hoc manual verification, resulting in low efficiency and potential omissions. Adverse event alerts are often reactive measures taken after incidents occur, making it difficult to preemptively identify potential safety risks—posing threats to participant rights and potentially leading to trial suspensions or terminations midway through the process. These overlapping pain points result in a failure rate exceeding 80% during traditional clinical trial phases, creating a formidable 'valley of death' that poses significant challenges in novel drug development. This severely hinders the rapid translation of innovative drugs and their clinical accessibility. AI technology addresses the core bottlenecks of clinical trials through precise matching, virtual simulation, and dynamic monitoring.

3.3.1. AI-Optimized Clinical Treatment

AI technology is optimizing clinical trial design and execution processes. The AI patient monitoring platform developed by AiCure ensures medication adherence by analyzing patients 'facial expressions and behavioral patterns, reduces data bias, and improves trial success rates [13]. Medable's decentralized clinical trial platform [14] supports remote patient participation, integrates wearable devices for health data monitoring, enhances patient convenience and data collection speed, and accelerates overall research progress. AI drives precision drug development through patient data analysis. For example, GNS Healthcare utilizes machine learning algorithms to analyze large-scale patient data [15] to identify individual differences and predict drug responses, enabling pharmaceutical companies to more accurately select medications tailored to specific patient populations and optimize treatment regimens.

In patient recruitment, AI leverages real-world data and natural language processing (NLP) analysis of electronic medical records to rapidly identify eligible patients, achieving a matching accuracy rate of 95%. The CTMS system at Nanfang Hospital of Southern Medical University, empowered by AI, reduced patient screening time by 97.8% and shortened the recruitment cycle by 60%, effectively addressing the inefficiency issue of "searching for a needle in a haystack" in traditional recruitment methods.

3.3.2. Outcomes

The advancement of virtual clinical trial technologies has further reduced R&D costs and risks. By constructing virtual patient models through digital twin technology, the efficacy and adverse reactions of drugs can be simulated across different populations. Certain scenarios can replace animal experiments and early-stage clinical trials, not only shortening trial cycles but also mitigating ethical controversies and resource consumption. The AI-driven transformation of clinical trials has been thoroughly validated through industry practice, demonstrating quantifiable improvements in efficiency and value: clinical trial cycles are on average shortened by 30%-60%, overall R&D costs reduced by 20%-40%, patient enrollment completion rates significantly increased, delay rates substantially decreased, and the conversion rate from Phase II to Phase III clinical trials markedly improved. Concurrently, data quality and compliance have been notably enhanced, better meeting regulatory requirements such as those from the U.S. Food and Drug Administration (FDA) and China's National Medical Products Administration (NMPA). Currently, AI is no longer merely an auxiliary tool in clinical trials but has become a core engine deeply integrated into critical stages including trial design, execution, quality control, and decision-making. It drives not only efficiency gains but also paradigm shifts in pharmaceutical R&D. In the future, with continuous AI technological iterations and deeper integration with big data, the Internet of Things (IoT), and precision medicine, clinical trials will evolve toward greater precision, efficiency, safety, and accessibility. This will persistently address core challenges in drug development, accelerate clinical translation of innovative therapies, and provide critical support for overcoming complex diseases and enhancing public health service capabilities.

4. Practical Outcomes and Industrial Impact of AI-Enabled Drug Development

4.1. Core Outcomes: Triple Optimization of Cycle Time, Cost, and Success Rate

The effectiveness of AI technology in shortening R&D cycles can be further substantiated through case studies and industry reports,as shown in Table 3.The "AI Pharmaceutical Industry Report" released by PharmaCube in 2025 reveals that the average development cycle for traditional innovative drugs spans 12 years, whereas with deep AI integration, this period can be compressed to 3-5 years. Some projects targeting well-defined therapeutic targets have even achieved breakthroughs in "drug development within 100 days" —for instance, an AI-driven pharmaceutical company successfully completed the entire process from target identification to Investigational New Drug (IND) application submission for a specific oncology target in just 98 days, demonstrating efficiency gains far surpassing conventional models. Regarding cost control, beyond the global annual R&D expenditure savings mentioned earlier, Deloitte's 2025 research report indicates that the average cost of developing a new drug has decreased from $2.23 billion under traditional models to below $1.5 billion with AI empowerment. The most significant cost reduction occurred during clinical trials, where the proportion of Phase III clinical trial expenses dropped from 40% to 28% due to AI-powered patient recruitment and virtual trial technologies.

Empirical data on improved success rates are equally abundant. A 2024 study published in the Journal of China Pharmaceutical University indicated that the introduction of AIDD (Artificial Intelligence Drug Design) and CADD (Computer-Aided Drug Design) technologies increased the overall success rate of drug development from 10% to 14%, with success rates reaching as high as 22% in specific fields such as antiviral drug development—thanks to AI's multi-dimensional prediction of molecular activity and toxicity, which reduced clinical failures caused by poor drugability. Additionally, the FDA-led CiPA program improved cardiac toxicity prediction accuracy to 86% by integrating AI models with multi-ion channel detection data, effectively lowering the probability of drug development being terminated due to safety concerns.

Table 3. Comparison of AI-empowered drug development data

Metric	Traditional Drug Development	AI Empowering Drug Development	Amplitude of optimization / Key highlights	Data sources
Average R&D cycle	12 years	3-5 years (some cases 98 days)	Reduction of 58%-75% with clearly defined targets in the "100-Day Drug Development" project	Pharmaceutical Magic Cube 2025 Industry Report
Average R&D cost	2.23 billion dollars	< 1.5 billion dollars	Decreased by 32.7%, the proportion of Phase III clinical trial costs declined	Deloitte 2025 Cost Report
Cost proportion in Phase III clinical trials	40%	28%	A decrease of 12 percentage points	Deloitte 2025 Cost Report
Overall R&D success rate	10%	14% (antiviral drugs 22%)	Increase by 40%, more than double in specific fields	Journal of China Pharmaceutical University 2024
Accuracy of cardiac toxicity prediction	-	86%	Effectively reduce the risk of safety termination	FDA CiPA Program

4.2. Industrial Impact: Rapid Development of Market Size and Innovation Ecosystem

The growth trends of the global AI pharmaceutical market can be further refined through sector-specific data. According to the industry analysis released by Zhiyuan Community in 2025, "target discovery" and "clinical trial optimization" are the two fastest-growing segments in the global AI pharmaceutical market, with market sizes reaching $820 million and $750 million respectively by 2025. It is projected that by 2026, the combined share of these two segments will exceed 60% of the global AI pharmaceutical market [16]. In the China market, apart from the 108 AI pharmaceutical companies mentioned in the original text, the industrial chain division has shown a trend toward specialization: approximately 30% of enterprises focus on AI algorithm development (such as molecular generation models and predictive models), 45% concentrate on drug development for specific disease areas (such as oncology and neurodegenerative diseases), while the remaining enterprises provide data services and experimental validation support, forming a complete collaborative ecosystem of "algorithm-development-validation".

In cross-industry collaboration, global partnerships between AI companies and traditional pharmaceutical firms have seen a 40% year-on-year increase since 2025. A notable example involves an AI company collaborating with multinational drugmakers to develop Alzheimer's disease treatments. By leveraging AI to optimize molecular structures, the candidate drug achieved a threefold increase in blood-brain barrier penetration rate while reducing toxicity by 50%. This collaborative model not only mitigates R&D risks for traditional pharmaceutical companies but also provides real-world validation scenarios for AI technologies, accelerating technological implementation and iterative improvements.

5. Core Challenges in AI-driven Drug Development

5.1. Data barriers: Insufficient supply of high-quality data

The scarcity of high-quality annotated data is particularly prominent in the field of modernization of traditional Chinese medicine. A 2024 study published in the Journal of China Pharmaceutical University pointed out that although AI has identified 220 million components from the genomes of 1,037 medicinal species, metabolic omics data for over 60% of medicinal plants (such as some rare and endangered species) remain incompletely collected. Moreover, only 15% of existing traditional Chinese medicine component data include annotations of "component-target-efficacy" associations, making it difficult for AI to accurately construct models of traditional Chinese medicine mechanisms of action. The issue of inconsistent multi-center data standards can be illustrated through specific scenarios: In a cross-hospital clinical trial project in 2025, due to differences in the "adverse reaction description" field format of electronic medical records across hospitals (some hospitals used structured data, while others used free text), the accuracy rate of patient enrollment matching for AI natural language processing models dropped from 95% in single-center trials to 78% in multi-center trials.

5.2. Technical Limitations: Key Technologies Require Breakthroughs

The limitations of protein structure prediction are particularly evident in the field of complex membrane proteins. A 2025 study by the Zhiyuan Community revealed that for G protein-coupled receptors (GPCRs) containing multiple transmembrane domains, the structural prediction bias rate of AI models (such as AlphaFold3) reached 18%, directly impacting the accuracy of molecular docking. A team designed small molecule inhibitors based on predicted structures, which exhibited only 30% of expected activity in experiments [17]. The accessibility of molecular synthesis also poses significant challenges. Statistics from the Journal of Medicinal Chemistry in 2024 indicate that approximately 25% of AI-generated candidate molecules are difficult to synthesize using existing organic synthesis techniques due to the presence of "complex ring systems" or "special functional groups." Even when successfully synthesized, these compounds incur costs 5-10 times higher than those of conventional molecules.

The lack of model interpretability has become a significant barrier in regulatory approval processes. The FDA's 2025 "Guidance on AI-Assisted Drug Development (Draft)" explicitly states that approximately 40% of AI-assisted drug development applications require additional data due to "inability to clearly explain model decision-making logic." A typical case involved an AI-designed anticancer drug that demonstrated favorable clinical trial efficacy but faced a six-month approval delay because it could not explain "why the molecule was more effective for patients with specific genetic mutations."

5.3. Talent Gap: Shortage of interdisciplinary professionals

The shortage of interdisciplinary talents can be reflected through educational qualifications and professional background data. The 2025 Global AI Pharmaceutical Talent Report shows that among practitioners in the global AI pharmaceutical field, those with both a "Master's degree or higher in Computer Science/Mathematics" and "pharmaceutical/medical-related work experience" account for less than 8%. In China, this proportion is only 5%, and 70% of interdisciplinary talents are concentrated in top-tier enterprises in first-tier cities such as Beijing, Shanghai, Guangzhou, and Shenzhen, leaving small and medium-sized R&D institutions facing the dilemma of "difficulty in recruitment and even greater difficulty in retaining talent" [18]. In terms of talent cultivation, although over 50 universities worldwide have introduced courses related to "AI drug development," the number of graduates in 2025 only meets 30% of market demand. Additionally, approximately 60% of graduates lack experimental validation capabilities and require 1-2 years of additional training from enterprises before they can work independently.

6. Conclusion

The "triple dilemma" in traditional drug development has become a core bottleneck constraining the advancement of the pharmaceutical industry, while AI technology's end-to-end empowerment provides an effective solution to this challenge. Through technological innovations in target discovery, molecular design, and clinical trials, AI has achieved shortened drug development cycles, reduced costs, and improved success rates, driving structural upgrades and ecological restructuring in the pharmaceutical industry. Although current AI-driven drug development faces challenges such as data barriers, technical limitations, and talent shortages, its promising prospects are underpinned by deepening interdisciplinary integration and the establishment of an independent R&D ecosystem. In the future, AI will continue to lead the technological revolution in drug development, delivering more efficient and safe innovative drugs for major disease treatments and providing core support for the enhancement of human health protection systems.

References

[1] Li, X., Yuan, Q., Zhu, F., Wang, L., Zhang, Y., & Chen, H. (2025). Current status and future prospects of artificial intelligence application in clinical practice. China Clinical Research, 38(5), 657-662. https://doi.org/10.13429/j.cnki.cjcr.2025.05.001

[2] Wu, M., Chen, X., Lin, L., Huang, J., & Zheng, Y. (2025). Application and prospects of artificial intelligence technology in drug design. China Prescription Drug, 23(24), 120-123.

[3] Zhang, J., Shao, X., & Fan, X. (2026). From predictive intelligence to generative intelligence: A new landscape of drug discovery and design led by artificial intelligence. Chinese Pharmaceutical Affairs, 40(1), 32-44. https://doi.org/10.16153/j.1002-7777.2025-11-0011

[4] Xu, Y., & Pei, J. (2017). Application of deep learning in cheminformatics. Big Data, 3(2), 45-66.

[5] Heid, E., Greenman, K. P., Chung, Y., Li, S., Graff, D. E., Vermeire, F. H., Wu, H., Green, W. H., & McGill, C. J. (2023). Chemprop: A machine learning package for chemical property prediction. Journal of Chemical Information and Modeling, 64(1), 9-17. https://doi.org/10.1021/acs.jcim.3c01250

[6] Wieder, O., Kohlbacher, S., Kuenemann, M., Garon, A., Ducrot, P., Seidel, T., & Langer, T. (2020). A compact review of molecular property prediction with graph neural networks. Drug Discovery Today: Technologies, 37, 1-12. https://doi.org/10.1016/j.ddtec.2020.11.009

[7] Scalia, G., Grambow, C. A., Pernici, B., Li, Y. P., & Green, W. H. (2020). Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. Journal of Chemical Information and Modeling, 60(6), 2697-2717. https://doi.org/10.1021/acs.jcim.9b00975

[8] Wang, W., & Huang, X. (2025). Artificial intelligence-assisted drug design. China Journal of Bioengineering, 45(11), 1-9. https://doi.org/10.13523/j.cb.202505046

[9] Chen, Y., Wang, Z., Zeng, X., Huang, Y., & Li, Y. (2023). Molecular language models: RNNs or transformers? Briefings in Functional Genomics, 22(4), 392-400. https://doi.org/10.1093/bfgp/elad012

[10] Liu, H., Zhu, B., Nie, S., Li, M., Zhou, J., Zhang, Y., Chen, W., & Wang, L. (2025). Advancing ADMET prediction through multiscale fragment-aware pretraining with Msformer-ADMET. Briefings in Bioinformatics, 26(5), bbaf506. https://doi.org/10.1093/bib/bbaf506

[11] Zhu, B., Li, Z., Jin, Z., Wang, X., Liu, H., Chen, G., & Zhang, Y. (2024). Knowledge-based in silico fragmentation and annotation of mass spectra for natural products with MassKG. Computational and Structural Biotechnology Journal, 23, 3327-3341. https://doi.org/10.1016/j.csbj.2024.08.022

[12] Zhou, G., Rusnak, D. V., Park, H. S., Yuan, Q., Wang, F., Zhang, L., & Wu, J. (2024). Artificial intelligence accelerated virtual screening platform for drug discovery. Nature Communications, 15, 7761. https://doi.org/10.1038/s41467-024-52033-7

[13] Ghozali, M. T. (2024). Predicting patient adherence in healthcare using artificial intelligence and machine learning techniques: A narrative review. In 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (pp. 1211-1215). IEEE.

https://doi.org/10.1109/I-SMAC61858.2024.10714872

[14] Crescenzo, R., Idoine, A., By, K., Viquez, A., & Chang, J. (2023). OP112 evaluation of a self-administered smart phone-based application as a wellness measure in a clinical trial of zanubrutinib. International Journal of Technology Assessment in Health Care, 39(S1), S31.

https://doi.org/10.1017/S026646232300115X

[15] Smalley, E. (2017). AI-powered drug discovery captures pharma interest. Nature Biotechnology, 35(7), 604-605. https://doi.org/10.1038/nbt.3914

[16] Zhiyuan Community. (2026). 1950s vs today: Productivity crisis in drug development and AI solutions [Report].

[17] Deloitte. (2025). Global drug R&D cost report 2025 [Report].

[18] Chen, B., Lü, L., Xu, X., Wang, Y., Zhang, M., & Liu, Z. (2024). Reflections on improving drug success rates with AIDD and CADD. Journal of China Pharmaceutical University, 55(3), 284-294.