Current Science
CS 5 (4), 3762-3773 (2025)
Received 20April 2025 | **Revised 19 May** 2025 | **Accepted 22 June**. 2025 | **Online Available 31 July** 2025
https://doi.org/ 10.52845/CS/2025-5-4-36

Current Science

Current Science

OPEN ACCESS

ISSN(O)2795-8639

Original Article

Applications of Artificial Intelligence in Drug Development and Diagnosis for Rare Diseases Current Status, Advantages, and Future Prospects

Wencheng Fu[1]*[#], Kayden Zhong[1]*, Bin Wen[1#]

[1]Shanghai Pharma Rare Disease Medicine Co., Ltd., Shanghai 201306, China

***Wencheng Fu and Kayden Zhong Made Equal Contributions to this Work.**

[*]Corresponding Author: Wencheng Fu, Bin Wen

Abstract:

Rare diseases, characterized by low prevalence and complex pathophysiological mechanisms, pose significant challenges in diagnosis and therapeutic development. Globally, an estimated 0.35 billion individuals are affected by rare diseases, with approximately 80% having a genetic basis. Conventional drug development approaches are often inadequate due to their high costs, long timelines, and low success rates. In recent years, rapid advancements in artificial intelligence (AI) have introduced transformative opportunities in the study and treatment of rare diseases. By integrating heterogeneous data sources—including genomics, proteomics, medical imaging, and electronic health records—AI has demonstrated substantial potential across multiple domains, such as biomarker discovery, target identification, virtual screening, drug repurposing, RNA-based therapeutics design, and clinical trial optimization. Notably, AI accelerates early-stage drug development, significantly reducing the time required for target validation and candidate selection while improving success rates. In diagnostics, AI leverages deep learning and natural language processing to enhance the accuracy and efficiency of genomic variant interpretation and phenotypic pattern recognition. Furthermore, AI contributes to clinical trial design by facilitating patient recruitment, optimizing trial protocols, and enabling the use of digital twin models to reduce the required sample size—an especially valuable capability given the limited patient populations in rare disease research. This review systematically examines key AI-driven strategies and their applications throughout the drug development pipeline for rare diseases. It also highlights how AI addresses critical challenges such as data scarcity, heterogeneity, and the need for personalized therapeutic interventions. Several representative case studies (e.g., AIT-101, PXT3003, REC-4881) illustrate the practical impact and future potential of AI in pharmaceutical innovation. As AI technologies continue to evolve and receive growing regulatory and policy support, they are poised to become a central force in advancing the diagnosis and treatment of rare diseases.

Keywords: AI, Rare Diseases, Drug Development, Biomarkers, Precision Medicine

Introduction

Rare diseases refer to a diverse group of conditions characterized by low prevalence and a relatively small number of affected individuals. According to the World Health Organization (WHO), a disease is classified as rare if it affects 0.65‰ to 1‰ of the global population [1]. Although individually uncommon, rare diseases collectively represent a significant public health challenge: over 5,000–7,000 distinct rare diseases have been identified, affecting approximately 350 million people worldwide. Of these, approximately 80% have a genetic origin and often manifest in childhood, leading to severe functional impairments or life-threatening complications.

The development of therapies for rare diseases

presents substantial challenges due to limited patient populations, insufficient clinical data, and complex pathophysiological mechanisms. Conventional drug development pipelines are particularly inefficient in this context, with an average timeline exceeding ten years and costs reaching billions of USD, yet yielding a success rate of less than 10% [2]. Moreover, diagnostic delays are common, with many patients facing years of inconclusive evaluations and misdiagnoses before receiving a definitive diagnosis.

In recent years, artificial intelligence (AI) has emerged as a powerful tool in biomedical research, offering transformative potential in the study and treatment of rare diseases. By enabling the integration of heterogeneous data sources, accelerating target identification, optimizing drug repurposing and clinical trial design, AI technologies have demonstrated significant advantages in both drug development and diagnostic decision-making. This paper provides a systematic review of AI applications in the field of rare diseases, focusing on its role across various stages of drug discovery, its utility in diagnostic support systems, and highlighting representative case studies that illustrate current progress and future directions.

2. AI Accelerating Drug Development for Rare Diseases: Key Strategies

2.1 Biomarker Discovery

Biomarker discovery has become a pivotal driver in advancing clinical research and precision medicine for rare and complex diseases. A biomarker is defined as a measurable biological characteristic that reflects normal physiological processes, pathophysiological changes, or responses to therapeutic interventions. These indicators can take diverse forms, including genetic mutations, protein expression levels, metabolite alterations, and imaging features. In the context of rare diseases, reliable biomarkers are essential for early diagnosis, disease stratification, patient selection, and monitoring therapeutic efficacy. For instance, mutations in the *HTT* gene, which encodes the huntingtin protein, are strongly associated with the onset and progression of Huntington's disease (HD). Moreover, emerging evidence highlights the utility of fluid biomarkers such as neurofilament light chain (NfL) and mutant huntingtin (mHTT) in the early detection and longitudinal monitoring of HD and other neurodegenerative disorders [3,4].

With continuous advancements in high-throughput sequencing technologies, mass spectrometry platforms, and medical imaging modalities, researchers now have access to vast amounts of multi-omics data—including genomic, transcriptomic, proteomic, and metabolomic profiles. These datasets provide a robust foundation for the systematic identification and validation of biomarkers. However, traditional analytical approaches often rely on univariate analysis and hypothesis-driven strategies, which are limited in their ability to uncover complex, multivariate relationships embedded within high-dimensional biological systems. In recent years, AI, particularly machine learning (ML) and deep learning techniques, has significantly enhanced the efficiency and accuracy of biomarker discovery. By integrating heterogeneous data from diverse sources—such as blood tests, cerebrospinal fluid analyses, and neuroimaging—AI algorithms enable the generation of individualized biomarker profiles. For instance, in the study of trisomy 18 syndrome, the Keith group applied a Characteristic Systemic Aberrant Expression (CSAX) approach, combining mathematical modeling with computational methods, to identify gene sets associated with dental development, immune function, and glucocorticoid metabolism. These findings were highly consistent with clinical observations, demonstrating AI's capacity to elucidate associations between gene expression abnormalities and phenotypic traits [5]. Similarly, in drug trials for amyotrophic lateral sclerosis (ALS), the Biosigner algorithm successfully identified sphingomyelin as a key metabolic biomarker correlated with disease progression [6]. In studies of West syndrome, the same algorithm identified serine and an unidentified metabolite, X363, as potential biomarkers [7].

Furthermore, AI has demonstrated significant potential in uncovering complex molecular regulatory networks. Carlier et al. integrated mechanistic modeling with data-driven approaches to identify three key parameters in the study of Neurofibromatosis Type 1 (NF1), among which the rate of cartilage formation (Pmc) was

established as a critical indicator distinguishing responders from non-responders to treatment. This finding provides a theoretical basis for optimizing therapeutic strategies involving bone morphogenetic protein (BMP) signaling [8]. In another example, the biomedical robot developed by DeAndrés-Galiana et al. showcased AI's capacity to integrate diverse data modalities, including genomics, epigenetics, and demographic information. Leveraging feature selection and ensemble learning techniques, this tool identified compact gene signatures with high predictive power. For instance, in chronic lymphocytic leukemia (CLL), a set of 13 key genes achieved a prediction accuracy of 93.25% for IGHV mutation status. In inclusion body myositis and polymyositis (IBM-PM), the prominence of HLA family genes suggested a potential role of viral infection in disease pathogenesis [9].

Deep learning models have also proven instrumental in identifying biomarkers at both the imaging and molecular levels. Convolutional neural networks (CNNs) are widely used for analyzing medical images and detecting visual biomarkers such as patterns of brain atrophy. Graph neural networks (GNNs), on the other hand, enable the reconstruction of gene regulatory networks, revealing dynamic alterations in signaling pathways associated with rare diseases. In Gaucher Disease research, an integrative approach combining molecular dynamics simulations with convolutional variational autoencoder (VAE) models enabled the first detailed mapping of the interaction interface between glucocerebrosidase-1 (GCase) and its chaperone protein saposin C. This work elucidated how pathogenic mutations destabilize this critical complex, thereby triggering disease onset [10].

2.2 Virtual Screening and Drug Repurposing

Identification of suitable drug targets and the subsequent screening of potential active compounds represent some of the most time-consuming and costly stages in new drug development. In recent years, AI has significantly enhanced the efficiency of this process by enabling accurate prediction of molecular interactions and compound-target binding affinities. Among the most promising AI-driven strategies are virtual screening and drug repurposing, both of which have demonstrated considerable potential in accelerating therapeutic discovery for rare diseases.

I. Virtual Screening: Efficient Identification of Potential Active Molecules

Virtual screening is a computational technique that enables the rapid identification of bioactive molecules from large chemical libraries containing millions of compounds. Traditional high-throughput screening methods rely on expensive and labor-intensive experimental assays, whereas AI-powered virtual screening leverages algorithms such as deep learning, generative adversarial networks (GANs), and variational autoencoders (VAEs) to markedly improve screening efficiency.

For example, Liu et al. developed MolFilterGAN, a novel molecular filtering approach based on progressive-enhancement GANs. This model expands the drug-like chemical space through a stepwise enhancement strategy, effectively distinguishing biologically active or drug-like molecules from randomly generated ones. As a result, the hit rate was significantly improved, and the workload for manual evaluation by medicinal chemists was substantially reduced [11]. Moreover, AI facilitates the identification of disease-associated genes and potential drug targets. In a 2022 study by Liu's team on dermatofibrosarcoma protuberans (DFSP), an integrative approach combining text mining (PubMed2Ensembl), GO/KEGG pathway enrichment analysis, and protein–protein interaction (PPI) network construction identified 18 core disease-related genes. Using the deep learning tool DeepPurpose, the researchers predicted drug–target interactions and ultimately selected 10 high-affinity candidate drugs [12]. This work exemplifies the power of AI in integrating multi-source data for precise target identification and efficient lead compound discovery.

II. Drug Repurposing: New Indications for Existing Therapies

Drug repurposing—also known as drug repositioning—involves identifying novel therapeutic applications for already approved drugs, thereby bypassing the lengthy and costly phases of preclinical development and early-stage clinical trials. AI accelerates this process by systematically analyzing existing pharmacological

data, including drug indications, adverse effects, molecular mechanisms, and clinical outcomes, to rapidly identify candidates with potential utility in rare diseases. A representative example is TxGNN, a model developed by researchers at Harvard Medical School [13]. Based on zero-shot learning, TxGNN was trained on a knowledge graph encompassing 17,080 diseases and 7,957 drug mechanisms of action. It successfully predicted potential therapeutic agents for multiple rare conditions. For instance, the model suggested deferasirox—a clinically approved iron chelator—for the treatment of Wilson's disease, based on its capacity to reduce hepatic copper accumulation. This prediction aligns with off-label usage patterns in the United States, demonstrating the practical value of AI-driven drug repurposing.

III. AI-Driven Development of RNA-Based Therapeutics

With the rapid advancement of RNA-targeted therapies, AI has become an essential tool in the design and optimization of RNA-based therapeutics. These include oligonucleotide drugs such as small interfering RNA (siRNA) and antisense oligonucleotides (ASOs), messenger RNA (mRNA) vaccines, and small molecules targeting RNA structures. In the field of oligonucleotide therapy, CAMP4 Therapeutics employs its proprietary RAP AI platform to identify regulatory RNA elements (regRNAs) and develop ASO-based therapeutics that modulate gene expression. For example, their investigational compound CMP-CPS-1 targets CPS1 (carbamoyl phosphate synthetase 1) mRNA to upregulate enzyme expression, offering a potential treatment for urea cycle disorders [14]. For small-molecule drugs targeting RNA, Atomic AI has developed ARES (Atomic Rotationally Equivariant Scorer), a deep learning model capable of accurately predicting RNA three-dimensional structures—an essential step in structure-based drug design [15]. Other companies, including Anima Biotech, Rgenta, and XiLi Technologies, are also making significant progress in AI-assisted discovery of RNA-targeting small molecules. In the domain of mRNA therapeutics, AI tools have been applied to optimize sequence design, enhance translational efficiency, and reduce immunogenicity. Platforms such as LinearDesign and Smart5UTR have demonstrated success in mRNA vaccine development [16,17], and hold promise for future application in protein replacement therapies for rare diseases.

3. Advances in AI Applications for Rare Disease Diagnosis

Rare diseases, characterized by low prevalence, complex phenotypes, and diverse etiologies, have long posed significant diagnostic challenges, often leading to delayed or missed diagnoses. In recent years, rapid advancements in AI have revolutionized the landscape of rare disease diagnosis, particularly through its applications in genomics and phenotypic analysis.

In the field of genomics, AI has significantly improved the efficiency of identifying pathogenic variants from next-generation sequencing (NGS) data. Traditional approaches rely heavily on manual curation of vast variant datasets—a process that is both time-consuming and prone to error. AI algorithms, including Bayesian networks, random forests, and deep neural networks, are now widely applied to variant annotation, pathogenicity scoring, and inheritance pattern recognition.

Databases such as ClinVar, CADD, and REVEL, when combined with AI models, have enhanced the ability to distinguish benign from pathogenic variants. For instance, CliniPred, introduced in 2018, employs machine learning algorithms integrated with allele frequency data from the gnomAD database to achieve high-accuracy prediction of the pathogenicity of non-synonymous SNVs, outperforming existing tools with superior AUC scores [18]. Subsequently, VarCoPP was developed to specifically assess the pathogenic potential of bi-allelic or two-hit variant combinations, thereby advancing the understanding of complex genetic mechanisms [19]. In 2021, Fabric GEM demonstrated how AI can integrate phenotypic information extracted via natural language processing (NLP) with genomic data to substantially narrow down candidate gene lists, accelerating the diagnostic workflow [20]. By 2024, AI-MARRVEL (AIM), developed at Baylor College of Medicine, leveraged over 3.5 million genetic variants to identify diagnosable cases within undiagnosed disease networks (UDN), successfully uncovering two novel disease-associated genes [21]. Similarly, BGI's GeneT model utilized large language modeling

techniques to simulate expert reasoning, achieving a recall rate of 98–99% and improving diagnostic efficiency by up to 20-fold [22].

Beyond genomics, AI also plays a pivotal role in phenotype-driven diagnosis by enabling automated symptom recognition through image analysis and text mining. The Human Phenotype Ontology (HPO) provides a standardized vocabulary that facilitates precise matching of clinical features by AI systems. CNNs have been widely adopted for facial dysmorphology detection, enabling accurate identification of conditions such as Down syndrome and Noonan syndrome, with reported accuracy exceeding 90% [23]. Additionally, NLP techniques allow AI to extract unstructured clinical notes from electronic health records (EHRs) and map them to HPO terms, supporting diagnostic decision-making. In 2019, URSAHD was developed to train machine learning models on whole-genome expression profiles, successfully identifying several understudied disease-associated genes, particularly in ectodermal disorders [24]. In 2020, Erping et al. introduced an AI-assisted telemedicine platform that significantly improved early diagnosis and postoperative management of congenital cataracts, optimizing healthcare resource utilization [25]. By 2024, GestaltMML had integrated facial images, demographic data, and clinical notes into a multi-modal diagnostic framework based on Transformer architecture, markedly reducing the number of candidate genes for rare genetic disorders [26].

4. Development Stages and Performance of AI in Pharmaceutical Innovation

In recent years, AI has increasingly transformed the drug discovery and development pipeline. From early target identification to clinical trial design and regulatory support, AI technologies are reshaping traditional pharmaceutical workflows. Based on current research advances and industry practices, the evolution of AI in drug development can be categorized into several key stages, each demonstrating distinct advantages and promising potential.

I. Drug Discovery and Preclinical Research

During the early stages of drug discovery, AI is primarily applied to target identification and validation, chemical space exploration, and prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles. By integrating vast amounts of scientific literature, genomic data, and experimental results, AI enables efficient identification and prioritization of novel therapeutic targets, significantly accelerating the validation process. Moreover, generative AI techniques have proven highly effective in designing novel molecular structures, thereby expanding the chemical space of candidate compounds and increasing the likelihood of identifying viable lead molecules. AI models also play a crucial role in predicting ADMET properties, allowing for the early elimination of compounds with potential safety concerns and reducing the risk of late-stage failures. A notable example is Insilico Medicine, which reported that its AI platform completed the process from target discovery to optimized lead compound selection within approximately 18 months, at a cost of less than USD 2.7 million—roughly one-tenth of conventional R&D expenditures [27]. According to a report by Boston Consulting Group, between 2015 and the end of 2023, over 75 AI-discovered drug candidates entered clinical trials, with an exponential growth trend and a compound annual growth rate exceeding 60% [28]. These figures underscore AI's growing impact in drug discovery and optimization, suggesting its potential to substantially shorten preclinical development timelines and reduce costs, ultimately enabling more AI-assisted therapeutics to advance into clinical evaluation.

II. AI in Clinical Trial Design and Execution

(i) AI Integration in Phase I to Phase II Clinical Trials

In Phase I clinical trials, AI plays a key role in dose optimization, pharmacokinetic modeling, and prediction of patient responses. Reinforcement learning algorithms, for instance, enable dynamic dose adjustments to balance therapeutic efficacy with safety considerations. As of the end of 2023, 24 AI-discovered drug candidates had completed Phase I trials, with 21 demonstrating sufficient safety and tolerability to progress further—a success rate of 80–90% [28]. This significantly exceeds the industry average success rate of approximately 67%, largely due to AI's ability to accurately assess ADMET properties during preclinical development, thereby enhancing candidate drug safety profiles prior to human

testing.

(ii) AI Support in Phase II to Phase III Clinical Trials

During Phase II and III trials, AI contributes primarily to trial design optimization and improved patient recruitment efficiency. Adaptive trial design methodologies powered by AI allow for real-time adjustment of trial parameters, enhancing overall efficiency. Additionally, predictive modeling enables the selection of patient populations most likely to benefit from treatment, thereby increasing enrollment success rates. Due to the scarcity of data in the field of rare diseases, we have not yet seen any experiments showcasing AI's assistance in clinical recruitment for rare diseases. However, we have observed promising results in this area for other disease fields. For example, researchers at Mass General Brigham demonstrated that the use of a generative AI tool, RECTIFIER, nearly doubled the enrollment rate compared to traditional manual screening methods in heart failure clinical trials [29]. These findings suggest that AI-assisted recruitment strategies have the potential to reduce the time required to complete patient enrollment by up to 50%. We believe that similar achievements will soon be applied in the field of rare diseases as well.

However, the current success rate of AI-developed drug candidates in Phase II trials remains at approximately 40% (based on data from 10 candidates), only slightly higher than the industry-wide average of 39% [28]. According to Boston Consulting Group, AI has not yet demonstrated a clear advantage at this stage, which may be attributed to limitations such as small sample sizes, limited population representativeness, and coarse categorization methods in outcome analysis. To address these challenges and improve success rates in later-phase clinical development, several AI-driven clinical trial simulation and outcome prediction platforms have emerged globally, including QuantHealth and inClinico. These tools are designed to forecast clinical trial outcomes and optimize study designs, ultimately enhancing the translational potential of investigational therapies.

III. AI in Market Authorization and Regulatory Approval

AI is increasingly being applied during the drug marketing authorization phase, particularly in the integration of clinical trial data, real-world evidence (RWE), and the automation of regulatory documentation. Leveraging NLP and advanced data analytics, AI facilitates the preparation and submission of regulatory dossiers, thereby improving efficiency and reducing time-to-market. Both the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have begun exploring the role of AI in drug evaluation. For instance, the FDA has established the Digital Health Center of Excellence (DHCoE) to advance the integration of AI technologies into drug development and regulatory processes, with a focus on ensuring their safe, effective, and compliant use. Despite AI's demonstrated success in early-stage drug discovery, significant challenges remain in later clinical phases and post-marketing surveillance. While AI-developed candidates have shown high success rates in Phase I trials, their performance in Phase II and III trials—and ultimately in regulatory approval—remains under evaluation. Moreover, even after market authorization, drugs may face withdrawal risks due to unforeseen adverse events or suboptimal clinical efficacy. Therefore, the long-term validation of AI-driven pharmaceutical innovations remains essential to ensure their safety, efficacy, and broader clinical utility.

5. Key Advantages of AI-Driven Drug Discovery in Rare Disease Research

Rare diseases, characterized by low prevalence and limited patient populations, have long been associated with significant unmet medical needs due to the lack of effective therapeutic options. However, rapid advancements in AI have introduced transformative potential into drug discovery, offering new hope for the development of treatments tailored to rare disease indications. AI not only accelerates the research timeline and improves success rates, but also effectively addresses challenges such as data scarcity and high disease heterogeneity. Furthermore, it supports the advancement of personalized medicine and precision interventions. The following section outlines the major advantages of AI-driven pharmaceutical technologies in the context of rare disease research.

I. Acceleration of Early-Stage Drug Discovery and Improved Efficiency and Success Rates

Traditional drug discovery processes are both time-consuming and associated with high attrition rates, often requiring 4–6 years to progress from target identification to candidate selection. In contrast, AI-assisted platforms can perform similar tasks within 1–2 years by efficiently processing large-scale biological datasets, thereby significantly shortening the early development timeline. This advantage is particularly evident in RNA-based therapeutics. RNA drugs offer broad targetability, rapid development cycles, and durable therapeutic effects, making them highly suitable for addressing the urgent need for novel therapies in rare diseases. AI has demonstrated strong capabilities in RNA sequence analysis, structure prediction, and functional modeling, with multiple AI-powered tools already deployed in drug development pipelines, substantially enhancing the efficiency of RNA drug discovery.

Moreover, AI contributes to risk reduction in drug development through early toxicity prediction and clinical trial optimization. By simulating interactions between compounds and molecular targets, AI models can identify promising candidate molecules while filtering out those with potential toxic effects. This enables more informed decision-making in lead selection and increases the likelihood of success in later-stage clinical trials.

II. Addressing Challenges of Data Scarcity and High Heterogeneity

The limited number and wide geographical distribution of patients with rare diseases result in a significant scarcity of high-quality data samples available for model training. Furthermore, substantial variability in disease phenotypes and genetic backgrounds across individuals leads to high data heterogeneity, which poses a major challenge for model generalizability. To address these limitations, AI offers advanced methodological solutions such as transfer learning and federated learning. These approaches enable researchers to leverage abundant datasets from related common diseases and transfer the learned knowledge to rare disease research, thereby mitigating the impact of limited sample sizes. In addition, AI facilitates the integration of multi-source, heterogeneous data—including genomics, medical imaging, and EHRs—to construct more comprehensive disease models. This cross-modal data fusion enhances understanding of disease mechanisms and supports the identification of novel therapeutic targets, further accelerating drug discovery for rare conditions.

III. Enabling Personalized Treatment and Precision Interventions

AI enables the development of individualized treatment strategies by incorporating patient-specific genomic profiles, phenotypic features, and prior treatment histories. Through the analysis of large-scale patient data, AI models can predict inter-individual variations in drug response, guiding decisions on dosage optimization and combination therapies. This capability not only improves therapeutic outcomes but also minimizes adverse effects—factors that are particularly critical in the context of rare diseases, where high inter-patient variability often renders one-size-fits-all treatment approaches ineffective. Moreover, AI supports drug repurposing by constructing medical knowledge graphs that uncover potential associations between existing drugs and rare diseases. This strategy significantly shortens the time required for drug development and reduces costs, offering new therapeutic options for patients with rare conditions who currently have limited or no approved treatments.

IV. Optimization of Clinical Trial Design and Execution: Enhancing Efficiency and Feasibility

Clinical trials represent the most time-consuming and resource-intensive phase of drug development, particularly in the context of rare diseases. AI has demonstrated substantial potential across all stages of clinical trial execution, offering transformative solutions to long-standing challenges.

Trial Design Optimization: AI algorithms such as AI Hint and SPOT can predict the likelihood of trial success, enabling researchers to design more scientifically sound protocols. These models also support dynamic adjustments based on interim results, thereby increasing the probability of trial success [30,31].

Accelerated Patient Recruitment: NLP enables AI systems to extract and interpret patient information from EHRs, facilitating rapid identification and matching of eligible participants. For instance, the U.S.-developed TrialGPT model achieves precise alignment between individual patient profiles and trial eligibility criteria, significantly shortening

recruitment timelines [32].

Enhanced Data Analysis: AI supports efficient and accurate interpretation of large-scale clinical data by extracting relevant features and patterns, thereby improving statistical analysis and outcome evaluation.

A notable innovation is the digital twin model developed by UnlearnAI, which has been approved for use in clinical statistical analysis. This model generates in silico control arms by simulating disease progression, potentially reducing the required sample size by up to 20% [33]. Such an approach holds particular significance for rare disease trials, where patient recruitment is inherently challenging. It not only reduces time and resource expenditure but also enhances the ethical standards of clinical research by minimizing the burden on real-world patients. In line with this trend, China's Center for Drug Evaluation (CDE), under the National Medical Products Administration (NMPA), released the Guidance on the Use of Natural History Studies in Rare Disease Drug Development in July 2023. This document underscores the importance of leveraging natural history data to support drug development. AI-driven digital twin models align well with this regulatory direction, enabling comprehensive utilization of longitudinal disease data and providing robust support for clinical trial design in rare diseases.

V. Policy Support and Industry Collaboration Driving Innovation

In recent years, governments and regulatory agencies worldwide have increasingly recognized the value of AI in addressing unmet medical needs in rare diseases. A range of supportive policies and initiatives have been introduced to facilitate its integration into pharmaceutical innovation. For example, China has explicitly emphasized the strategic importance of integrating AI with biopharmaceutical research, encouraging the development of novel therapeutics through national innovation programs. Concurrently, growing collaboration between technology firms and pharmaceutical companies has accelerated the implementation of AI-powered approaches in rare disease drug discovery. This cross-sectoral cooperation model fosters synergies between computational innovation and biomedical expertise, paving the way for faster translation of AI-based discoveries into clinical applications. Such collaborative efforts are expected to drive sustainable progress in the field, creating a virtuous cycle of technological advancement and therapeutic development.

6. Case Studies of Leading AI-Driven Pharmaceutical Companies and Projects

In recent years, the rapid advancement of AI has significantly transformed the pharmaceutical landscape, particularly in the development of novel therapeutics for rare and complex diseases. The following four case studies highlight representative AI-driven companies and their successful drug discovery projects, illustrating how AI accelerates target identification, optimizes preclinical research, and expedites clinical development.

Case 1: AIT-101 [34]

AI Therapeutics' investigational drug candidate AIT-101 (LAM-002A) represents a novel therapeutic approach for amyotrophic lateral sclerosis (ALS). As a new class of potent and highly selective lipid kinase PIKfyve inhibitors, AIT-101 is capable of crossing the blood-brain barrier and acting directly on neurons within the central nervous system. Inhibition of PIKfyve leads to activation of the transcription factor TFEB, which promotes the clearance of toxic protein aggregates via the autophagy-lysosome pathway. In a Phase IIa clinical trial involving ALS patients carrying the C9ORF72 mutation, AIT-101 demonstrated promising efficacy. Over a 12-week treatment period, increased target engagement was observed through upregulation of pharmacodynamic biomarkers, accompanied by a 73% reduction in toxic protein aggregates. Moreover, the drug and its three active metabolites were successfully delivered to the brain and exhibited favorable safety and tolerability profiles. In addition, AIT-101 showed significant functional improvement in a TDP-43 mutant mouse model of ALS, suggesting potential therapeutic utility across multiple ALS subtypes. These findings provide strong preclinical and early clinical evidence supporting further development of AIT-101. This program highlights the potential of AI-driven drug discovery in addressing neurodegenerative diseases and offers new hope for patients with ALS.

Case 2: PXT3003 [35]

Pharnext has made significant strides in AI-powered drug development with PXT3003, a combination therapy for Charcot-Marie-Tooth disease type 1A (CMT1A). Demonstrating promising results in a Phase III trial, PXT3003 was granted FDA Fast Track designation in early 2025. The therapy consists of baclofen, naltrexone, and sorbitol. By leveraging AI to model CMT disease networks, Pharnext bypassed traditional de novo drug discovery and completed preclinical development in just three years—a marked reduction from the typical 8–10-year timeline. After screening 57 candidate drugs in silico, followed by experimental validation in vitro and in animal models, the optimal three-drug combination was selected. PXT3003 not only stabilizes disease progression but also promotes cellular regeneration, leading to measurable improvements in disability scores. Additionally, it has been granted priority review status in China, indicating its potential for accelerated global patient access.

Case 3: REC-4881 [36]

Recursion Pharmaceuticals utilized its AI-powered drug discovery platform, Recursion OS, to identify REC-4881—a non-ATP-competitive allosteric inhibitor of MEK1/2. Originally developed as a kinase inhibitor, REC-4881 was later found to exhibit therapeutic potential in familial adenomatous polyposis (FAP), a rare genetic disorder with limited treatment options. Through phenotypic screening in APC-knockdown cells, researchers discovered that REC-4881 suppresses aberrant MAPK signaling, thereby reducing polyp formation and delaying malignant transformation. The compound has received Orphan Drug and Fast Track designations from the FDA and initiated a Phase II trial for FAP in Q2 2025. REC-4881 exhibits favorable gut-targeted pharmacokinetics, making it a promising candidate for treating APC-driven gastrointestinal malignancies. This project exemplifies how Recursion's AI-driven discovery system—integrating high-throughput screening, machine learning, and phenotypic modeling—can accelerate the journey from unknown mechanisms to clinical validation, showcasing the transformative power of AI in modern drug development.

Conclusion

AI is rapidly transforming the landscape of rare disease research, diagnosis, and treatment, offering unprecedented depth and precision across multiple domains. From biomarker identification to RNA-based drug design, from virtual screening to clinical trial optimization, AI has demonstrated significant potential in enhancing both the efficiency of drug development and the accuracy of diagnostic approaches. Particularly in the field of rare diseases—historically hindered by data scarcity and high research costs—AI provides novel technical pathways through its advanced data-processing capabilities and sophisticated algorithmic models, enabling breakthroughs that were previously unattainable.

In pharmaceutical development, AI has shown substantial advantages at various stages. For de novo drug discovery, AI accelerates early-phase research through target identification, molecular design, and ADMET prediction, significantly shortening the overall development timeline. In the context of drug repurposing, AI enables rapid screening of potential indications, reducing the time required for target validation and preclinical design, and in some cases allowing candidates to bypass Phase I trials and proceed directly to clinical evaluation. Moreover, AI-powered digital twin technology has emerged as a promising tool in clinical trials, contributing to reduced sample size requirements, improved patient recruitment, and enhanced trial efficiency.

Nevertheless, the clinical value and therapeutic efficacy of AI-developed interventions remain under evaluation. While several AI-assisted compounds have demonstrated strong affinity and activity in preclinical studies, their performance in late-stage clinical trials, particularly in Phase III, requires further validation. Similarly, the effectiveness of drug repurposing strategies must be substantiated through real-world evidence and rigorous post-marketing surveillance.

Despite its promise, the widespread application of AI in rare disease research faces several challenges. First, the lack of high-quality, standardized datasets remains a critical barrier to model training and performance. Rare disease data are often fragmented, heterogeneous, and constrained by ethical and privacy considerations, making the establishment of shared, interoperable databases a foundational challenge. Second, the "black-box" nature of many AI models limits their

acceptance in regulatory decision-making. Therefore, advancing explainable AI (XAI) techniques to improve model transparency and interpretability will be essential for future regulatory compliance.

Furthermore, interdisciplinary integration and collaborative innovation are expected to drive deeper AI adoption in this domain. Synergies between AI and emerging technologies such as synthetic biology, single-cell sequencing, and CRISPR-based gene editing will expand its applicability and support the realization of precision medicine in rare disease management. Although AI currently sees broader application in non-rare disease areas—largely due to more abundant market incentives, funding, and data resources—the pace of AI implementation in rare disease research is accelerating, supported by growing policy incentives and technological advancements. Emerging platforms such as BioMap's xTrimo, XtalPi's ID4Inno™, Unlearn AI's digital twin systems, brain-computer interfaces, AI-driven mobile health solutions, and robotic surgery technologies exemplify the expanding frontiers of AI in healthcare. These innovations not only offer renewed hope for patients with rare diseases but also inject transformative momentum into the broader healthcare ecosystem.

Looking ahead, AI holds the potential to establish a patient-centered ecosystem for rare disease diagnosis and therapy, enabling efficient translation from bench to bedside. By integrating multi-source, heterogeneous data, streamlining drug development pipelines, improving diagnostic accuracy, and facilitating personalized treatment strategies, AI can help address the longstanding challenge of no available therapies in rare diseases. However, achieving this vision will require sustained collaboration among academia, industry, and regulatory bodies to collectively address technical, ethical, and legal challenges. Only through coordinated efforts can AI fulfill its role as a key driver of progress in rare disease therapeutics, ultimately delivering life-changing benefits to millions of patients worldwide.

Author Contributions

W.F., K.Z. and B.W. were responsible for conceptualization, drafting the original manuscript, review, editing, and supervision. All authors have reviewed and approved the final version of the manuscript for publication.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Funding Statement: No funding.

Footnotes: Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

References:

1. Chen K., Yao L., Liu Z. Towards government-funded special biomedical research programs to combat rare diseases in China. Biosci Trends. 2015 Apr;9(2):88-90. doi: 10.5582/bst.2015.01048. .
2. Sun D., Gao W., Hu H., Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022 Jul;12 (7):3049-3062. doi:10.1016/j.apsb.2022.02.0 02.
3. Morena E., Romano C., Marconi M., Diamant S., Buscarinu MC., Bellucci G., Romano S., Scarabino D., Salvetti M., Ristori G. Peripheral Biomarkers in Manifest and Premanifest Huntington's Disease. Int J Mol Sci. 2023 Mar 23;24(7):6051. doi:10.3390/ ijms24076051.
4. Caron N.S., Banos R., Aly A.E., Xie Y., Ko S., Potluri N., Anderson C., Black HF., Anderson LM., Gordon B., et al. Cerebrospinal fluid mutant huntingtin is a biomarker for huntingtin lowering in the striatum of Huntington disease mice. Neurobiol Dis. 2022 May;166:105652. doi: 10.1016/j.nbd.2022.105652.
5. Noto K., Majidi S., Edlow AG., Wick HC., Bianchi DW., Slonim DK. CSAX: Characterizing Systematic Anomalies in eXpression Data. J Comput Biol. 2015 May; 22(5):402-13. doi: 10.1089/cmb.2014.0155.

6. Blasco H., Patin F., Descat A., Garçon G., Corcia P., Gelé P., Lenglet T., Bede P., Meininger V., Devos D., et al. A pharmaco-metabolomics approach in a clinical trial of ALS: Identification of predictive markers of progression. PLoS One. 2018 Jun 5;13(6):e0 198116. doi: 10.1371/journal.pone.0198116.
7. Baraniuk J.N. Cerebrospinal fluid metabolomics, lipidomics and serine pathway dysfunction in myalgic encephalomyelitis/ chronic fatigue syndroome (ME/CFS). Sci Rep. 2025;15(1):7381. Published 2025 Mar 3. doi:10.1038/s41598-025-91324-1.
8. Carlier A., Vasilevich A., Marechal M., de Boer J., Geris L. In silico clinical trials for pediatric orphan diseases. Sci Rep. 2018 Feb 6;8(1):2465. doi:10.1038/s41598-018-20737-y .
9. deAndrés-Galiana E.J., Fernández-Martínez J.L., Sonis S.T. Design of Biomedical Robots for Phenotype Prediction Problems. J Comput Biol. 2016 Aug; 23(8):678-92. doi:10.1089/ cmb.2016.0008.
10. Romero R., Ramanathan A., Yuen T., Bhowmik D., Mathew M., Munshi L.B., Javaid S., Bloch M., Lizneva D., Rahimova A., et al. Mechanism of glucocerebrosidase activation and dysfunction in Gaucher disease unraveled by molecular dynamics and deep learning. Proc Natl Acad Sci U S A. 2019 Mar 12;116(11):5086-5095. doi:10.1073/ pnas.1818411116.
11. Liu X., Zhang W., Tong X., Zhong F., Li Z., Xiong Z., Xiong J., Wu X., Fu Z., Tan X., et al. MolFilterGAN: a progressively augmented generative adversarial network for triaging AI-designed molecules. J Cheminform. 2023 Apr 8;15(1):42. doi: 10.1186/s13321-023-007 11-1.
12. Liu M., Yang F., Xu Y. Identification of Potential Drug Therapy for Dermatofi brosarcoma Protuberans with Bioinformatics and Deep Learning Technology. Curr Comput Aided Drug Des. 2022;18(5):393-405. doi: 10.2174/1573409918666220816112206.
13. Huang K., Chandak P., Wang Q., Havaldar S., Vaid A., Leskovec J., Nadkarni G.N., Glicksberg B.S., Gehlenborg N., Zitnik M. A foundation model for clinician-centered drug repurposing. Nat Med. 2024 Dec; 30(12): 36 01-3613. doi: 10.1038/s41591-024-03233-x.
14. Genetown. CAMP4 Therapeutics Announces Dosing of First Participant in Phase 1 Clinical Study of CMP-CPS-001 [Internet]. BioSpace; [cited 2024 Mar 21]. Available from: biospace.com/camp4-therapeutics-announces-dosing-of-first-participant-in-phase-1-clinical -study-of-cmp-cps-001.
15. Townshend R.J.L., Eismann S., Watkins A.M., Rangan R., Karelina M., Das R., Dror R.O. Geometric deep learning of RNA structure. Science. 2021 Aug 27;373(6558): 1047-1051. doi: 10.1126/science.abe5650. Erratum in: Science. 2023 Jan 27;379 (6630): eadg6616. doi: 10.1126/science.adg6616.
16. Zhang H., Zhang L., Lin A., Xu C., Li Z., Liu K., Liu B., Ma X., Zhao F., Jiang H., et al. Algorithm for optimized mRNA design improves stability and immunogenicity. Nature. 2023 Sep;621(7978):396-403. doi: 10.1038/s41586-023-06127-z.
17. Tang X., Huo M., Chen Y., Huang H., Qin S., Luo J., Qin Z., Jiang X., Liu Y., Duan X., et al. A novel deep generative model for mRNA vaccine development: Designing 5' UTRs with N1-methyl-pseudouridine modification. Acta Pharm Sin B. 2024 Apr; 14(4):1814-1826. doi: 10.1016/j.apsb.2023.11.003.
18. Alirezaie N., Kernohan K.D., Hartley T., Majewski J., Hocking T.D. ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. Am J Hum Genet. 2018 Oct 4;103(4):474-483. doi: 10.1016/j.ajhg.2018.08.005.
19. Papadimitriou S., Gazzo A., Versbraegen N., Nachtegael C., Aerts J., Moreau Y., Van Dooren S., Nowé A., Smits G., Lenaerts T. Predicting disease-causing variant combinations. Proc Natl Acad Sci U S A. 2019 Jun 11;116(24):11878-11887. doi: 10.10 73/pnas.1815601116.
20. De La Vega F.M., Chowdhury S., Moore B., Frise E., McCarthy J., Hernandez E.J., Wong T., James K., Guidugli L., Agrawal P.B., et al. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med. 2021 Oct 14; 13(1):153. doi: 10.1186/s13073-021-00965-0.
21. Mao D., Liu C., Wang L., Ai-Ouran R., Deisseroth C., Pasupuleti S., Kim S.Y., Li L., Rosenfeld J.A., Meng L., et al. AI-MARRVEL - A Knowledge-Driven AI

System for Diagnosing Mendelian Disorders. NEJM AI. 2024 May;1(5):10.1056/aioa 2300009. doi: 10.1056/aioa2300009.
22. Liang, L., Chen, Y., Wang, T., Jiang, D., Jin, J., Pang, Y., Na, Q., Liu, Q., Jiang, X., Dai, W., et al. Genetic Transformer: An Innovative Large Language Model Driven Approach for Rapid and Accurate Identification of Causative Variants in Rare Genetic Diseases [J]. MedRxiv, 2024 : 2024.07.18.24310666.
23. Sherif F.F., Tawfik N., Mousa D., Abdallah M.S., Cho Y.I. Automated Multi-Class Facial Syndrome Classification Using Transfer Learning Techniques. Bioengineering (Basel). 2024 Aug 13;11(8):827. doi: 10.3390 /bioengineering11080827.
24. Lee Y.S., Krishnan A., Oughtred R., Rust J., Chang C.S., Ryu J., Kristensen V.N., Dolinski K., Theesfeld C.L., Troyanskaya O.G. A Computational Framework for Genome-wide Characterization of the Human Disease Landscape. Cell Syst. 2019 Feb 27;8 (2):152-162.e6. doi: 10.1016/j.cels.2018. 12.010.
25. Long E., Chen J., Wu X., Liu Z., Wang L., Jiang J., Li W., Zhu Y., Chen C., Lin Z., et al. Artificial intelligence manages congenital cataract with individualized prediction and telehealth computing. NPJ Digit Med. 2020 Aug 28;3:112. doi: 10.1038/s41746-020-00 319-x.
26. Wu D., Yang J., Liu C., Hsieh TC., Marchi E., Blair J., Krawitz P., Weng C., Chung W., Lyon GJ., et al. GestaltMML: Enhancing Rare Genetic Disease Diagnosis through Multimodal Machine Learning Combining Facial Images and Clinical Texts. ArXiv [Preprint]. 2024 Apr 22:arXiv:2312.15320v2.
27. Ren F., Aliper A., Chen J., Zhao H., Rao S., Kuppe C., Ozerov I.V., Zhang M., Witte K., Kruse C., et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. Nat Biotechnol. 2025 Jan; 4 3(1):63-75. doi:10.1038/s41587-024-02143-0.
28. Kp Jayatunga M., Ayers M., Bruens L., Jayanth D., Meier C. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. Drug Discov Today. 2024 Jun;29(6):104009. doi: 10.1016/ j.drudis.2024.104009.
29. Unlu O., Shin J., Mailly C.J., Oates M.F., Tucci M.R., Varugheese M., Wagholikar K., Wang F., Scirica B.M., Blood A.J., et al. Retrieval Augmented Generation Enabled Generative Pre-Trained Transformer 4 (GPT-4) Performance for Clinical Trial Screening. medRxiv [Preprint]. 2024 Feb 8:2024.02. 08.24302376. doi: 10.1101/2024.02.08.2430 2376.
30. Fu T., Huang K., Xiao C., Glass L.M., Sun J. HINT: Hierarchical interaction network for clinical-trial-outcome predictions. Patterns (N Y). 2022 Feb 3;3(4):100445. doi: 10.1016/j. patter.2022.100445.
31. Hutson M. How AI is being used to accelerate clinical trials. Nature. 2024 Mar; 627(8003):S2-S5. doi: 10.1038/d41586-024-00753-x.
32. Jin Q., Wang Z., Floudas C.S., Chen F., Gong C., Bracken-Clarke D., Xue E., Yang Y., Sun J., Lu Z. Matching patients to clinical trials with large language models. Nat Commun. 2024 Nov 18;15(1):9074. doi: 10.1038/s414 67-024-53081-z.
33. Vanderbeek A.M., Ross J.L., Miller D.P., Schuler A. Prognostic Covariate Adjustment for Binary Outcomes Using Stratification [Preprint]. arXiv:2212.09903; 2022.
34. AI Therapeutics. AI Therapeutics Announces Positive Results from Phase 2a Biomarker-driven Trial of AIT-101 in Patients with C9ORF72 Amyotrophic Lateral Sclerosis (ALS) [Internet]. Rochester, NY: Globe Newswire; 2023 Apr 5 [cited 2023 Oct 10]. Available from: https://www.rochesterfirst. com/business/press-releases/globenewswire/ 8802258/ai-therapeutics-announces-positive-results-from-phase-2a-biomarker-driven-trial-of-ait-101-in-patients-with-c9orf72-amyotrophic-lateral-sclerosis-als/.
35. Jennings M.J., Lochmüller A., Atalaia A., Horvath R. Targeted Therapies for Hereditary Peripheral Neuropathies: Systematic Review and Steps Towards a 'treatabolome'. J Neuromuscul Dis. 2021;8(3):383-400. doi: 10.3233/JND-200546.
36. Recursion Pharmaceuticals Inc. Evaluate REC-4881 in Patients With FAP (TUPELO) [Internet]. Bethesda (MD): National Library of Medicine (US); 2025 Apr 29 [cited 2025 Oct 15]. Available from: https://clinicaltrials. gov/study/NCT05552755