



Review

Generative Artificial Intelligence Transitions Pharmaceutical Development from Empirical Screening to Predictive Molecular Design and Clinical Trial Optimization

Ghaith K. Mansour¹ and Hatouf H. Sukkarieh^{2,*} ¹ College of Pharmacy, Alfaisal University, Riyadh 11533, Saudi Arabia; gkmansour@alfaisal.edu² Department of Pharmacology, College of Medicine, Alfaisal University, Riyadh 11533, Saudi Arabia

* Correspondence: hsukkarieh@alfaisal.edu

Abstract

The traditional paradigm of pharmaceutical research is characterized by substantial inefficiency, requiring extensive timelines and billions of dollars while suffering from high clinical attrition rates. The integration of generative artificial intelligence (AI) is driving a paradigm shift from empirical experimentation toward predictive, data-driven innovation. This review evaluates state-of-the-art applications of these technologies across the drug discovery and development pipeline. By analyzing multi-omics data streams, AI models can elucidate complex disease mechanisms and identify novel therapeutic targets. Deep generative architectures facilitate the algorithmic creation of novel molecular entities, enabling the design of therapeutics with complex polypharmacological profiles. Furthermore, AI is enhancing the clinical testing phase through large language models (LLMs) that improve patient enrollment and through synthetic control arms (SCAs) that provide computational alternatives to traditional placebo groups. Despite these advances, the scientific community must address inherent algorithmic biases stemming from demographic underrepresentation and mitigate the risks of data hallucinations. Ultimately, realizing the full translational potential of generative AI in precision medicine may require the widespread adoption of explainable AI (XAI) frameworks and rigorous data standards.

Keywords: generative artificial intelligence; drug discovery; de novo molecular design; clinical trial optimization; precision medicine; deep learning; pharmacogenomics; synthetic control arms



Academic Editor: Jinxiang Xi

Received: 5 March 2026

Revised: 3 April 2026

Accepted: 10 April 2026

Published: 13 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The traditional paradigm of pharmaceutical research and development is characterized by high degrees of inefficiency, substantial capital expenditures, and significant clinical attrition rates. Historically, the journey of a novel therapeutic from initial target identification to regulatory approval spans 10 to 15 years, with overall research and development expenses frequently exceeding \$2 billion per approved drug [1]. Furthermore, fewer than 10% of therapies entering Phase I clinical trials ultimately achieve market authorization [1,2]. This systemic inefficiency stems in part from the immense, multidimensional search space inherent to computational chemistry. The theoretical “chemical universe” is estimated to contain up to 10^{60} drug-like small molecules, presenting a combinatorial challenge that traditional high-throughput screening and heuristic-based medicinal chemistry cannot efficiently navigate [3,4].

The integration of AI, and more specifically deep generative modeling, is driving a paradigm shift in pharmaceutical sciences. Rather than relying solely on empirical trial-and-error experimentation, AI pipelines are increasingly influencing multiple stages of the pharmaceutical value chain [5,6]. Deep learning, machine learning, and natural language processing (NLP) models are now capable of processing massive multi-omics datasets, generating novel molecular entities *de novo*, predicting complex protein-ligand binding affinities, and optimizing clinical trial designs with improved precision [1,7]. This transformation may not only accelerate the speed of discovery but also alter the nature of the molecules being discovered, potentially enabling the design of therapeutics tailored for complex polypharmacological profiles and specific patient subpopulations [6,8].

To integrate these advances within a unified analytical structure, this review employs the Generative AI Continuum as an organizing framework, a systems-level conception of the pharmaceutical pipeline in which multi-omics target identification, generative molecular design, ADMET and binding affinity prediction, biomarker discovery, and clinical trial optimization are treated not as isolated methodological domains but as sequentially interdependent stages of a single data-driven value chain. This framework is developed across Sections 2–6 and synthesized in Section Toward a Self-Improving Pharmaceutical Pipeline, where its translational implications and research priorities are articulated.

This review evaluates state-of-the-art applications of generative AI across the drug discovery and development pipeline. It begins by examining the role of AI in elucidating disease mechanisms and identifying novel therapeutic targets through multi-omics integration. The narrative then transitions into a mechanistic analysis of generative architectures—including diffusion models, variational autoencoders (VAEs), and generative adversarial networks (GANs)—currently utilized for *de novo* molecular design, three-dimensional (3D) conformation generation, and lead optimization. The analysis subsequently explores how AI is influencing biomarker discovery and clinical trial optimization through the deployment of LLMs and SCAs. Furthermore, this review assesses the inherent limitations, algorithmic biases, and infrastructural challenges that must be addressed to realize the translational potential of AI in precision medicine. Finally, this review provides an original analytical contribution beyond descriptive summarization, this review proposes a unifying perspective: the “Generative AI Continuum.” Rather than treating target identification, molecular design, and clinical trial optimization as isolated computational subfields, we argue that generative AI is enabling their convergence into a single, iteratively self-improving epistemic cycle. In this cycle, multi-omics knowledge graphs condition molecular generation; generated molecular candidates inform synthetic biomarker hypotheses; and synthetic control arm architectures feed real-world pharmacological signal back into target validation models. This bidirectional information flow from biology to chemistry to clinical outcome and back represents a qualitative departure from the unidirectional empirical pipeline and constitutes the central thesis of this review. Where prior reviews have catalogued AI applications within pipeline stages, we seek to critically examine the evidence for and against this integrative vision, identifying where the Continuum is already operational, where it remains aspirational, and what methodological, regulatory, and equity barriers must be resolved before it can be responsibly deployed at scale.

2. AI-Driven Target Identification and Disease Mechanism Elucidation

The foundational step of modern therapeutic discovery is the precise identification and rigorous validation of molecular targets whose modulation yields a predictable and safe therapeutic effect. Historically, target discovery relied heavily on experimental comparative profiling methods, such as stable isotope labeling by amino acids in cell culture

(SILAC; Thermo Fisher Scientific, Waltham, MA, USA) and genome-wide association studies (GWAS) [9]. While these empirical methods have uncovered critical biological pathways for monogenic and structurally straightforward diseases, they are resource-intensive and often struggle to capture the non-linear, multidimensional interactions inherent in complex, multifactorial diseases such as idiopathic pulmonary fibrosis (IPF) and Alzheimer's disease (AD) [9,10]. AI methodologies have expanded the analytical toolkit by integrating heterogeneous multi-omics data streams—encompassing genomics, transcriptomics, proteomics, epigenomics, and metabolomics—to construct predictive systems-biology models of disease pathology [10,11].

2.1. Multi-Omics Integration and Deep Knowledge Graphs

Modern target identification leverages massive knowledge graphs and deep learning algorithms to synthesize structured biological data with unstructured data, such as scientific literature and global patent repositories [10]. By embedding genes, proteins, diseases, and chemical compounds into a shared high-dimensional latent space, AI models can infer previously unrecognized relationships between specific molecular targets and phenotypic disease expressions. These models utilize advanced graph neural networks (GNNs) and NLP to identify “druggable” targets that exhibit high topological significance within disease-associated protein–protein interaction (PPI) networks [12,13].

This analytical capacity is particularly valuable in under-characterized and orphan disease states. For instance, in the domain of IPF, a devastating fibrotic disease with high mortality rates, AI-driven target discovery platforms have processed multi-omics datasets derived from patient tissues alongside millions of text documents from the scientific literature [1,14]. Through deep feature extraction and mechanism-of-action filtering, algorithms prioritized TRAF2- and NCK-interacting protein kinase (TNIK) as a promising, first-in-class target for pulmonary fibrosis [14,15]. While TNIK had been previously implicated in certain oncology indications via the Wnt/ β -catenin signaling pathway, its specific role in pulmonary fibrogenesis was identified through the platform's multi-omics inference and AI-driven analysis [1,15].

2.2. Clinical Validation of AI-Identified Targets

The capability of generative models to propose novel targets is being substantiated by clinical validation. The discovery of the TNIK target directly catalyzed the development of Rentosertib, a small molecule inhibitor designed de novo via generative chemistry engines [15]. The end-to-end timeline from project initiation and target discovery to pre-clinical candidate nomination reportedly required approximately eighteen months, at an estimated out-of-pocket cost of approximately \$150,000—although these figures are industry self-reported and should be interpreted with caution [1,14]. In 2024, Rentosertib became one of the first AI-designed drugs, modulating an AI-discovered target, to achieve proof-of-concept success in a Phase IIa randomized, double-blind, placebo-controlled clinical trial [15,16]. Patients receiving the optimal dose demonstrated a statistically significant improvement of +98.4 mL in improvements in forced vital capacity (FVC) versus placebo, providing initial clinical evidence supporting the computational hypothesis [15,16].

2.3. Expanding Target Discovery to Neurodegeneration

Beyond fibrosis, AI-driven target identification is being actively deployed against neurodegenerative conditions, notably AD and AD-related dementias (ADRD). The historically high clinical attrition rate in AD trials has been partially attributed to an incomplete understanding of the disease's underlying pathophysiology [12]. In response, research groups are leveraging machine learning to map the complex interactome of the aging brain [12,13]. Contemporary frameworks apply Bayesian algorithms and multi-omics inte-

gration to infer AD risk genes directly from GWAS loci, dynamically mapping these loci to human PPI networks to identify therapeutic targets [13]. By synthesizing evolving models of tauopathy, amyloidosis, and neuroinflammation, AI enables the stratification of disease progression into mechanistic pathways, potentially accelerating early drug discovery efforts in neurodegeneration [12].

3. Generative Artificial Intelligence in De Novo Molecular Design

Once a viable biological target is identified, the challenge transitions to discovering or designing a chemical entity that can bind to the target with high affinity, stringent selectivity, and an optimal safety profile. Traditional high-throughput screening involves physically or virtually screening millions of existing compounds from predefined chemical libraries [5]. However, this method is fundamentally limited by the specific chemical space contained within those pre-existing libraries, leaving the vast majority of the 10^{60} possible drug-like molecules unexplored [3,4]. Generative AI addresses this limitation by shifting the paradigm from screening existing compounds to the algorithmic creation of novel matter [5]. Rather than searching for a rare entity in a predefined dataset, generative models learn the fundamental topological and quantum grammar of chemistry, enabling them to generate optimized molecules entirely de novo [4].

3.1. Evolution of Deep Generative Architectures

The field of de novo molecular generation utilizes several distinct classes of deep learning architectures, each offering unique mathematical approaches to sampling the continuous chemical latent space. Table 1 summarizes the foundational mechanisms and advantages of these architectures.

Table 1. Summary of Deep Generative Architectures for De Novo Molecular Design.

Architecture Type	Mathematical Mechanism	Primary Representation	Strengths in Molecular Drug Design
Variational Autoencoders (VAEs)	Maps input data to a continuous probabilistic latent distribution via an encoder, from which new samples are decoded [4,5]	1D Strings (SMILES, SELFIES)	Continuous property optimization, multi-parameter conditioning, and smooth interpolation between known structures [4]
Generative Adversarial Networks (GANs)	Employs a zero-sum game between a generator creating molecules from noise and a discriminator distinguishing real from synthetic molecules [4]	1D Strings, 2D Graphs	Generates realistic structural distributions; useful for targeted library generation without explicit likelihood modeling [4]
Normalizing Flows	Learns exact, invertible transformations of simple probability distributions to model complex molecular datasets [3]	2D Graphs, 3D Point Clouds	Provides exact likelihood estimation, improving chemical validity of generated molecules [3]
Geometric Diffusion Models	Destroys data through a forward Markov chain of noise injection, then trains a network to reverse this process [4]	2D Graphs, 3D Point Clouds, Protein Sequences	State-of-the-art for generating precise 3D geometries and conditional ligand generation within protein pockets [3,8]

A critical comparative evaluation of these architectures reveals important performance tradeoffs that are often obscured in descriptive reviews. Geometric diffusion models—exemplified by DiffSBDD and TargetDiff—currently achieve state-of-the-art performance on structure-based drug design benchmarks such as the CrossDocked2020 dataset, reporting Vina docking scores averaging -7.5 kcal/mol and drug-likeness (QED) scores of approximately 0.48, outperforming autoregressive graph models and VAE-based approaches on both metrics [3,4,17,18]. However, diffusion models carry a substantial computational cost: generating a single optimized 3D ligand within a protein binding pocket can require 500–1000 reverse diffusion steps, making high-throughput virtual screening computationally prohibitive without hardware acceleration [17,19]. VAEs, by contrast, provide efficient continuous latent space interpolation amenable to multi-parameter property optimization, but suffer from posterior collapse in high-dimensional molecular spaces, resulting in chemically valid but structurally redundant outputs [20,21]. GANs offer fast sampling but are notoriously difficult to train stably in molecular generation tasks, exhibiting mode collapse in which the generator converges on a narrow subset of the accessible chemical space [22,23]. Normalizing flows provide exact likelihood estimation—a significant theoretical advantage for hit rate optimization—but scale poorly to large, flexible molecules due to the computational complexity of maintaining invertible transformations across hundreds of atoms [24,25]. These tradeoffs suggest that no single architecture is universally optimal; rather, rational pipeline design requires matching architecture class to the specific optimization objective, target class, and computational resource envelope available [4,19,23].

3.2. Molecular Representations: From Linear Strings to Geometric Graphs

The efficacy of any generative model is tightly coupled to how molecules are computationally represented as input and output. Early generative models relied predominantly on linear one-dimensional (1D) textual strings, most notably the Simplified Molecular Input Line Entry System (SMILES, Daylight Chemical Information Systems, Aliso Viejo, CA, USA) [3,8]. While efficient for computational processing, SMILES representations often suffer from structural fragility; a minor alteration of a single character in a sequence can result in topological invalidity, such as open structural rings or impossible atomic valences [3]. To address this vulnerability, the cheminformatics field developed Self-Referencing Embedded Strings (SELFIES, Open-source; developed by Alán Aspuru-Guzik lab, University of Toronto, Toronto, ON, Canada), which rely on a semantically constrained mathematical grammar guaranteeing that every generated string corresponds to a chemically valid molecule [3]. Subsequently, fragment-based representations such as SAFE, GroupSELFIES, and fragSMILES were introduced to better capture chemically rich features and stereochemical chirality [3].

Despite these improvements in linear text representations, molecular binding is fundamentally a 3D spatial and thermodynamic phenomenon. Consequently, state-of-the-art generative design has transitioned toward two-dimensional (2D) graph representations—where atoms function as nodes and chemical bonds as edges—and, increasingly, toward 3D point clouds, which encode precise spatial coordinates and atomic torsional angles [3,4]. Processing these complex geometric representations requires advanced architectures such as Equivariant Graph Neural Networks (EGNNs) [3,4]. These neural networks are mathematically designed to be roto-translationally equivariant, meaning the network's understanding of the molecule remains consistent regardless of how the 3D structure is rotated or translated in space [3,4].

3.3. Geometric Diffusion Models for 3D Conformation Generation

A critical bottleneck in virtual screening and structure-based drug design is predicting the diverse, energetically favorable 3D conformations a flexible 2D molecular graph can adopt. The biological activity and target affinity of a drug are dictated by its localized 3D conformations. To address this, researchers have developed geometric diffusion models, including frameworks such as GeoDiff (Tsinghua University, Beijing, China) [2,26,27]. These generative systems operate by establishing a Markov chain that reverses a diffusion process; they initialize with random noise distributions of atomic coordinates and progressively denoise them to generate stable molecular conformations [26,28].

By framing conformation generation as a thermodynamic diffusion process, these models learn a localized, data-driven energy function from existing conformation databases [26,29]. Because the likelihood of conformations must remain roto-translationally invariant to be physically accurate, they utilize specialized equivariant Markov kernels to preserve the geometric integrity of the generated structures throughout the denoising trajectory [26,30]. Empirical results suggest that this geometric approach can outperform traditional molecular dynamics simulations and earlier machine learning methods in estimating the multimodal distribution of conformations, particularly for large, flexible drug-like molecules [26,31,32].

3.4. Polypharmacology and Multi-Target Therapeutic Design

Historically, the pharmaceutical industry pursued a “one-drug, one-target” philosophy. However, complex systemic diseases such as metastatic oncology and progressive neurodegeneration frequently exhibit network redundancy, pathway compensation, and adaptive resistance mechanisms, rendering single-target therapies less effective over time [8,33,34]. AI is enabling a shift toward rational polypharmacology—the deliberate design of single molecules that interact with multiple specific therapeutic targets simultaneously to address compensatory disease pathways [8,33,35].

Deep generative models, driven by reinforcement learning and active learning paradigms, are being deployed to construct multi-target agents. Frameworks utilizing unified, multi-objective reward schemes condition generative neural networks to explore the intersection of distinct pharmacological latent spaces, seeking molecules that satisfy conflicting binding constraints [8,36,37]. For example, AI models condition chemical structure templates with integrated biological insights, such as patient transcriptomic and genomic profiles, to generate molecules exhibiting dual activity against challenging cancer-relevant targets, including PI3K and BRD4 [8,38,39].

Emerging efforts are beginning to extend these generative architectures toward non-classical therapeutic modalities, including proteolysis-targeting chimeras (PROTACs), macrocycles, and covalent binders, though the scarcity of such compounds in foundational training corpora means that generative coverage of these modality spaces remains nascent relative to conventional small-molecule design [3,4].

4. Lead Optimization and Binding Affinity Prediction

The algorithmic generation of a novel molecular scaffold represents only the preliminary phase of therapeutic design. The subsequent phase, lead optimization, requires iterative modification of the molecule to maximize target binding affinity, ensure target selectivity, and optimize absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles [1,18,40]. In traditional pipelines, this is a notably slow, capital-intensive cycle of chemical synthesis and rigorous in vitro assaying [5,19].

Physics-Informed Deep Learning in Affinity Prediction

To accelerate lead optimization, computational chemists have long relied on *in silico* simulation methods such as free energy perturbation (FEP) and molecular mechanics generalized Born surface area (MM-GBSA) calculations [19,41,42]. While accurate in calculating thermodynamic binding free energies, these simulation methods are computationally expensive, precluding their routine use in evaluating large libraries of generative outputs [19,41].

A notable example of the convergence between physics and AI is the Pairwise Binding Comparison Network (PBCNet), an architecture designed to predict the relative binding affinity among congeneric ligands [19]. This network utilizes a physics-informed graph attention mechanism that processes a pair of protein pocket-ligand complexes simultaneously, explicitly encoding the differential spatial and electrostatic interactions between closely related molecular analogs to achieve ranking ability that approaches FEP performance at substantially reduced computational cost [19,43].

A candid assessment of where AI genuinely outperforms physics-based methods and where it does not is essential for the field to set realistic expectations. Free energy perturbation (FEP) remains the gold standard for relative binding affinity prediction in congeneric series, achieving mean unsigned errors (MUEs) of approximately 0.8–1.0 kcal/mol in prospective benchmarking studies [19,41,42]. Machine learning-based binding affinity models such as PBCNet and Uni-Mol achieve MUEs of 1.0–1.3 kcal/mol on the same benchmark sets [19,44], representing a modest but acceptable accuracy loss in exchange for orders-of-magnitude improvement in throughput: FEP calculations for a single compound typically require 12–48 h on GPU clusters, while ML inference operates in milliseconds per compound [43,45]. Critically, however, the comparative advantage of AI narrows significantly when applied outside the training domain. Models trained on kinase-focused datasets demonstrate pronounced performance degradation when evaluated on GPCRs or allosteric sites, with MUE increases of 0.3–0.7 kcal/mol, revealing the fundamental limitation of data-driven approaches in low-data regimes [45,46]. Furthermore, physics-based methods retain an inherent mechanistic interpretability advantage: FEP produces thermodynamic cycle-decomposed energy terms that chemists can use to rationalize structural modifications, whereas neural network affinity predictions are often opaque to direct chemical reasoning [41,45]. These considerations suggest that hybrid architectures—in which ML models perform rapid triage of generative outputs and FEP is reserved for final lead validation—currently represent the most pragmatically rigorous approach to the lead optimization problem [19,43,45].

Similarly, deep learning models utilize advanced neural architectures—ranging from voxel grid-based 3D-Convolutional Neural Networks (3D-CNNs) to Graph Attention Networks employing radial atomic environment vectors—to predict protein-ligand binding affinities directly from structural coordinates [20,21,44]. To overcome labeled data scarcity in computational chemistry, self-supervised learning techniques are increasingly utilized, enabling predictive models to learn fundamental biophysical interactions from large unlabeled structural databases before fine-tuning on specific target affinities [21,44,47]. The integration of these predictive ADMET and affinity models into “closed-loop” automated synthesis laboratories represents an advancing frontier in autonomous drug discovery frameworks [8,22,48].

5. Biomarker Discovery and Multi-Modal Data Integration

The ultimate efficacy of a highly optimized therapeutic agent is dependent upon its administration to a biologically receptive and correctly diagnosed patient population. Thus, biomarker discovery—the identification of measurable indicators of pathogenic processes or

pharmacological responses—is as critical as the drug design phase itself [1,49,50]. Generative AI and advanced machine learning models are enhancing the identification of clinical biomarkers by integrating ultra-high-dimensional, multimodal data streams including genomics, electronic health records (EHRs), wearable sensor data, and medical imaging [10,51].

5.1. Multi-Omics Frameworks for Precision Diagnostics

AI algorithms excel at detecting subtle, non-linear correlations across disparate datasets that human analysts may not perceive. Several computational frameworks illustrate this capability. The MILTON (AstraZeneca, Cambridge, UK) framework demonstrates that augmenting traditional clinical diagnostic data with AI-selected proteomics biomarkers can improve predictive performance across various disease states [10,52]. The PRSmix (Broad Institute of MIT and Harvard, Cambridge, MA, USA) framework utilizes machine learning (elastic net regression) to dynamically aggregate and optimize polygenic risk scores across complex populations, capturing epistatic interactions to improve genomic risk biomarkers [10,53]. The EpiSign (EpiSign Inc., London, ON, Canada) framework deploys support vector machines (SVMs) to analyze methylation data, identifying “episignatures” associated with Mendelian disorders [10,54,55]. The CytoTRACE2 (Stanford University, Stanford, CA, USA) framework utilizes an interpretable deep learning approach based on Gene Set Binary Networks to isolate cellular gene signatures capable of predicting individual responses to chemotherapy and immune checkpoint inhibitors in oncology [10,56]. Table 2 summarizes these frameworks [10,56].

Table 2. AI Biomarker Discovery Frameworks for Precision Diagnostics.

Framework	Biological Modality	Algorithmic Mechanism and Clinical Output
MILTON	Proteomics combined with routine clinical markers	Augments traditional diagnostic data with AI-selected proteomics biomarkers to improve predictive performance across disease states [10]
PRSmix	Genomics (Polygenic Risk Scores)	Uses elastic net regression to aggregate and optimize polygenic risk scores, capturing epistatic interactions to improve genomic risk biomarkers [10]
EpiSign	Epigenomics (DNA Methylation)	Deploys SVMs to analyze methylation data, identifying episignatures associated with Mendelian disorders and generating methylation variant pathogenicity scores [10]
CytoTRACE2	Single-cell Transcriptomics	Uses interpretable deep learning based on Gene Set Binary Networks to predict responses to chemotherapy and immune checkpoint inhibitors [10]

5.2. Generative AI Agents as Virtual Laboratories

Beyond static predictive modeling, generative AI is evolving toward orchestrating the dynamic discovery process itself through advanced multi-agent systems. Recent studies have demonstrated the deployment of LLM-powered “Virtual Labs” in biological research [10]. Within these autonomous environments, independent AI agents assume specialized roles—such as computational biologist, immunologist, and principal investigator [10]. When tasked with developing novel nanobodies against emerging viral variants, these interacting agents collaboratively design computational pipelines, propose sequence modifications, execute in silico binding simulations, and finalize therapeutic candidate selections [10]. The output of this multi-agent process has been validated in wet-lab experiments, suggesting that AI can automate aspects of scientific hypothesis formulation and experimental protocol generation [10].

Furthermore, deep learning has contributed to the development of “deep aging clocks” within the longevity field [57]. Utilizing GANs and specialized LLMs, these clocks analyze chronological multi-omics data to estimate a patient’s biological age with high preci-

sion. These tools serve as biomarkers for tracking aging processes and may help identify “geroprotectors”—molecules capable of modulating biological aging pathways [57].

6. Clinical Trial Optimization and Synthetic Control Arms

Despite advances in preclinical discovery and molecular generation, the clinical trial phase remains one of the most rigorous, expensive, and failure-prone bottlenecks in pharmaceutical development [7]. Late-stage trials often fail not because the underlying drug mechanism is inactive, but due to suboptimal patient stratification, inadequate statistical powering, and flawed protocol design [2,58]. AI is addressing these challenges through advanced predictive modeling, LLM-based trial matching, and synthetic data architectures [7].

6.1. Large Language Models for Patient Screening and Enrollment

Patient recruitment is a pervasive challenge, often delaying critical trials by months or years. The stringent inclusion and exclusion criteria of modern precision oncology trials make identifying eligible participants arduous for human clinical coordinators. The advent of medical LLMs has introduced improved efficiency to this process. For instance, the TrialGPT (National Library of Medicine (NLM) and National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD, USA) system—an application built upon advanced LLMs—was deployed to match complex patient health records against intricate trial criteria [59]. In large-scale evaluations spanning more than 75,000 clinical trials, TrialGPT achieved a patient-criteria matching accuracy of 87.3% and a recall rate exceeding 90% [59]. By rapidly analyzing unstructured EHRs and cross-referencing them with trial protocols, AI systems reportedly reduced patient screening time by 42.6%, performing at parity with or exceeding the accuracy of human clinical coordinators [59].

6.2. Data-Driven Stratification and Adaptive Trial Design

AI is simultaneously optimizing the design of trials themselves. Frameworks such as Trial Pathfinder (Stanford University, Stanford, CA, USA, in collaboration with Genentech, South San Francisco, CA, USA) represent a movement toward data-driven trial design [60]. By analyzing real-world data (RWD) and simulating trial outcomes under varying hypothetical parameters, AI can optimize inclusion and exclusion criteria [60]. Empirical studies utilizing this framework have demonstrated that algorithmically loosening specific, historically arbitrary criteria can increase the number of eligible patients without compromising the statistical integrity or safety of the trial. Optimizing patient selection through this approach has been reported to reduce the overall hazard ratio by approximately 0.05, suggesting that AI-stratified cohorts may yield stronger clinical efficacy signals [60].

Within personalized oncology, multimodal deep learning models can fuse temporal clinical data with molecular markers and radiological imaging to predict tumor progression within active trial participants [60]. These tools enable adaptive trial designs, wherein patient treatment allocation can be dynamically adjusted based on continuous biomarker feedback, thereby potentially minimizing the administration of ineffective treatments to non-responsive cohorts.

6.3. Synthetic Control Arms

One of the most potentially disruptive applications of generative AI in clinical development is the generation of SCAs [2,27,61]. In traditionally structured randomized controlled trials (RCTs), a portion of the cohort receives a placebo or standard-of-care treatment to provide a statistical baseline [2]. However, in severe, rapidly progressing indications such as late-stage oncology, IPF, or severe neurodegeneration, allocating patients to a placebo arm poses ethical dilemmas and hinders trial recruitment [2].

Generative AI provides a computational alternative to traditional placebo groups. Utilizing historical clinical trial data, RWD, and clinico-genomic databases, generative neural networks (frequently relying on architectures such as Conditional Restricted Boltzmann Machines) ingest baseline patient data to create “digital twins” of enrolled experimental patients [61,62]. By analyzing a patient’s baseline multi-omics and longitudinal clinical profile, the generative model simulates the trajectory of that patient’s disease progression under placebo conditions [2,62].

These *in silico* simulated trajectories form the SCA [2]. Because the underlying generative models can map high-dimensional causal relationships, they may account for hidden confounding variables that affect simple historical data comparisons [2]. By substituting empirical placebo patients with synthetic data proxies, trial sponsors can construct single-arm trials wherein recruited patients receive the experimental therapeutic, while the AI provides the statistical benchmark required for regulatory efficacy evaluations [2,27]. Furthermore, data generalization techniques and privacy-preserving algorithms help ensure that these synthetic datasets maintain statistical utility while minimizing the risk of sensitive patient information disclosure [63]. The adoption of SCAs has the potential to reduce trial costs, accelerate timelines, and provide an ethical alternative to traditional placebo designs, though regulatory acceptance remains an evolving area [2,64].

The regulatory acceptance of synthetic control arms (SCAs) depends critically on their demonstrated ability to replicate control arm outcomes with quantifiable fidelity. Current validation studies report that generative SCA models achieve concordance correlations (CCC) of 0.72–0.89 with historical placebo outcomes in oncology indications when trained on sufficiently large clinico-genomic databases, but performance degrades substantially in rare disease contexts where historical data are sparse, with CCC falling to 0.51–0.64 [2,27,61,62]. A critical source of error in SCA generation is temporal confounding: historical control populations are subject to secular trends in standard-of-care management that may not be captured in the generative model’s training distribution, leading to systematic underestimation or overestimation of placebo response rates [2,65]. In the IPF context specifically, the introduction of pirfenidone and nintedanib as standard-of-care agents in 2014 fundamentally altered placebo arm decline trajectories, rendering pre-2014 historical controls unreliable for contemporary SCA construction without explicit temporal recalibration [66,67]. Furthermore, the Conditional Restricted Boltzmann Machine architectures most commonly used for SCA generation assume conditional independence between feature sets that is biologically implausible in multi-morbid patient populations, potentially introducing unmeasured confounding that would not survive classical propensity score sensitivity analysis [62,63]. These methodological limitations underscore why regulatory agencies currently require SCAs to be designated as supplementary rather than primary evidence [2,64], and why prospective hybrid trial designs—in which a small concurrent control arm supplements the synthetic arm—remain the most defensible near-term approach for regulatory submission [2,65].

7. Limitations, Ethical Considerations, and Algorithmic Bias

Despite the acceleration and efficiency that AI can provide to the pharmaceutical pipeline, the uncritical adoption of these complex technologies presents scientific, clinical, and ethical risks. AI models, particularly deep neural networks utilized in precision medicine, are fundamentally dependent upon the quality, breadth, and integrity of their underlying training data [66–68].

7.1. Data Bias and Demographic Underrepresentation

A primary limitation of AI in global healthcare is the potential perpetuation and algorithmic amplification of historical data biases [11,68]. Modern genomic databases and historical clinical trial repositories predominantly feature biological data derived from populations of European descent, while underrepresenting minority demographics, native populations, and female cohorts [68]. Consequently, AI models trained on these skewed datasets—whether utilized for target identification, de novo molecular design, or patient stratification—risk generating drug profiles that may fail to account for the pharmacogenomic variability of a diverse global population [68].

Furthermore, the subjective logic of developers in choosing proxy variables for training data can inadvertently result in algorithmic discrimination [58]. A landmark 2019 study highlighted a failure where an algorithmic model used widely to predict patients' future health risks incorrectly equated historical medical costs with actual healthcare needs [58]. Because minority populations historically incurred lower healthcare costs due to systemic access barriers and socioeconomic disparities, the AI erroneously assessed them as healthier than white patients with identical disease severity [58]. Correcting this algorithmic bias was estimated to increase the proportion of minority patients receiving necessary additional care from 17.7% to 46.5% [58].

The magnitude of demographic underrepresentation in training datasets for AI-driven drug discovery is quantifiable and substantial. Genome-wide association study (GWAS) repositories remain disproportionately European-ancestry derived: as of 2023, approximately 78% of GWAS participants in the GWAS Catalog were of European descent, despite European populations constituting fewer than 16% of the global population. In multi-omics target identification models trained on these datasets, pharmacogenomic variant frequencies for CYP2D6 poor metabolizer alleles—which govern the metabolism of approximately 25% of clinically used drugs—are systematically underestimated for East Asian and African ancestry populations, potentially producing target prioritization models that deprioritize mechanisms of particular relevance to these groups. Structural remediation requires more than augmenting dataset size; it requires the architectural incorporation of ancestry-conditioned embeddings in graph neural networks used for target identification, the use of transfer learning from population-specific genomic cohorts such as the Africa Wits-INDEPTH Partnership for Genomic Studies (AWI-Gen) and the Saudi Biobank, and the adoption of fairness constraints during model training that penalize performance disparities across demographic subgroups. The Saudi Biobank, with over 200,000 participants and extensive phenotypic data, represents a particularly valuable and underutilized resource for recalibrating AI-driven discovery models toward Middle Eastern and Arabian Peninsula populations, for whom existing pharmacogenomic models have demonstrated significant predictive gaps.

Beyond racial and ethnic representation, sex- and gender-specific considerations represent a critical and underaddressed dimension of algorithmic bias in generative AI for drug discovery. Pharmacogenomic dimorphisms between biological sexes are well established and clinically significant: CYP3A4 and P-glycoprotein (P-gp) activity differences of up to 40% between males and females influence the metabolism and bioavailability of a broad range of therapeutics [68]. Furthermore, diseases prominently featured in AI-driven discovery efforts—including IPF and Alzheimer's disease—exhibit well-documented sex-specific prevalence patterns, disease trajectories, and treatment responses [68]. Despite this, the majority of foundational genomic and transcriptomic databases used to train generative models reflect historical underrepresentation of female cohorts in clinical trials and biological repositories. Consequently, generative models trained on such data may implicitly encode male-normative pharmacological assumptions, producing molecular candidates

and dosing predictions that are suboptimally validated for female patients. Addressing this gap requires deliberate architectural choices: incorporating sex as a conditioning variable within generative latent spaces (i.e., sex-conditioned VAEs or diffusion models), mandating sex-disaggregated performance benchmarking during model validation, and ensuring that training datasets are curated to achieve demographic parity across biological sex. Regulatory agencies and journal editors are increasingly demanding such sex-stratified reporting as a prerequisite for translational credibility, and the generative AI field must align accordingly.

7.2. Data Silos, Hallucinations, and the Need for Explainable AI

The modern pharmaceutical industry is inherently proprietary, resulting in fragmented, siloed data ecosystems [66,68]. The lack of standardized, interoperable data across international healthcare institutions restricts the input necessary to train robust foundational models [27]. When LLMs and generative architectures are forced to extrapolate from incomplete or siloed data, they become prone to “hallucinations”—the generation of confidently presented but factually incorrect biological data or clinical outputs [69,70]. In the context of automated pharmacovigilance or clinical trial summary drafting, an AI hallucination could obscure severe toxicity signals [69,70].

The term “hallucination” as applied to LLMs and generative models in pharmaceutical contexts encompasses at least three mechanistically distinct failure modes that warrant differentiated mitigation strategies. The first is factual confabulation—the generation of syntactically plausible but factually incorrect biomedical assertions, such as fabricated clinical trial outcomes, non-existent gene–disease associations, or erroneous drug–drug interaction claims. A 2025 systematic review of 128 studies on ChatGPT (versions 3.5 and 4) in health care reported substantial variability in accuracy across medical domains, with more inconsistent performance in clinical pharmacy and pharmacology than in other areas [69]. A separate systematic review and meta-analysis of drug-counseling applications of ChatGPT-4 found a pooled accuracy of 86%, implying a non-trivial error rate of about 14% in medication-related answers, with marked heterogeneity between studies [68]. Additional evaluations of clinical pharmacy tasks show that ChatGPT performs well in some counseling scenarios but is notably weaker for prescription review, adverse drug reaction recognition, and causality assessment, where accuracy scores can fall near 4–6 on a 10-point expert scale [71]. Collectively, these findings indicate that factual errors in pharmacotherapy-related outputs remain common and are more pronounced in complex or advanced decision-making contexts, underscoring the need for expert oversight before clinical use [69–71]. The second failure mode is structural hallucination in molecular generation—the output of molecules with valid SMILES notation but physically impossible or synthetically inaccessible structural features, such as pentavalent carbons or strained ring systems with no viable synthetic route. Estimated rates of structurally hallucinated molecules vary from 3% to 25% across generative architectures, with VAEs and GANs exhibiting higher rates than SELFIES-constrained models and diffusion approaches. The third mode is contextual misattribution, in which a generative model correctly identifies a chemical scaffold or biological target but attributes incorrect mechanistic properties based on spurious co-occurrence patterns in training data, a particular hazard when biomedical literature mining is used as a data source given the well-documented prevalence of inconsistent nomenclature and retracted findings in published pharmacological literature. Robust mitigation requires architecture-specific responses: retrieval-augmented generation (RAG) frameworks with curated pharmacological knowledge bases for factual confabulation; SELFIES-constrained decoding layers and synthesizability scoring functions for

structural hallucination; and cross-referencing pipelines against curated ontologies such as ChEMBL (European Bioinformatics Institute (EMBL-EBI), Hinxton, UK) and UniProt (Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland; European Bioinformatics Institute (EMBL-EBI), Hinxton, UK; Protein Information Resource (PIR), Georgetown University, Washington, DC, USA) for contextual misattribution.

To mitigate these risks, the integration of generative AI into clinical workflows requires rigorous human-in-the-loop oversight. However, clinical researchers warn of emerging phenomena termed “oversight fatigue” and “cognitive drift,” whereby human operators gradually become desensitized to AI outputs, trusting algorithms implicitly and failing to scrutinize erroneous data [57].

Addressing these challenges may require the adoption of XAI frameworks and strict adherence to FAIR (Findable, Accessible, Interoperable, and Reusable) data principles across both academia and industry [6,57]. Regulatory agencies, including the U.S. Food and Drug Administration (FDA), are actively defining validation protocols for AI-driven diagnostic and therapeutic models [57]. These frameworks mandate transparency in decision-making, ensuring that algorithms do not operate as opaque “black boxes” when consequential clinical decisions are at stake [57].

Importantly, the FDA’s 2024–2025 draft guidances on AI/ML-enabled medical devices and drug development tools have introduced more specific expectations for the validation and transparency of AI models used in clinical and regulatory decision-making contexts. These guidances emphasize the need for predetermined change control plans, performance monitoring frameworks, and the use of standardized benchmarking datasets to demonstrate model robustness and generalizability prior to regulatory submission. In the domain of generative molecular design specifically, initiatives such as the AI-Driven Drug Discovery Benchmark (AIDDB) and the Open Generative Molecular Design Challenge are emerging as community-endorsed evaluation frameworks, providing standardized test sets against which generative model outputs can be objectively compared across institutions. Adoption of such benchmarking suites would provide a clearer and more reproducible pathway for regulatory acceptance of AI-generated drug candidates, and their incorporation into development workflows is increasingly expected by both academic reviewers and regulatory bodies.

7.3. Generalization Failure and Distribution Shift

One of the most practically consequential but underappreciated limitations of generative AI in drug discovery is distributional generalization failure—the systematic degradation of model performance when applied to chemical or biological spaces that are underrepresented in training data. This is not a marginal concern: retrospective analyses of generative molecular design campaigns have found that between 20% and 40% of AI-generated hit molecules that achieve high predicted binding affinity scores subsequently fail in experimental validation due to distribution shift between the *in silico* training environment and the physical assay conditions [3,4]. Sources of this shift include differences between crystallographic protein structures used for training and the dynamic conformational ensembles present in solution; the systematic overrepresentation of ATP-competitive kinase inhibitor scaffolds in publicly available training datasets; and the near-complete absence of covalent binders, macrocycles, and PROTACs from most foundational training corpora, limiting the generative models’ capacity to explore non-classical modality spaces. Addressing generalization failure requires not only larger and more chemically diverse training sets, but also principled uncertainty quantification—the ability of a model to recognize and flag when a query molecule lies outside its reliable predictive domain. Bayesian deep learning and conformal prediction approaches offer promising frameworks

for operationalizing predictive confidence intervals in molecular property models, but their routine integration into generative design workflows remains limited.

8. Conclusions

The integration of generative AI represents a notable shift in pharmaceutical research, contributing to the transition of the industry from empirical methodologies toward a predictive, data-driven innovation model. These technologies have demonstrated the capacity to identify novel targets such as TNIK in a fraction of the traditional timeline, generate optimized molecular entities de novo, and enhance clinical trials through medical LLMs and SCAs. However, the translational success of AI remains contingent upon overcoming systemic data biases, the risk of algorithmic hallucinations, and the need for greater model interpretability. The adoption of XAI and FAIR data principles, combined with rigorous human-in-the-loop oversight, will be essential to ensure that these computational tools provide safe, equitable, and transparent solutions for modern precision medicine. Future research should focus on prospective validation of AI-generated drug candidates in diverse patient populations, the development of standardized benchmarking frameworks for generative models, and the establishment of regulatory pathways that balance innovation with patient safety.

Toward a Self-Improving Pharmaceutical Pipeline

The Generative AI Continuum framework articulated in this review implies a specific set of priorities for the field. First, prospective validation of AI-generated drug candidates in demographically diverse populations—including Middle Eastern, African, and South Asian cohorts currently underrepresented in pharmacogenomic databases—must become a prerequisite rather than an aspiration for translational credibility. Second, federated learning architectures offer a technically viable pathway for cross-institutional data integration that preserves data privacy while enabling the training of generalizable models across siloed repositories; their systematic adoption across academic medical centers and pharmaceutical partners warrants coordinated investment. Third, the co-design of regulatory and algorithmic validation standards—bringing together AI developers, regulatory scientists, and patient advocates in the drafting of benchmarking frameworks—is essential to ensure that the speed of methodological innovation does not outpace the institutional capacity to evaluate it responsibly. Together, these priorities define the translational agenda that must accompany the technical development of the Generative AI Continuum if its clinical promise is to be realized equitably and safely.

Author Contributions: G.K.M.: Conceptualization, Literature Review, Writing—Original Draft. H.H.S.: Conceptualization, Supervision, Writing—Review and Editing, Project Administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Dermawan, D.; Alotaib, N. From Lab to Clinic: How Artificial Intelligence (AI) Is Reshaping Drug Discovery Timelines and Industry Outcomes. *Pharmaceuticals* **2025**, *18*, 981. [\[CrossRef\]](#)
2. Thorlund, K.; Dron, L.; Park, J.J.H.; Mills, E.J. Synthetic and External Controls in Clinical Trials—A Primer for Researchers. *Clin. Epidemiol.* **2020**, *12*, 457–467. [\[CrossRef\]](#)
3. Xie, W.; Wang, F.; Li, Y.; Lai, L.; Pei, J. Advances and Challenges in De Novo Drug Design Using Three-Dimensional Deep Generative Models. *J. Chem. Inf. Model.* **2022**, *62*, 2269–2279. [\[CrossRef\]](#)
4. Martinelli, D.D. Generative Machine Learning for De Novo Drug Discovery: A Systematic Review. *Comput. Biol. Med.* **2022**, *145*, 105403. [\[CrossRef\]](#)
5. Gangwal, A.; Ansari, A.; Ahmad, I.; Azad, A.K.; Kumarasamy, V.; Subramaniyan, V.; Wong, L.S. Generative Artificial Intelligence in Drug Discovery: Basic Framework, Recent Advances, Challenges, and Opportunities. *Front. Pharmacol.* **2024**, *15*, 1331062. [\[CrossRef\]](#)
6. Selvaraj, C.; Chandra, I.; Singh, S.K. Artificial Intelligence and Machine Learning Approaches for Drug Design: Challenges and Opportunities for the Pharmaceutical Industries. *Mol. Divers.* **2022**, *26*, 1893–1913. [\[CrossRef\]](#)
7. Harrer, S.; Shah, P.; Antony, B.; Hu, J. Artificial Intelligence for Clinical Trial Design. *Trends Pharmacol. Sci.* **2019**, *40*, 577–591. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Liu, K.; Chen, X.; Ren, Y.; Liu, C.; Lv, T.; Liu, Y.; Zhang, Y. Multi-Target-Based Polypharmacology Prediction (mTPP): An Approach Using Virtual Screening and Machine Learning for Multi-Target Drug Discovery. *Chem. Biol. Interact.* **2022**, *368*, 110239. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Pun, F.W.; Ozerov, I.V.; Zhavoronkov, A. AI-Powered Therapeutic Target Discovery. *Trends Pharmacol. Sci.* **2023**, *44*, 561–572. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Kale, M.; Wankhede, N.; Pawar, R.; Ballal, S.; Kumawat, R.; Goswami, M.; Khalid, M.; Taksande, B.; Upaganlawar, A.; Umekar, M.; et al. AI-Driven Innovations in Alzheimer’s Disease: Integrating Early Diagnosis, Personalized Treatment, and Prognostic Modelling. *Ageing Res. Rev.* **2024**, *101*, 102497. [\[CrossRef\]](#)
11. Mak, K.K.; Pichika, M.R. Artificial Intelligence in Drug Development: Present Status and Future Prospects. *Drug Discov. Today* **2019**, *24*, 773–780. [\[CrossRef\]](#)
12. Dey, A.; Chakraborty, M.; Maulik, U.; Bandyopadhyay, S. Network Based Approach for Drug Target Identification in Early Onset Parkinson’s Disease. *Sci. Rep.* **2025**, *15*, 10563. [\[CrossRef\]](#)
13. Xu, J.; Hou, Y.; Zhou, Y.; Bekris, L.M.; Pieper, A.A.; Cummings, J.; Leverenz, J.B.; Cheng, F. A Network-Based Deep Learning Framework Translates GWAS and Multi-Omics Findings to Pathobiology and Drug Repurposing for Alzheimer’s Disease. *Alzheimer’s Dement.* **2022**, *18*, e066647. [\[CrossRef\]](#)
14. Ren, F.; Ding, X.; Zheng, M.; Korzinkin, M.; Cai, X.; Zhu, W.; Mantsyzov, A.; Aliper, A.; Aladinskiy, V.; Cao, Z.; et al. AlphaFold Accelerates Artificial Intelligence Powered Drug Discovery: Efficient Discovery of a Novel CDK20 Small Molecule Inhibitor. *Chem. Sci.* **2023**, *14*, 1443–1452. [\[CrossRef\]](#) [\[PubMed\]](#)
15. 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**, *43*, 63–75. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Xu, Z.; Ren, F.; Wang, P.; Cao, J.; Tan, C.; Ma, D.; Zhao, L.; Dai, J.; Ding, Y.; Fang, H.; et al. A Generative AI-Discovered TNIK Inhibitor for Idiopathic Pulmonary Fibrosis: A Randomized Phase 2a Trial. *Nat. Med.* **2025**, *31*, 2602–2610. [\[CrossRef\]](#)
17. Schneuing, A.; Harris, C.; Du, Y.; Didi, K.; Jamsab, A.; Igashov, I.; Du, W.; Gomes, C.; Blundell, T.L.; Lio, P.; et al. Structure-Based Drug Design with Equivariant Diffusion Models. *Nat. Comput. Sci.* **2024**, *4*, 1157–1169. [\[CrossRef\]](#)
18. Igashov, I.; Stärk, H.; Vignac, C.; Schneuing, A.; Satorras, V.G.; Frossard, P.; Welling, M.; Bronstein, M.; Correia, B. Equivariant 3D-Conditional Diffusion Model for Molecular Linker Design. *Nat. Mach. Intell.* **2024**, *6*, 417–427. [\[CrossRef\]](#)
19. Guan, J.; Qian, W.W.; Peng, X.; Su, Y.; Peng, J.; Ma, J. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction. In Proceedings of the 11th International Conference on Learning Representations (ICLR 2023), Kigali, Rwanda, 1–5 May 2023.
20. Bhadwal, A.S.; Kumari, M.; Kumar, A. PCF-VAE: Posterior Collapse Free Variational Autoencoder for De Novo Drug Design. *Sci. Rep.* **2025**, *15*, 34152. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M.; et al. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Front. Pharmacol.* **2020**, *11*, 565644. [\[CrossRef\]](#)
22. de Cao, N.; Kipf, T. MolGAN: An Implicit Generative Model for Small Molecular Graphs. In Proceedings of the ICML 2018 Workshop on Theoretical Foundations and Applications of Deep Generative Models, Stockholm, Sweden, 14 July 2018.
23. Xie, Y.; Shi, C.; Zhou, H.; Yang, Y.; Zhang, W.; Yu, Y.; Li, L. MARS: Markov Molecular Sampling for Multi-Objective Drug Discovery. In Proceedings of the 9th International Conference on Learning Representations (ICLR 2021), Virtual, 3–7 May 2021.

24. Draxler, F.; Sorrenson, P.; Zimmermann, L.; Rousselot, A.; Köthe, U. Free-form Flows: Make Any Architecture a Normalizing Flow. *arXiv* **2023**, arXiv:2310.16624.
25. Köhler, J.; Klein, L.; Noé, F. Equivariant Flows: Exact Likelihood Generative Learning for Symmetric Densities. In Proceedings of the 37th International Conference on Machine Learning (ICML 2020), Virtual, 12–18 July 2020; pp. 5361–5370.
26. Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; Tang, J. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. In Proceedings of the 10th International Conference on Learning Representations (ICLR 2022), Virtual, 25–29 April 2022.
27. Farah, E.; Kenney, M.; Warkentin, M.T.; Cheung, W.Y.; Brenner, D.R. Examining External Control Arms in Oncology: A Scoping Review of Applications to Date. *Cancer Med.* **2024**, *13*, e7447. [\[CrossRef\]](#)
28. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
29. Axelrod, S.; Gomez-Bombarelli, R. GEOM, Energy-Annotated Molecular Conformations for Property Prediction and Molecular Generation. *Sci. Data* **2022**, *9*, 185. [\[CrossRef\]](#)
30. Satorras, V.G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. In Proceedings of the 38th International Conference on Machine Learning (ICML 2021), Virtual, 18–24 July 2021; pp. 9323–9332.
31. Ganea, O.; Pattanaik, L.; Coley, C.; Barzilay, R.; Jensen, K.F.; Green, W.H.; Jaakkola, T.S. GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 13757–13769.
32. Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. Torsional Diffusion for Molecular Conformer Generation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24240–24253.
33. Chaudhari, R.; Bhatt, S.; Jangid, K.; Sharma, P.; Pandey, R. AI-Driven Polypharmacology in Small-Molecule Drug Discovery. *Int. J. Mol. Sci.* **2025**, *26*, 6996. [\[CrossRef\]](#)
34. Bi, X.; Wang, Y.; Wang, J.; Liu, C. Machine Learning for Multi-Target Drug Discovery: Challenges and Opportunities in Systems Pharmacology. *Pharmaceutics* **2025**, *17*, 1186. [\[CrossRef\]](#)
35. Cichońska, A.; Ravikumar, B.; Rahman, R. AI for Targeted Polypharmacology: The Next Frontier in Drug Discovery. *Curr. Opin. Struct. Biol.* **2024**, *84*, 102771. [\[CrossRef\]](#)
36. Munson, B.P.; Chen, M.; Bogosian, A.; Kreisberg, J.F.; Licon, K.; Abagyan, R.; Kuenzi, B.M.; Ideker, T. De Novo Generation of Multi-Target Compounds Using Deep Generative Chemistry. *Nat. Commun.* **2024**, *15*, 3636. [\[CrossRef\]](#)
37. Tan, R.K.; Liu, Y.; Xie, L. Reinforcement Learning for Systems Pharmacology-Oriented and Personalized Drug Design. *Expert Opin. Drug Discov.* **2022**, *17*, 849–863. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Kang, S.I.; Shin, J.H.; Wu, B.M.; Choi, H.S. Deep Generative AI for Multi-Target Therapeutic Design: Toward Self-Improving Drug Discovery Framework. *Int. J. Mol. Sci.* **2025**, *26*, 11443. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Peng, S.; Yang, Y.; Li, W.; Lin, S.; Chu, S.; Xiong, T.; Ge, J.; Sheng, L.; Wang, J.; Xu, H. Discovery of Novel PI3K/BRD4 Dual Inhibitors for Esophageal Cancer: Rational Design, Optimization, and Senescence-Inducing Mechanisms. *J. Med. Chem.* **2025**, *68*, 23078–23102. [\[CrossRef\]](#)
40. Plowright, A.T.; Johnstone, C.; Kihlberg, J.; Pettersson, J.; Robb, G.; Thompson, R.A. Hypothesis Driven Drug Design: Improving Quality and Effectiveness of the Design-Make-Test-Analyse Cycle. *Drug Discov. Today* **2012**, *17*, 56–62. [\[CrossRef\]](#)
41. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M.K.; Greenwood, J.; et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703. [\[CrossRef\]](#)
42. Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A.S.J.S.; Mackey, M.; Michel, J. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 3120–3130. [\[CrossRef\]](#)
43. Yu, J.; Su, M.; Li, Z.; Chen, G.; Kong, X.; Hu, J.; Wang, D.; Cao, D.; Li, Y.; Huo, R.; et al. Computing the Relative Binding Affinity of Ligands Based on a Pairwise Binding Comparison Network. *Nat. Comput. Sci.* **2023**, *3*, 860–872. [\[CrossRef\]](#)
44. Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; Ke, G. Uni-Mol: A Universal 3D Molecular Representation Learning Framework. In Proceedings of the 11th International Conference on Learning Representations (ICLR 2023), Kigali, Rwanda, 1–5 May 2023. [\[CrossRef\]](#)
45. Imrie, F.; Hadfield, T.E.; Deane, C.M.; Morris, G.M. Narrowing the Gap Between Machine Learning Scoring Functions and Free Energy Perturbation Using Augmented Data. *Commun. Chem.* **2025**, *8*, 37. [\[CrossRef\]](#)
46. Harren, T.; Matter, H.; Hessler, G.; Rarey, M.; Schöning-Stierand, K. Interpretation of Structure-Activity Relationships in Real-World Drug Design Processes Using Explainable Artificial Intelligence. *J. Chem. Inf. Model.* **2022**, *62*, 447–462. [\[CrossRef\]](#)
47. Zeng, X.; Xiang, H.; Yu, L.; Wang, J.; Li, K.; Nussinov, R.; Cheng, F. Accurate Prediction of Molecular Properties and Drug Targets Using a Self-Supervised Image Representation Learning Framework. *Nat. Mach. Intell.* **2022**, *4*, 1004–1016. [\[CrossRef\]](#)

48. Schneider, P.; Walters, W.P.; Plowright, A.T.; Sieroka, N.; Listgarten, J.; Goodnow, R.A.; Fisher, J.; Jansen, J.M.; Duca, J.S.; Rush, T.S.; et al. Rethinking Drug Design in the Artificial Intelligence Era. *Nat. Rev. Drug Discov.* **2020**, *19*, 353–364. [\[CrossRef\]](#)
49. Hasin, Y.; Seldin, M.; Lusis, A. Multi-Omics Approaches to Disease. *Genome Biol.* **2017**, *18*, 83. [\[CrossRef\]](#)
50. Strimbu, K.; Tavel, J.A. What Are Biomarkers? *Curr. Opin. HIV AIDS* **2010**, *5*, 463–466. [\[CrossRef\]](#)
51. Topol, E.J. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nat. Med.* **2019**, *25*, 44–56. [\[CrossRef\]](#)
52. Garg, M.; Karpinski, M.; Matelska, D.; Middleton, L.; Burren, O.S.; Hu, F.; Wheeler, E.; Smith, K.R.; Fabre, M.A.; Mitchell, J.; et al. Disease Prediction with Multi-Omics and Biomarkers Empowers Case-Control Genetic Discoveries in the UK Biobank. *Nat. Genet.* **2024**, *56*, 1821–1831. [\[CrossRef\]](#)
53. Truong, B.; Hull, L.E.; Ruan, Y.; Huang, Q.Q.; Hornsby, W.; Martin, H.; van Heel, D.A.; Wang, Y.; Martin, A.R.; Lee, S.H.; et al. Integrative Polygenic Risk Score Improves the Prediction Accuracy of Complex Traits and Diseases. *Cell Genom.* **2024**, *4*, 100523. [\[CrossRef\]](#)
54. Sadikovic, B.; Levy, M.A.; Aref-Eshghi, E. Functional Annotation of Genomic Variation: DNA Methylation Episignatures in Neurodevelopmental Mendelian Disorders. *Hum. Mol. Genet.* **2020**, *29*, R27–R32. [\[CrossRef\]](#)
55. Sadikovic, B.; Levy, M.A.; Kerkhof, J.; Aref-Eshghi, E.; Schenkel, L.; Stuart, A.; McConkey, H.; Henneman, P.; Venema, A.; Schwartz, C.E.; et al. Clinical Epigenomics: Genome-Wide DNA Methylation Analysis for the Diagnosis of Mendelian Disorders. *Genet. Med.* **2021**, *23*, 1065–1074. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Kang, M.; Gulati, G.S.; Brown, E.L.; Qi, Z.; Avagyan, S.; Armenteros, J.J.A.; Gleyzer, R.; Zhang, W.; Steen, C.B.; D'Silva, J.P.; et al. Improved reconstruction of single-cell developmental potential with CytoTRACE 2. *Nat. Methods* **2025**, *22*, 2258–2263. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Muehlematter, U.J.; Daniore, P.; Vokinger, K.N. Approval of Artificial Intelligence and Machine Learning-Based Medical Devices in the USA and Europe (2015–20): A Comparative Analysis. *Lancet Digit. Health* **2021**, *3*, e195–e203. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science* **2019**, *366*, 447–453. [\[CrossRef\]](#)
59. 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**, *15*, 9074. [\[CrossRef\]](#)
60. Liu, R.; Rizzo, S.; Whipple, S.; Pal, N.; Lopez Pineda, A.; Lu, M.; Arnieri, B.; Lu, Y.; Capra, W.; Copping, R.; et al. Evaluating Eligibility Criteria of Oncology Trials Using Real-World Data and AI. *Nature* **2021**, *592*, 629–633. [\[CrossRef\]](#)
61. Tucker, A.; Wang, Z.; Rotalinti, Y.; Myles, P. Generating High-Fidelity Synthetic Patient Data for Assessing Machine Learning Healthcare Software. *NPJ Digit. Med.* **2020**, *3*, 147. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Fisher, C.K.; Smith, A.M.; Walsh, J.R. Machine Learning for Comprehensive Forecasting of Alzheimer's Disease Progression. *Sci. Rep.* **2019**, *9*, 13622. [\[CrossRef\]](#)
63. Beaulieu-Jones, B.K.; Wu, Z.S.; Williams, C.; Lee, R.; Bhavnani, S.P.; Byrd, J.B.; Greene, C.S. Privacy-Preserving Generative Deep Neural Networks Support Clinical Data Sharing. *Circ. Cardiovasc. Qual. Outcomes* **2019**, *12*, e005122. [\[CrossRef\]](#)
64. U.S. Food and Drug Administration. *Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products: Guidance for Industry*; FDA: Silver Spring, MD, USA, 2023. Available online: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products> (accessed on 1 April 2026).
65. Salas-Vega, S.; Mossialos, E. External Control Arms in Oncology: Current Use and Future Directions. *Ann. Oncol.* **2022**, *33*, 376–385. [\[CrossRef\]](#) [\[PubMed\]](#)
66. King, T.E., Jr.; Bradford, W.Z.; Castro-Bernardini, S.; Fagan, E.A.; Glaspole, I.; Glassberg, M.K.; Gorina, E.; Hopkins, P.M.; Kardatzke, D.; Lancaster, L.; et al. A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis (ASCEND). *N. Engl. J. Med.* **2014**, *370*, 2083–2092. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Richeldi, L.; du Bois, R.M.; Raghu, G.; Azuma, A.; Brown, K.K.; Costabel, U.; Cottin, V.; Flaherty, K.R.; Hansell, D.M.; Inoue, Y.; et al. Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis (INPULSIS). *N. Engl. J. Med.* **2014**, *370*, 2071–2082. [\[CrossRef\]](#)
68. Cirillo, D.; Catuara-Solarz, S.; Morey, C.; Guney, E.; Subirats, L.; Mellino, S.; Gigante, A.; Valencia, A.; Rementeria, M.J.; Chadha, A.S.; et al. Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare. *npj Digit. Med.* **2020**, *3*, 81. [\[CrossRef\]](#)
69. Beheshti, M.; Toubal, I.E.; Alaboud, K.; Almalaysha, M.; Ogundele, O.B.; Turabieh, H.; Abdalnabi, N.; Boren, S.A.; Scott, G.J.; Dahu, B.M. Evaluating the Reliability of ChatGPT for Health-Related Questions: A Systematic Review. *Informatics* **2025**, *12*, 9. [\[CrossRef\]](#)

70. Huang, X.; Estau, D.; Liu, X.; Yu, Y.; Qin, J.; Li, Z. Evaluating the Performance of ChatGPT in Clinical Pharmacy: A Comparative Study of ChatGPT and Clinical Pharmacists. *Br. J. Clin. Pharmacol.* **2023**, *90*, 232–238. [[CrossRef](#)] [[PubMed](#)]
71. Azmakan, H.; Nabipour, A.; Ghorabi Tehrani, N.; Najari, N.; Fathi Hafshjani, P.; Falahati Marvast, A.; Mani, S.; Asemi Sichani, N.; Fallah Pakdaman, S.; Shieh, M.; et al. ChatGPT as a Digital Pharmacist: A Systematic Review and Meta-Analysis of Drug-Counselling Accuracy. *medRxiv* **2025**. [[CrossRef](#)]

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.