



Contents lists available at [Journal of Pharmacological Horizons](https://www.lakshmi-biosciences.com)
Journal of Pharmacological Horizons
 Journal homepage: www.lakshmi-biosciences.com



REVIEW ARTICLE

Artificial Intelligence in Drug Discovery and Development: From Computational Promise to Clinical Translation – A Pharmacological Perspective

Sonali Arora^a, Sachin Sharma^b, Ravi Jayant^c*

^a Assistant Professor, Sunder Deep Pharmacy College, SDGI Global University, Ghaziabad, Uttar Pradesh, India

^b Assistant Professor, Department of Pharmaceutics, ITS College of Pharmacy, Muradnagar, Ghaziabad, Uttar Pradesh, India

^c Professor, Department of Pharmaceutics, NKBR College of Pharmacy and Research Centre, Meerut, Uttar Pradesh, India

ARTICLE HISTORY

OPEN ACCESS

EDITED BY

Lakshmi Bioscience Team;

REVIEWED BY

Dr. Sagarika Majhi;

Dr. Amit Lather;

Dr. Iti Chauhan;

*CORRESPONDENCE

Ravi Jayant, NKBR College of Pharmacy and Research Centre, Meerut, Uttar Pradesh, India

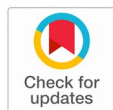
RECEIVED 1 May 2026

REVISED 1 June 2026

ACCEPTED 1 July 2026

PUBLISHED 30 August 2026

DOI: <https://doi.org/10.67796/jph.v1i2.16>



COPYRIGHT

© 2026 Arora S. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Abstract: The use of artificial intelligence (AI) now spans the drug discovery and development pipeline, from target identification to post-marketing surveillance, yet the evidence for its pharmacological and clinical impact is uneven between applications. This review synthesizes 2023–2026 evidence on AI methods, platforms, clinical translation, and regulation, and identifies where AI has produced measurable impact and where performance claims remain unvalidated. We conducted a structured narrative review of PubMed-indexed literature, Google Scholar, ScienceDirect, Nature Portfolio journals, and regulatory sources (FDA, EMA), limited to English-language peer-reviewed articles. Industry and trade-press reports were screened for context only and are not used as evidence. The review was not registered and reports no PRISMA flow diagram, record counts, or formal risk-of-bias appraisal. The most reproducible gains are in early discovery (structure prediction, virtual screening, and ADMET triage), where training labels are close to the pharmacological property of interest. A peer-reviewed asset-level analysis reports Phase I success of 80–90% for AI-nominated candidates against historical industry averages of roughly 40–65%, but Phase II success (about 40%, small sample) is comparable to historical averages. Larger industry counts use inconsistent definitions and are not used as evidence, and no AI-designed drug is approved. Rentosertib, an AI-discovered TNIK inhibitor, improved forced vital capacity over placebo in a small, 12-week randomized Phase IIa trial, and a Phase III trial was registered in 2026 and no efficacy data are available. The FDA draft guidance of January 2025 excludes discovery-stage uses and handles continuously updated models only through general lifecycle maintenance. AI has produced reproducible gains in specific early-pipeline tasks, but prospective clinical and regulatory validation remains the main unresolved constraint. We propose tiered explainability, predetermined change control plans, and continuous validation protocols as pathways for overseeing adaptive models, and a pharmacological evidence checklist for AI-originated candidates.

Keywords: Artificial intelligence; machine learning; generative AI; large language models; drug discovery; pharmacokinetics; pharmacodynamics; safety pharmacology; drug repurposing; pharmacovigilance; regulatory science.

1. INTRODUCTION

1.1 Background and Rationale

Conventional small-molecule drug development is one of the least efficient processes in applied science. Bringing one approved medicine to market is commonly estimated to take

10–15 years and, depending on the cost-of-capital assumptions used, one to several billion US dollars [1,2,3]. Clinical-stage attrition still exceeds 90% in most therapeutic areas [4]. Attrition is commonly driven by lack of efficacy, unacceptable safety, or unfavorable pharmacokinetics in humans [4,5]. These structural inefficiencies, rather than any

single technological gap, explain the rapid uptake of artificial intelligence (AI) and machine learning (ML) across the pipeline over the past five years. A large and fast-growing literature now describes AI applications in target identification, molecular design, preclinical safety assessment, clinical trial operations and post-marketing surveillance [6,7].

1.2 The Chemistry–Biology Evidentiary Divide

Bender and Cortés-Ciriano argued that enthusiasm for AI in drug discovery has often outpaced the evidence normally required in pharmaceutical science [8,9]. Their central distinction is between chemistry-facing tasks and biology-facing tasks. In chemistry-facing tasks, such as predicting whether a compound binds a target or estimating a physicochemical property, the training labels lie close to the outcome of interest. In biology-facing tasks, such as predicting efficacy and safety in a living organism, the labels are distant proxies for the clinical outcome. A related expert perspective on the grand challenges of AI in small-molecule drug discovery likewise noted that clear impact in discovery projects is still awaited [10].

In pharmacological terms, the first group corresponds to molecular pharmacology and pharmacokinetic properties. The second corresponds to pharmacodynamics, integrative physiology and clinical pharmacology. We use this divide as the organizing lens of the review. Structure prediction and virtual screening show the most reproducible gains, whereas target validation and clinical efficacy prediction remain provisional and contested.

1.3 Aim and Review Questions

This review addresses three questions. **RQ1:** Which AI methodologies (machine learning, deep learning, generative AI, large language models and reinforcement learning) are applied across the drug discovery and development pipeline, and how strong is the evidence for each? **RQ2:** What is the current state of clinical and regulatory translation for AI-originated therapeutics, and how does it compare with conventional development benchmarks? **RQ3:** Which structural challenges (data quality and equity, interpretability, synthesizability, regulatory readiness and environmental sustainability) limit real-world translation, and what near-term developments may address them?

1.4 Contribution and Positioning

Earlier reviews of AI in drug discovery, such as those of Paul et al. [6] and Vamathevan et al. [7], predate 2023. They therefore do not cover transformer-based foundation models, the 2025 FDA draft guidance, or peer-reviewed randomized

efficacy data for an AI-originated candidate. More recent scoping work, such as Shakeri and Far [11], examines generative AI, graph neural networks, reinforcement learning and proof-of-concept quantum computing applications in drug discovery. A 2025 critical review of small-molecule discovery likewise concludes that AI complements rather than replaces established methods and reports failures alongside successes [12]. This review is organized by pipeline-stage pharmacology and adds clinical-translation figures and regulatory analysis.

This review adds three elements. First, it organizes the evidence by pipeline stage and by the pharmacological question each stage addresses. Second, it synthesizes quantitative benchmarks while separating historical reference benchmarks from evidence inside the 2023–2026 window. Third, it proposes concrete pathways for regulating continuously updated AI systems. We do not claim novelty in the individual methods or studies discussed, which come from the existing literature. The contribution lies in the synthesis, the critical framing and the regulatory proposals.

2. METHODS

2.1 Review Design

This is a structured narrative review. It was not registered and was not conducted as a systematic or PRISMA-ScR scoping review. The design was chosen deliberately. The evidence base mixes peer-reviewed primary studies, peer-reviewed reviews, regulatory documents and, for clinical-pipeline status, trial-registry records and sponsor announcements, which are flagged where used. A scoping protocol restricted to peer-reviewed literature would exclude the regulatory and registry material. Consequently, the review reports no PRISMA flow diagram, no database record counts and no formal risk-of-bias appraisal. The consequences of this choice are discussed under Limitations.

2.2 Information Sources and Search Strategy

Literature was identified through iterative, topic-targeted searches of PubMed-indexed literature, Google Scholar, ScienceDirect, and Nature Portfolio journals. Primary regulatory sources (FDA guidance documents and EMA publications) were also searched. Preprints were used only as background context. Industry and trade-press reports were screened for context only; no quantitative claim in this review rests on them.

Searches combined core terms (“artificial intelligence”, “machine learning”, “generative AI”, “large language models”, “reinforcement learning”) with pipeline terms

(“target identification”, “virtual screening”, “ADMET”, “drug repurposing”, “clinical trial”, “pharmacovigilance”, “regulatory”) and domain qualifiers (“digital twin”, “nanomedicine”, “precision medicine”). Searches prioritized publications from January 2023 to August 2026, and regulatory and trial-registry status was rechecked in August 2026. Landmark earlier publications (for example, the original AlphaFold and ReLeaSE papers) were retained where they remain the primary methodological reference.

2.3 Eligibility Criteria

We included English-language, peer-reviewed original research and reviews on AI methods applied to any stage of drug discovery or development, primary regulatory guidance and, for clinical-pipeline status only, trial-registry records and sponsor announcements (flagged where used). We excluded preprints used as primary evidence, marketing material without methodological detail, and sources whose quantitative claims were unverifiable or internally inconsistent. Where such claims were important, they are noted as points of disagreement between sources.

Pre-2023 benchmark studies are included only as historical reference anchors. They are marked in the benchmark table and are not presented as evidence of the 2023–2026 state of the art.

2.4 Data Synthesis

Findings were synthesized narratively and organized by AI methodology and by pipeline stage. Where sources gave different quantitative estimates for the same quantity, such as the number of AI-originated drugs in clinical trials, both estimates are reported with their methodological basis instead of a single reconciled figure.

2.5 Use of Secondary Evidence

A recent scoping review by Shakeri and Far [11] covers AI in healthcare and drug discovery. We cite it only as a pointer to the wider literature. No count, percentage or figure from that review is used as evidence here. The maturity assessments in Section 4.1 and the challenges in the Discussion are the authors’ qualitative synthesis of the cited primary and review literature, and they are not quantified.

3. CONCEPTUAL BACKGROUND

This section gives the minimum technical background needed to read the Results, with attention to what each approach means for pharmacological questions.

3.1 AI Paradigms

Generative AI (GAI) covers models that learn a data distribution and generate new samples, such as molecular graphs or SMILES strings. They suit *de novo* design, where the aim is a novel compound with a desired pharmacological profile. Common families are generative adversarial networks [13], variational autoencoders (VAEs) [14,15,16], diffusion models [17] and transformer-based generators [18]. They are often combined with reinforcement learning (RL) to steer generation towards potency and property targets [19].

Large language models (LLMs) are transformer-based networks trained on large text corpora [18,20,21]. In drug discovery they are used mainly for literature mining, hypothesis generation and clinical text mining, and, through chemistry-tuned variants, for support in molecular design. Their outputs need the same verification as any biology-facing prediction.

Reinforcement learning (RL) frames learning as sequential decision-making that maximizes a cumulative reward [19]. In drug discovery it is used mainly to guide molecule generation [22,23]. In clinical settings it has been explored for treatment optimization and individualized dosing [24].

Quantum computing (QC) may help with molecular simulation and combinatorial optimization [25,26], but current noisy intermediate-scale quantum (NISQ) hardware limits it. Hybrid quantum–classical approaches are considered the most realistic near-term route [27]. Deep learning in general builds on representation learning [28].

3.2 Pharmacological Context

Drug discovery has grown from a chemistry-driven activity into a multidisciplinary field combining pharmacology, clinical science, molecular biology, genomics and bioinformatics. Widely cited estimates put the capitalized cost of a new medicine above US\$2.6 billion [1,3]. High-throughput screening (HTS) tests large chemical libraries against targets, but its cost and low hit rates limit access, particularly in academic settings. This is the practical rationale for computational methods.

3.3 Molecular Representations

AI models need molecules in machine-readable form, and the representation determines which methods can be used [29]. String formats such as SMILES [30] and the more robust SELFIES [31] suit sequence models and generative design. Two-dimensional graphs allow graph neural networks but omit stereochemistry. Three-dimensional representations capture geometry relevant to binding but need high-quality

structures. Fingerprints such as ECFP [32] and MACCS keys [33] remain common in QSAR because they are simple and interpretable. Table 1 summarizes these options.

Table 1. Molecular representation strategies and their relevance to AI-driven drug discovery.

Representation	Examples	Typically used by	Strengths	Limitations	Ref.
String	SMILES, SELFIES	RNNs, transformers, RL	Simple and scalable	SMILES invalidity; stereochemistry issues	[30,31]
2D graph	Atom–bond graph	GNNs, graph RL	Natural structure representation	Limited 3D geometry	[29]
3D graph	Coordinates, distances, angles	Equivariant GNNs, docking models	Captures molecular geometry	Computationally expensive	[29]
Fingerprints	ECFP, MACCS	QSAR, classical ML	Efficient and interpretable	Less expressive than graph methods	[32,33]
Descriptors	Physicochemical, topological, quantum	ML, screening	Chemically meaningful	Requires feature engineering	[29]

4. RESULTS

4.1 AI Methodologies and Their Evidentiary Maturity

Table 2 summarizes the main AI methodology families, their principal drug-discovery use and an evidence-based maturity assessment drawn from the reviewed literature.

Table 2. AI methodology families identified in the reviewed literature, with an evidentiary maturity assessment drawn from the reviewed literature.

AI methodology	Primary drug-discovery application	Representative methods/tools	Evidentiary maturity	Ref.
Classical machine learning (SVM, random forest, gradient boosting)	QSAR, ADMET prediction, toxicity screening	eToxPred, ProCTOR, TargeTox	Mature; well validated on retrospective benchmarks, limited on truly novel chemical space	[34]
Deep learning (CNN, GNN, MPNN)	Molecular property prediction, protein–ligand scoring	DeepTox, PotentialNet, GNN-based QSAR	Moderately mature; strong benchmark performance, variable external validation	[34,35,36]
Transformer-based structure prediction	Protein and complex structure prediction	AlphaFold2, AlphaFold3	High maturity for single-chain proteins; Nobel-recognized; lower reliability for large conformational change	[37,38,39]
Large language models	Biomedical text mining, repurposing, molecular design support	Clinical and chemistry-tuned LLMs; TxGNN (graph foundation model)	Growing maturity for knowledge synthesis; hallucination and validation gaps for molecular design	[20,21,40]
Generative AI (VAE, GAN, diffusion, autoregressive)	De novo molecular and protein design	REINVENT, JT-VAE, ORGANIC/ORGAN	Rapidly maturing for generation; synthesizability and experimental validation remain limiting	[15,16,22,41,42]

AI methodology	Primary drug-discovery application	Representative methods/tools	Evidentiary maturity	Ref.
Reinforcement learning	Molecular optimization, lead refinement, treatment-regime optimization	ReLeaSE, MolDQN, GCPN	Widely used to steer generative design; reward-design and validation gaps persist	[19,23,43]
Quantum computing / quantum ML	Molecular simulation, binding-mechanism calculation	VQE-based hybrid pipelines, QKDTI	Least mature; mostly conceptual, simulation-based or proof of concept as of 2026	[25,26,27]

4.2 Applications Across the Drug Discovery Pipeline

The literature clusters around five pipeline stages: target identification and validation, virtual screening and hit discovery, lead optimization and de novo design, ADMET and

safety pharmacology, and clinical and post-marketing applications. Each stage is read below through the pharmacological question it addresses (Table 3). Fig. 1 shows how the stages relate to one another and to the cross-cutting themes.

Table 3. Pharmacological question addressed at each pipeline stage, the AI contribution and the strength of current evidence.

Pipeline stage	Core pharmacological question	AI contribution	Evidence level	Ref.
Target identification and validation	Does modulating the target change disease biology in humans (pharmacodynamics)?	Multi-omics integration, knowledge graphs, target scoring	Provisional; definitive validation only after approval of a drug acting on the target	[44,45]
Hit discovery	Does the compound bind the target with adequate affinity and selectivity?	Deep-learning virtual screening, affinity prediction	Strongest and most reproducible; depends on target class	[46,47,48,49,50]
Lead optimization	Can potency, selectivity and drug-like properties be balanced?	Generative models with RL, multi-objective design	Rapidly maturing; synthesizability and experimental validation limit it	[41,42,43,51]
ADMET and safety pharmacology	What are the absorption, clearance, metabolic and off-target safety liabilities?	Graph- and transformer-based QSAR	Moderate; weaker outside the training chemical space	[34,35,36]
Clinical pharmacology and trials	Which patients and doses give a favorable exposure–response and benefit–risk balance?	Patient stratification, digital twins, trial optimization	Emerging; heterogeneous study designs	[52,53,54,55]
Pharmacovigilance	Can adverse drug reactions be detected earlier?	NLP and ML on EHR and spontaneous reports	Moderate; sensitivity still limited	[56,57]

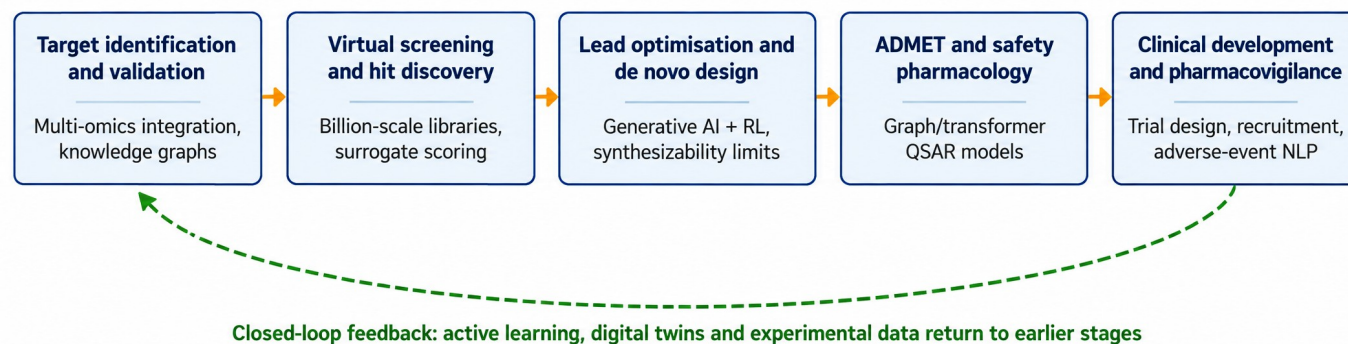


Fig. 1. Schematic overview of AI applications across the drug discovery and development pipeline, showing the five stages examined, the closed-loop feedback role of active learning and digital twins, and the cross-cutting themes that apply throughout.

4.2.1 Target Identification and Validation

Target identification increasingly combines multi-omics data and knowledge graphs [45] to prioritize disease-relevant targets at a scale that manual curation cannot match. The pharmacological question is whether modulating the target changes disease biology in humans. A target is considered definitively validated only when a drug acting on it is approved, often years after computational prioritization, so most AI-prioritized targets remain provisionally validated at best [44]. Target-scoring algorithms are not uniformly better than simple heuristics. A druggable, safe and differentiated target needs multi-dimensional assessment rather than a single AI-generated score [44].

4.2.2 Virtual Screening and Hit Discovery

Deep-learning-augmented virtual screening now routinely evaluates libraries of hundreds of millions to billions of make-on-demand compounds. Surrogate scoring models trained on docking or experimental data triage candidates before costlier physics-based or experimental evaluation [46,47,48,49]. This has widened the chemical space accessible to hit discovery. Gains in enrichment over classical docking depend on the compound and target class and are not universal, which fits the chemistry-facing evidentiary strength described above. Prospective ultra-large library docking campaigns have yielded experimentally confirmed hits [58]. Binding to one target does not establish selectivity. Off-target binding to related proteins, for example within a kinase family, underlies both adverse effects and polypharmacology [59], and single-target screening models do not capture it.

4.2.3 Lead Optimization and De Novo Molecular Design

Variational autoencoders, generative adversarial networks and, increasingly, diffusion models are used for de novo design under multi-objective constraints: potency, selectivity, synthesizability and ADMET [41,42]. The main unsolved

problem is no longer the validity of generated structures but their synthesizability. Molecules from earlier diffusion and VAE models often have poor synthetic accessibility, unstable ring systems or reactive functional groups. RL is frequently combined with generative models to steer this optimization. Synthetic accessibility and drug-likeness are usually approximated by computed scores such as the synthetic accessibility score [60] and the quantitative estimate of drug-likeness (QED) [61], but such proxies do not show that a molecule can actually be made [62]. Diffusion-based generative models have also been extended to de novo protein design [63].

Experimental proof of concept exists. A generative platform produced potent inhibitors of the kinase discoidin domain receptor 1 (DDR1) within weeks [64], deep-learning screening identified a structurally novel antibiotic that was active in mice [65], and explainable deep learning identified a new antibiotic structural class [66]. These are early, target-specific demonstrations and are not evidence of clinical benefit. They also show that prospective validation is technically possible, even though a scoping review of generative AI and LLMs in medication safety found that none of its 30 included studies tested models prospectively [56].

4.2.4 ADMET, Pharmacokinetics and Safety Pharmacology

Graph- and transformer-based QSAR models predict ADMET endpoints better than classical descriptor-based QSAR [34]. In pharmacological terms, they support early estimation of absorption, clearance and metabolic liability. One example is inhibition of cytochrome P450 enzymes, which underlies many drug–drug interactions. They also flag off-target safety liabilities such as hERG channel block, a recognized safety-pharmacology concern because of its link to QT prolongation [67]. Deep-learning models have been built specifically for

hERG liability [68], and directed message-passing networks are a widely used architecture for property prediction [69]. Performance remains sensitive to the chemical diversity of the training data, and extrapolation beyond the training distribution is a documented source of error [35,36].

4.2.5 Quantitative Performance Benchmarks

Table 4 collects quantitative benchmarks reported in individual studies, grouped by pipeline stage. Three cautions

apply. First, figures are reported as published. Apart from the pharmacovigilance row, which is itself a published meta-analysis, no pooling or cross-study statistical synthesis was performed, so the table is not a meta-analysis of AI performance. Second, metrics are not comparable across rows because datasets, splitting strategies (random versus scaffold) and evaluation protocols differ. Third, rows marked † are pre-2023 reference benchmarks kept as historical anchors; they do not describe the 2023–2026 state of the art.

Table 4. Reported quantitative benchmarks for AI/ML methods across the drug discovery and development pipeline, grouped by stage.

Reported metric	AI/ML result	Comparator	Source
A. Early discovery			
Virtual screening (structure-based, DUD-E): hit rate in top 1% of ranked library	55.6% (RF-Score-VS)	16.2% (AutoDock Vina)	Wójcikowski et al., Sci Rep 2017 [50] †
Structure prediction (AlphaFold2, CASP14): median GDT_TS; high-accuracy domains (GDT_TS > 70)	92.4 median; 87/92 domains high-accuracy	68.5 median (first AlphaFold, CASP13); ~40 median, free-modeling category (CASP12)	Jumper et al., Nature 2021 [37]; CASP14 assessment 2021 [70] †
De novo generation (REINVENT 4, structure-based PDK1 demonstration): hits defined as docking score ≤ −8 kcal/mol and QED ≥ 0.7	Cumulative number of hits rose over 50 RL epochs (6400 molecules at batch size 128)	RL from the prior compared with RL from a transfer-learning agent in the same demonstration	Loeffler et al., J Cheminform 2024 [51]
Drug repurposing (TxGNN, zero-shot): accuracy gain over 8 benchmarked methods, 17,080 diseases	+49.2% (indications); +35.1% (contraindications)	Best of 8 prior repurposing methods	Huang et al., Nat Med 2024 [40]
B. Clinical translation and post-marketing			
AI-nominated candidates: Phase I success rate	80–90%	40–65% (historical industry average)	Jayatunga et al., Drug Discov Today 2024 [71]
AI-nominated candidates: Phase II success rate	~40% (small sample)	Comparable to historical industry averages, not superior	Jayatunga et al., Drug Discov Today 2024 [71]
Rentosertib Phase IIa: change in forced vital capacity at 12 weeks, 60 mg once daily (n = 18) versus placebo (n = 17)	+98.4 mL (95% CI 10.9 to 185.9)	−20.3 mL (95% CI −116.1 to 75.6; placebo)	Xu et al., Nat Med 2025 [72]
Pharmacovigilance: adverse drug event prediction from EHR data, pooled estimates (random-effects meta-analysis)	Sensitivity 0.65 (95% CI 0.65–0.66); specificity 0.89 (95% CI 0.89–0.90); AUC 0.807	Diagnostic odds ratio 12.11 (95% CI 8.17–17.95)	Hu et al., Front Pharmacol 2024 [57]

Three observations follow from Table 4 (Fig. 2). First, the Phase I versus Phase II comparison is the most important nuance. AI-nominated candidates show a clear advantage in

Phase I, which is mainly a safety and pharmacokinetic test, but the advantage narrows to parity with historical averages in Phase II, which is an efficacy test [71]. This fits the chemistry–

biology divide: Phase I outcomes are close to the pharmacokinetic and safety properties that models are trained to predict, whereas Phase II efficacy lies further downstream and is harder to predict computationally.

Second, benchmark performance depends on how the data are split. Random-split validation can overstate real-world performance on novel chemical space, so scaffold-split or, better, prospective validation gives a more honest estimate

[35,36]. Benchmark suites and time-split validation were introduced to support such evaluation [73,74].

Third, AlphaFold2 (median GDT_TS 92.4 at CASP14, Fig. 3) and TxGNN (+49.2% zero-shot accuracy, Table 4) show some of the largest and most reproducible gains in the literature. Both are structural or relational prediction tasks (atomic coordinates and graph-based drug–disease association), not predictions of in vivo biological response.

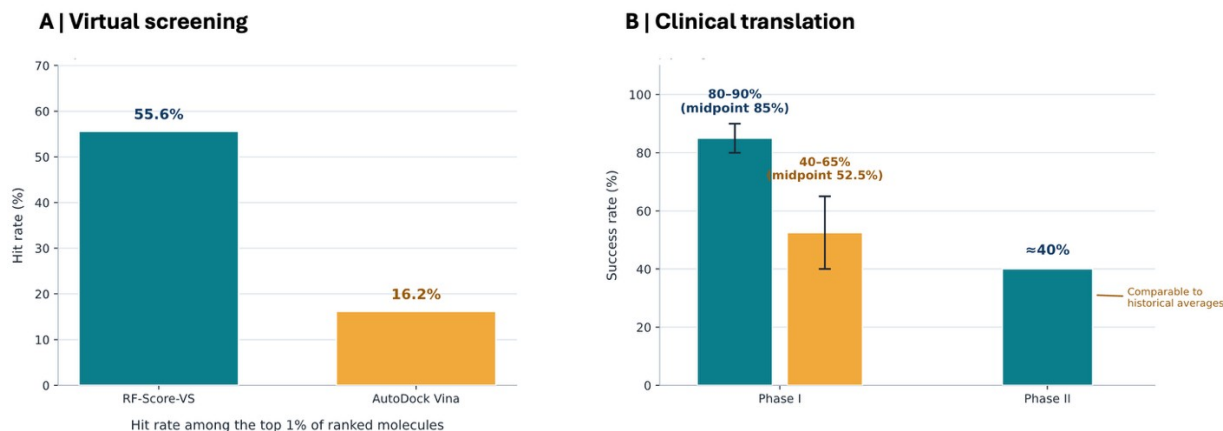


Fig. 2. Selected quantitative benchmarks from Table 4: virtual screening hit rate (left; pre-2023 reference benchmark) and Phase I and Phase II success rates for AI-nominated candidates versus historical averages (right). Bars show midpoints of reported ranges.

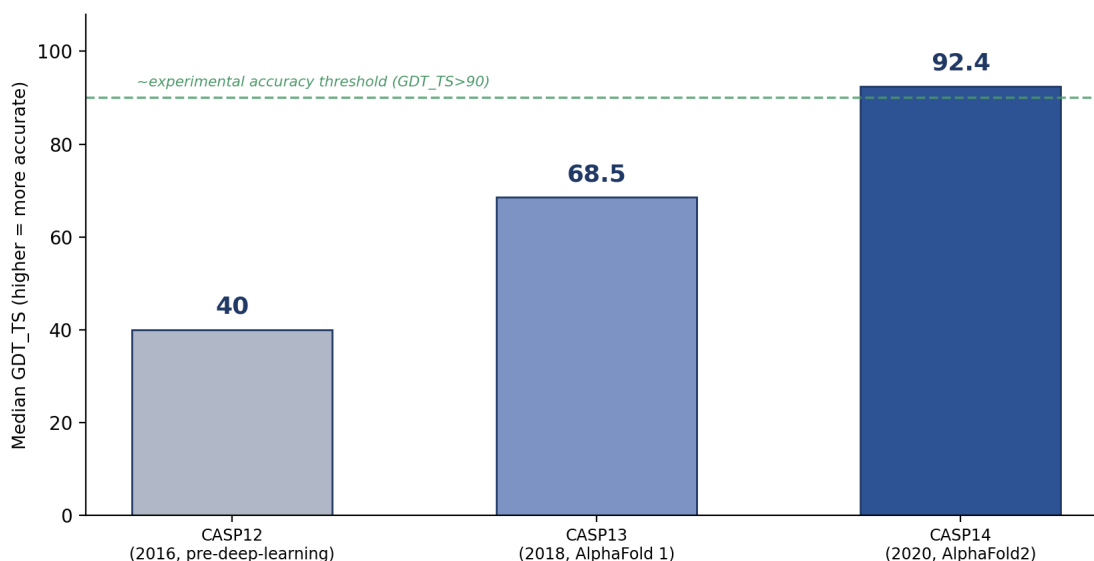


Fig. 3. Median structure-prediction accuracy (GDT_TS) across three successive CASP assessments, showing the step change attributable to AlphaFold and AlphaFold2. Data as reported in Table 4.

4.3 AI-Enabled Platforms and Tools

Table 5 summarizes representative tools and platforms. It includes only tools for which peer-reviewed benchmark or

clinical evidence was found in the reviewed literature, and it reports the evidence and its limits rather than vendor claims. Platforms whose performance claims are proprietary or not independently benchmarked are not listed.

Table 5. Representative AI tools and platforms with peer-reviewed benchmark or clinical evidence.

Tool / platform	Core AI approach	Primary application	Reported evidence and limitations	Ref.
AlphaFold2 / AlphaFold3	Attention-based and diffusion-based architectures	Protein structure and biomolecular complex prediction	Median GDT_TS of 92.4 at CASP14 for AlphaFold2; AlphaFold3 extends prediction to biomolecular complexes. Predictions are valuable hypotheses that do not replace experimental structure determination, and accuracy is reported for structure, not for in vivo pharmacological response.	[37,38,39,75]
RF-Score-VS	Random-forest scoring function	Structure-based virtual screening	Hit rate of 55.6% in the top 1% of ranked molecules versus 16.2% for AutoDock Vina on the full DUD-E data sets. Pre-2023 benchmark on one dataset family.	[50]
REINVENT 4	Open-source generative design with RL	De novo molecule generation and optimization	In a structure-based demonstration (PDK1), the cumulative number of hits, defined by docking score and QED thresholds, rose over 50 RL epochs. Hits are computational, not experimental.	[51]
TxGNN	Graph foundation model	Zero-shot drug repurposing	Accuracy gains of 49.2% (indications) and 35.1% (contraindications) over eight benchmarked methods. Prospective clinical confirmation remains rare.	[40]
DeepTox	Deep neural networks	Toxicity prediction	Reported state-of-the-art performance on the Tox21 Data Challenge. Performance depends on the chemical diversity of the training data.	[35,36]
Pharma.AI (PandaOmics, Chemistry42)	Generative DL with multi-omics target discovery	End-to-end target-to-candidate discovery	Produced rentosertib. Efficacy evidence so far rests on one small Phase IIa trial.	[72,76]

4.4 Clinical Translation: A Quantitative Snapshot

Quantifying clinical translation is hard because “AI-discovered” and “AI-enabled” are defined inconsistently across data sources. The most rigorous peer-reviewed accounting is an asset-level analysis published in Drug Discovery Today in 2024 [71]. It compared AI-originated assets entering human trials with historical industry benchmarks for Phase I and Phase II success (Table 4). It reports Phase I success of about 80–90% for AI-nominated candidates, against historical industry averages of roughly 40–65%. By December 2023, 24 AI-discovered molecules had completed Phase I and 10 had completed Phase II. Phase II success was about 40%, comparable to historical averages, and the analysis cautions that this sample is small. Because it was published in 2024, its data do not cover later trials.

Industry trackers report larger counts of AI-designed or AI-enabled programs in clinical development than the peer-

reviewed analysis. These counts use broader and inconsistent inclusion criteria, are not reproducible and are not reconciled with the peer-reviewed figure, so no industry number is used as evidence in this review. As of August 2026, we found no report of regulatory approval of an AI-discovered drug.

Rentosertib (formerly ISM001-055 / INS018-055) is an oral small-molecule inhibitor of the kinase TNK1, developed for idiopathic pulmonary fibrosis (IPF). Both its target and its chemical structure were AI-nominated. Preclinical characterization, including Phase I pharmacokinetic and safety data, was published independently [76]. The Phase IIa trial was randomized, double-blind and placebo-controlled, was run at 22 sites in China, and was published in Nature Medicine [72]. Its primary endpoint was safety (the percentage of patients with at least one treatment-emergent adverse event), which was similar across arms. Among the 71 patients, the 60 mg once-daily arm (n = 18) showed a mean forced vital

capacity (FVC) change of +98.4 mL (95% CI 10.9 to 185.9), against -20.3 mL (95% CI -116.1 to 75.6) in the placebo arm (n = 17) over 12 weeks. FVC is the standard functional endpoint in IPF, so the result is a clinically interpretable pharmacodynamic signal, but it was a secondary outcome with wide, overlapping confidence intervals. It is, to our knowledge, the clearest peer-reviewed controlled clinical evidence for an AI-originated drug to date, but it remains a small, 12-week, early-phase study.

The Phase III trial (GENESIS-IPF-3, NCT07687459) is registered by the sponsor as a randomized, double-blind,

placebo-controlled, 52-week study of once-daily oral rentosertib. The registry record was first submitted in June 2026 [77]. The record lists about 320 planned participants in China, and the primary endpoint is the annual rate of decline in forced vital capacity [77]. No Phase III efficacy data are available, and rentosertib remains investigational and unapproved.

4.5 Regulatory Landscape

Table 6 and Fig. 4 summarize the main regulatory milestones from 2023 to 2026.

Table 6. Selected regulatory milestones for AI in drug development, United States and European Union, 2023–2026.

Date	Development	Significance	Ref.
May 2023	FDA discussion paper on AI/ML in drug development (revised Feb 2025)	First cross-center articulation of AI regulatory considerations	[78]
Sep 2024	EMA reflection paper on the use of AI in the medicinal product lifecycle	Parallel, less formally codified regulatory science agenda	[79]
Jan 2025	FDA draft guidance: “Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products”	Establishes a risk-based, seven-step AI credibility framework	[80]
2026	Critical appraisals of the FDA draft guidance (for example Niazi 2026)	Identify scope gaps (discovery-stage AI, adaptive and continuously updated models) and recommend tiered explainability standards	[81]

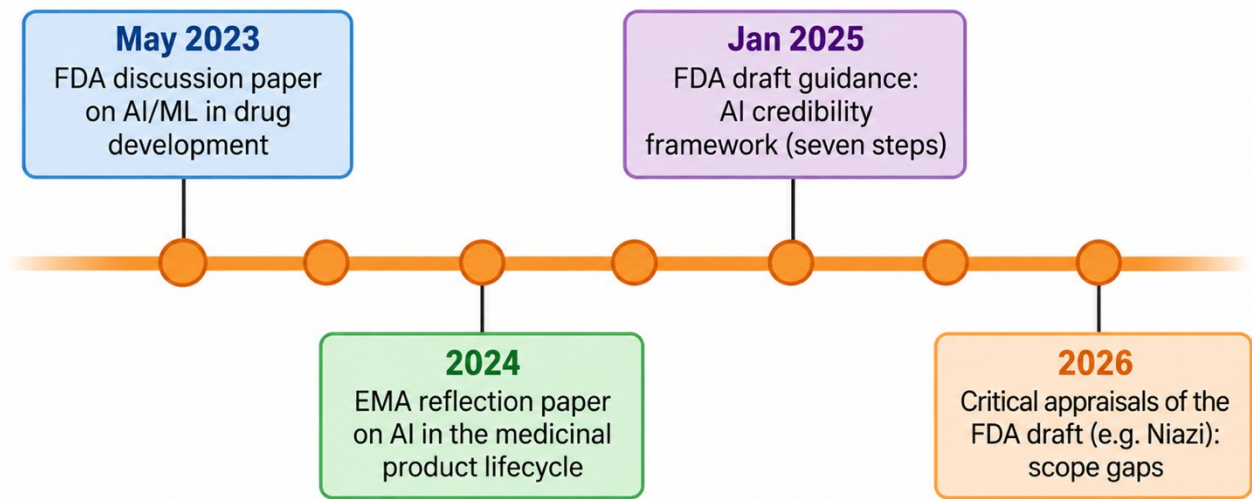


Fig. 4. Timeline of regulatory milestones for AI in drug development, 2023–2026, corresponding to Table 6. Markers denote FDA (blue), EMA (green) and independent critical-appraisal (orange) developments.

4.5.1 United States Food and Drug Administration

The FDA framework is risk-based. Sponsors must define the question of interest and the context of use of the AI model, assess model risk as a function of the model’s influence and

the consequence of an error, and run a credibility assessment plan matched to that risk [80]. The guidance applies to AI outputs that support regulatory decisions on safety, effectiveness or quality. It excludes AI used only for drug

discovery or for operational efficiency. This leaves early-stage discovery AI, the most heavily represented application area in the literature, largely outside formal oversight and governed by internal industry quality standards. Regulators already receive AI/ML components in submissions, and their number rose between 2016 and 2021 [82]. The guidance addresses model changes through lifecycle maintenance, but critical appraisals consider this insufficient for continuously updated (adaptive) models [81].

4.5.2 European Medicines Agency and International Harmonization

The EMA has pursued a parallel but less formally codified agenda. It emphasizes a lifecycle approach to AI reliability, human oversight, and integration with existing Good Clinical Practice and pharmacovigilance frameworks [79]. The FDA, EMA and ICH share a structural problem: regulatory science was built around static, validated tools, whereas many AI models are updated continuously. This creates tension between keeping a model current and the fixed-in-time validation that a regulatory submission presumes.

4.6 Drug Repurposing

Drug repurposing, which finds new indications for existing molecules with characterized safety, offers a comparatively low-risk route because repurposed candidates bypass much of the preclinical safety burden [40]. AI-driven repurposing uses knowledge graphs linking drugs, targets, pathways and diseases, transcriptomic signature matching and LLM-assisted literature mining to generate and rank hypotheses. TxGNN, for example, supports zero-shot repurposing through graph foundation modeling with interpretable multi-hop explanations [40]. Activity has been highest in oncology, neurodegeneration and infectious disease. Prospective clinical confirmation of AI-nominated repurposing hypotheses is still rare compared with the volume of computational predictions. Because repurposed drugs have known pharmacokinetics, dose ranges and safety profiles, early development can focus on whether the new indication responds. Polypharmacology, meaning activity at several targets, is often both the pharmacological basis of a repurposing signal and a source of adverse effects.

4.7 Personalized Medicine, Digital Twins and Model-Informed Drug Development

AI models that integrate genomic, transcriptomic and clinical data support patient stratification for enrichment designs, selection of therapy by molecular subtype and discovery of predictive and prognostic biomarkers. These are among the

more mature uses of AI, helped by large biobank and electronic health record (EHR) infrastructure.

Digital twins are continuously updated, multiscale virtual representations of a patient, organ system or disease process. They have moved from concept to applied research since 2023, including precision-medicine twins for Alzheimer's disease drug discovery [52] and generative-AI twins for clinical trial simulation [53]. In pharmacometric terms they extend model-informed drug development, which is the subject of the ICH M15 guideline, adopted in January 2026 [83]. AI is also entering pharmacometric and quantitative systems pharmacology workflows, and good-practice recommendations have been published [84,85]. Combining quantitative systems pharmacology with machine learning can reveal patterns in multidimensional data, but quantifying uncertainty, bias and error remains a central challenge [86]. In the authors' view, digital twins need formal verification, uncertainty quantification and validation standards before they can support regulatory decisions, in line with the credibility principles of model-informed drug development [83].

4.8 Nanomedicine and Advanced Drug Delivery

AI is increasingly applied to nanoparticle-based drug delivery [87] in four ways: predictive screening of nanocarrier–drug systems to reduce animal testing; optimizing drug–nanocarrier pairing for permeation, targeting and ADMET performance; generative design of nanocarriers for patient-specific data; and AI-guided repurposing of nanoformulated agents. A convergence of AI, nanotechnology and digital twins has been proposed for next-generation cancer theranostics [88]. It remains a research direction, not a validated clinical pathway.

4.9 Clinical Development and Pharmacovigilance

AI applications in clinical trials include protocol and eligibility-criteria optimization, EHR-based patient matching, decentralized and sensor-based monitoring, and prediction of trial outcomes and feasibility [54,55]. Reported effect sizes come from heterogeneous designs and should be read as directionally positive, not as generalizable performance guarantees.

In pharmacovigilance, natural language processing and deep learning mine EHR narratives, social media and spontaneous-report systems for adverse drug event signals. A 2025 scoping review of generative AI and LLMs in mitigating medication-related harm found promise for early adverse-event identification. It also noted that these tools have not been tested prospectively in real-world use, which the authors identify as the main near-term priority [56]. The pooled

sensitivity of 0.65 for EHR-based event prediction (Table 4) means that about one in three events would still be missed [57]. Such tools therefore suit signal support, not stand-alone detection.

4.10 Case Studies: Multi-Omics and Network Pharmacology

A recent case study shows how multi-omics integration and network pharmacology are combined with AI for target prioritization. The PI4AD framework [89] integrates multi-omics data, systems biology and artificial neural networks to prioritize Alzheimer's disease targets. It recovered clinically validated targets (including APP and ESR1) and identified nodal genes such as HRAS and MAPK1 within Ras-centered pathways implicated in synaptic dysfunction and neuroinflammation.

4.11 Cross-Cutting Themes

4.11.1 The Translational Trust Gap

A 2026 structured review of immuno-oncology AI models [90] found that high algorithmic accuracy leads to clinical adoption only with external or prospective validation, adequate interpretability, and regulatory and infrastructural readiness for continuously updated models. This pattern recurs across AI-driven drug discovery. The review recommends moving from model-centric optimization to system-level accountability, which applies equally to target identification.

4.11.2 Population-Level Validity and Equity

Biomedical training data are often concentrated in well-studied targets, diseases and populations, which limits how far model predictions generalize. A documented example is a widely used health-care algorithm that underestimated the needs of Black patients because it used cost as a proxy for need [91]. In drug discovery, the same problem arises when genomic or clinical training data under-represent some populations, which may affect target selection and the generalizability of exposure-response and safety predictions. Data quality is a related concern: merging bioactivity measurements for the same target from different assay sources has been shown to add considerable noise to training sets [92].

4.11.3 Environmental Sustainability (Green AI)

A 2026 review of "Green AI" in healthcare [93] identifies the energy use, electronic waste and carbon emissions of training and deploying large models as a growing concern, and outlines energy-efficient architectures and green-computing practices as mitigation. Large generative and transformer-based foundation models, the same classes increasingly used for

molecular design, are among the most computationally intensive AI systems in use, and the energy cost of training large models has been quantified in language-model research [94]. Their environmental footprint is therefore a legitimate design consideration.

4.11.4 Evidence-Based AI

AI outputs are too often judged by machine-learning metrics such as AUROC and accuracy. Clinical epidemiology has long required external validity, calibration, reproducibility and demonstrated effect on patient-relevant outcomes of any diagnostic or therapeutic tool. Reporting guidance for prediction models that use machine learning [95] and for clinical trials of AI interventions [96] now reflects these standards, and aligning AI development with them is a precondition for clinical trust.

5. DISCUSSION

5.1 Where AI Has Demonstrated Real Impact

The clearest and most reproducible gains cluster where training labels are close to the outcome of interest: structure prediction, docking-adjacent virtual screening and ADMET triage. This agrees with the chemistry-biology distinction of Bender and Cortés-Ciriano [8,9] and with earlier calls for an evidence-disciplined approach to AI in medicine [97,98]. Clinical translation, by contrast, is still early. Even the most favorable peer-reviewed accounting [71] describes a field in which most tracked AI-originated assets have not progressed beyond Phase I, and none has reached approval.

This does not show that AI drug discovery is failing. Phase I success rates well above historical averages are a meaningful, if provisional, positive signal. It does show that claims that AI has already "revolutionized" drug development go beyond the current clinical record.

5.2 Structural Challenges by AI Paradigm

Challenges are organized by paradigm because shared issues (validation gaps, interpretability and data quality) have different practical consequences for RL, LLM, QC and GAI applications. The points below are the authors' qualitative synthesis of the cited literature, not frequency counts.

5.2.1 Reinforcement Learning

Reward design is a recurring difficulty. Composite, context-dependent objectives make simple reward proxies vulnerable to reward hacking, and non-differentiable targets such as synthetic accessibility and drug-likeness complicate credit assignment. Oracle quality and distribution shift add to this: a

narrow QSAR or docking model used as a reward oracle can mislead the search when the generative policy drifts outside the oracle's valid domain. Published demonstrations usually score success by computational thresholds, for example docking score and QED, and not by experimental activity [51].

5.2.2 Large Language Models

Reliability is the main concern. Models can produce fluent but incorrect content, including fabricated citations or unsupported target–disease associations, which is risky for hypothesis generation and repurposing. Transparency and governance, privacy and data security (especially for clinical text and longitudinal EHR data), and bias across under-represented populations and diseases also matter. In medication-safety applications, a scoping review found promise for early identification of adverse drug events but no prospective testing of the models [56].

5.2.3 Quantum Computing

QC challenges cluster around hardware immaturity (NISQ devices limit circuit depth and reliability [27]), limited scalability from proof-of-concept demonstrations to drug-sized systems, and validation gaps, because many evaluations lack rigorous comparison with experimental measurements.

Workflow complexity (integration overhead in hybrid pipelines) and adoption barriers (wall-clock performance often slower than classical methods, plus workforce and infrastructure costs) complete the picture. No single bottleneck dominates, which reflects QC's early stage.

5.2.4 Generative AI

The concerns about reliability apply equally to molecule-generating models. The main GAI-specific concern is limited real-world validation, because most systems are evaluated retrospectively or on proxy benchmarks (validity, novelty, diversity, QED and docking score) instead of prospective or wet-laboratory studies, although exceptions exist [64,65,66]. A transparency gap arises from proprietary training data and undisclosed architectures, and novelty claims are often contested when outputs resemble known compounds. Poor synthesizability and weak cross-target generalization add to these concerns and agree with the synthesizability problem described in Section 4.2.3.

5.2.5 Cross-Cutting Structural Challenges

Table 7 summarizes six challenges that recur across all four paradigms.

Table 7. Cross-cutting structural challenges to AI-driven drug discovery, synthesized across the reviewed literature.

Challenge	Core issue	Representative evidence
Data quality and equity	Training data skewed towards well-studied targets, diseases and high-income populations	Algorithmic bias in a widely used health-care algorithm (2019) [91]; noisy merged bioactivity data (2025) [92]
Interpretability	Deep and transformer models remain low in interpretability relative to regulatory stakes	FDA risk-based credibility framework (2025) [80]
Synthesizability	Generated molecules are often synthetically inaccessible or unstable	Generative-AI molecular design review (2024) [41]
Prospective validation	Retrospective and in silico validation dominates; wet-lab and prospective validation are rare	No prospective testing in a scoping review of generative AI and LLMs in medication safety [56]; prospective examples exist in antibiotic discovery [65]
Regulatory scope gaps	Discovery-stage and adaptive models fall outside the guidance (discovery) or are handled only through general lifecycle maintenance (adaptive models)	Niazi (2026) critical appraisal of the FDA draft guidance [81]
Environmental sustainability	High energy and carbon cost of large generative and foundation models	Green AI in healthcare review (2026) [93]

5.3 Comparative Maturity of AI Paradigms

Drawing on the evidence reviewed here, RL and GAI show the strongest methodological maturity for molecular design and optimization. LLMs show the strongest near-term readiness

for knowledge-intensive tasks (literature mining, hypothesis generation and evidence synthesis), not for direct molecular generation. QC is the least mature paradigm for practical drug discovery. Across all paradigms, the main barrier is the limited

progression from in silico evaluation to prospective experimental and clinical validation. The immuno-oncology analysis [90] reached the same finding independently, which suggests it is a structural property of the field and not an artifact of one sub-domain.

5.4 Proposed Pathways for Regulating Continuously Updated AI

Current FDA draft guidance leaves two gaps that matter for clinical translation: discovery-stage AI, which is out of scope, and AI systems that are updated continuously, which are covered only by general lifecycle-maintenance expectations [80,81]. The regulatory logic of fixed-in-time validation does not fit models that learn from new data. Table 8 therefore proposes five concrete pathways. They are the authors'

proposals, built from existing regulatory instruments and published critiques; they are not existing requirements.

The pathways follow one principle: regulatory scrutiny should scale with the influence of the model on a decision and the consequence of an error, as in the FDA risk matrix [80], and should be defined before the model changes, not after. A predetermined change control plan (PCCP) has already been formalized by the FDA for AI-enabled device software [99], and the EMA reflection paper takes a lifecycle approach [79]. Neither has been applied to AI used in drug development. For safety pharmacology, AI predictions such as hERG liability should support, and not replace, the nonclinical and clinical QT assessment framework [67]. Explainable AI methods that could support such tiered documentation have been reviewed for drug discovery [100].

Table 8. Proposed pathways to close the regulatory gap for discovery-stage and continuously updated AI in drug development (authors' proposals).

Gap	Proposed mechanism	Key elements	Basis
Explainability that does not scale with risk	Tiered explainability tied to context of use	Low-influence, low-consequence uses: documentation and performance reporting. High-influence uses (for example dose selection or safety conclusions): feature attribution, mechanistic plausibility check and uncertainty quantification.	[66,80,81,100]
Model changes after submission	Predetermined change control plan for drug-development AI	Pre-specified permitted modifications, scope of retraining data, performance acceptance thresholds, rollback criteria and version identifiers.	[79,99]
Performance drift over time	Continuous validation protocol	Locked-version submission; scaffold- or time-split and prospective validation; calibration reporting; drift monitoring with defined triggers for revalidation.	[57,71,74]
Unregulated discovery-stage AI	Proportionate voluntary quality framework	Model documentation, data provenance, version control and decision logs, without imposing the burden of a full credibility assessment on exploratory work.	[81]
Weak link to pharmacological models	Alignment with model-informed drug development	Link AI outputs to PK/PD and exposure–response reasoning; define context of use and model credibility along the lines of ICH M15; treat AI safety predictions as supportive evidence.	[67,83]

Harmonization between the FDA, EMA and ICH would reduce the risk that sponsors face divergent expectations for the same adaptive model.

5.5 A Pharmacological Evidence Checklist for AI-Originated Candidates

A pharmacologist reviewing an AI-originated candidate should ask the same questions as for any other investigational drug. The origin of the molecule changes how it was found,

not what must be shown. Table 9 lists the pharmacological evidence that should accompany such a candidate, what AI currently contributes at each point, and the gap that must be

closed with experimental or clinical data. The checklist is the authors' proposal, built from the evidence reviewed above.

Table 9. Pharmacological evidence checklist for AI-originated candidates (authors' proposal).

Pharmacological element	Current AI contribution	Evidence needed beyond the model
Target engagement and mechanism of action	Target prioritization from multi-omics data; binding and affinity prediction	Biochemical and cellular confirmation of binding and selectivity; evidence that modulating the target changes disease biology
Pharmacokinetics (absorption, distribution, clearance, half-life)	ADMET models trained on historical compounds	Measured pharmacokinetics in preclinical species and humans; scaffold-split or prospective checks on novel chemical space
Metabolism and drug–drug interactions	Prediction of cytochrome P450 liability	In vitro and clinical interaction studies for the relevant enzymes and transporters
Pharmacodynamic biomarker and clinical endpoint	Patient stratification; digital-twin simulation	A validated biomarker or endpoint (for example FVC in IPF) measured in a controlled trial
Dose–exposure–response	Model-informed simulation; digital twins	Prospective dose-ranging data and pharmacometric analysis consistent with ICH M15 [83]
Safety pharmacology	Prediction of off-target and hERG liability	Nonclinical and clinical QT assessment as described in ICH E14/S7B [67]; AI prediction as supportive evidence only
Special populations and equity	Limited; training data are skewed [91]	Exposure and response data in under-represented populations
Pharmacovigilance	Machine-learning signal detection (pooled sensitivity 0.65 [57])	Prospective evaluation and human review of signals [56]
Model credibility and change control	Credibility frameworks are emerging [80]	Defined context of use, version control, predetermined change control (Table 8) and transparent reporting (TRIPOD+AI [95]; CONSORT-AI for trials [96])

5.6 Implications for Stakeholders

Pharmacologists and clinical pharmacologists. Pharmacological evidence remains the decisive test: measured exposure, target engagement, dose–response and safety margins. AI predictions can prioritize experiments but cannot replace these measurements (Table 9).

Researchers and method developers. The chemistry–biology divide suggests that effort will yield the largest near-term returns when directed at biology-facing tasks such as target validation and in vivo efficacy prediction, not at further refinement of mature tasks such as docking and structure prediction.

Industry and investors. The gap between the peer-reviewed asset-level analysis and larger industry-tracked counts shows that “AI-discovered” is an inconsistently defined and imperfectly audited category. Valuation or investment

decisions that rely on unreconciled industry figures should treat them with caution.

Regulators. Exclusion of discovery-stage AI from current guidance leaves the stage with the largest concentration of applications under internal quality standards. Extending risk-based credibility frameworks to this stage, without disproportionate burden on exploratory research, is a plausible near-term priority [81].

Clinicians and patients. AI-originated therapeutics should be judged with the same clinical-trial scrutiny as any other investigational drug. An AI-driven origin tells us about the discovery process; it is not evidence of superior safety or efficacy.

5.7 Future Directions

Three developments are likely to decide whether current AI investment produces approved therapeutics. First,

standardized and, ideally, prospective validation protocols with transparent reporting, which is also the main gap identified for generative AI in medication safety [56]. Second, regulatory frameworks that extend to discovery-stage and adaptive AI systems, for example through the pathways in Table 8. Third, closer integration of generative and predictive AI with experimental feedback loops (active learning, closed-loop discovery and digital twins), so that computational gains are tested against biological ground truth and do not substitute for it.

6. LIMITATIONS

This is a narrative, unregistered review without a PRISMA flow diagram, record counts or formal risk-of-bias appraisal, so selection may favor well-known or positive results. Clinical-pipeline figures rest on one small asset-level analysis, and efficacy evidence comes from one short Phase IIa trial.

7. CONCLUSION

AI has measurably reshaped early-stage drug discovery. It has produced structure-prediction tools such as AlphaFold3 and screening and generative platforms that have yielded experimentally confirmed hits [58,64,65], although the evidence reviewed does not show an effect on development timelines. Clinical and regulatory translation is real but early. As of 2026, a peer-reviewed asset-level analysis and larger, differently defined industry counts describe AI-originated therapeutics in human trials, with rentosertib the most advanced (Phase III trial registered in 2026, no efficacy data yet) and no AI-designed drug yet approved.

The most consequential unresolved challenges are translational and not computational: prospective validation, interpretability matched to regulatory stakes, equitable and representative training data, and regulatory frameworks able to oversee continuously updated models. The tiered explainability, predetermined change control and continuous validation pathways proposed here are one route to closing the last gap. Future progress will be measured not by benchmark performance but by the standard that has always governed pharmaceutical innovation: reproducible, prospectively validated clinical benefit.

AUTHORS' CONTRIBUTIONS

Sonali Arora: conceptualization, investigation, writing – original draft, writing – review and editing. Sachin Sharma: investigation, data curation, writing – review and editing, visualization. Ravi Jayant: conceptualization, supervision, writing – review and editing, project administration. All

authors have accepted responsibility for the content of the manuscript and approved the final version for publication.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

No new datasets were generated or analyzed during the current study. All data supporting the findings of this review are available within the published literature cited in the References section.

FUNDING

The authors received no specific funding for this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

USE OF AI IN MANUSCRIPT PREPARATION

Grammarly tools were used to assist with language editing. All scientific content, interpretations, and conclusions were reviewed and approved by the authors, who take full responsibility for the accuracy of the manuscript.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

1. Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA*. 2020;323(9):844–853. doi:10.1001/jama.2020.1166 [Correction published 2022.]
2. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ*. 2016;47:20–33. doi:10.1016/j.jhealeco.2016.01.012
3. Avorn J. The \$2.6 billion pill methodologic and policy considerations. *N Engl J Med*. 2015;372(20):1877–1879. doi:10.1056/NEJMp1500848
4. 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;12(7):3049–3062. doi:10.1016/j.apsb.2022.02.002
5. Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov*. 2015;14(7):475–486. doi:10.1038/nrd4609

6. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today*. 2021;26(1):80–93. doi:10.1016/j.drudis.2020.10.010
7. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18(6):463–477. doi:10.1038/s41573-019-0024-5
8. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. *Drug Discov Today*. 2021;26(2):511–524. doi:10.1016/j.drudis.2020.12.009
9. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data used for AI in drug discovery. *Drug Discov Today*. 2021;26(4):1040–1052. doi:10.1016/j.drudis.2020.11.037
10. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al. Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov*. 2020;19(5):353–364. doi:10.1038/s41573-019-0050-3
11. Shakeri E, Far B. Artificial intelligence in healthcare and drug discovery: a scoping review of current trends, challenges, and future directions. *Intell Based Med*. 2026;15:100442. doi:10.1016/j.ibmed.2026.100442
12. Niazi SK. Artificial intelligence in small-molecule drug discovery: a critical review of methods, applications, and real-world outcomes. *Pharmaceuticals (Basel)*. 2025;18(9):1271. doi:10.3390/ph18091271
13. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Adv Neural Inf Process Syst*. 2014;27:2672–2680. arXiv:1406.2661.
14. Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv. 2013;1312.6114.
15. Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. *Proc Mach Learn Res*. 2018;80:2323–2332. arXiv:1802.04364.
16. Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci*. 2018;4(2):268–276. doi:10.1021/acscentsci.7b00572
17. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst*. 2020;33:6840–6851. arXiv:2006.11239.
18. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30:5998–6008. arXiv:1706.03762.
19. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529–533. doi:10.1038/nature14236
20. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930–1940. doi:10.1038/s41591-023-02448-8
21. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259–265. doi:10.1038/s41586-023-05881-4
22. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. *J Cheminform*. 2017;9:48. doi:10.1186/s13321-017-0235-x
23. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv*. 2018;4(7):eaap7885. doi:10.1126/sciadv.aap7885
24. Yu C, Liu J, Nemati S, Yin G. Reinforcement learning in healthcare: a survey. *ACM Comput Surv*. 2021;55(1):5. doi:10.1145/3477600
25. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. *Nature*. 2017;549(7671):195–202. doi:10.1038/nature23474
26. Cao Y, Romero J, Olson JP, Degroote M, Johnson PD, Kieferová M, et al. Quantum chemistry in the age of quantum computing. *Chem Rev*. 2019;119(19):10856–10915. doi:10.1021/acs.chemrev.8b00803
27. Preskill J. Quantum computing in the NISQ era and beyond. *Quantum*. 2018;2:79. doi:10.22331/q-2018-08-06-79
28. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444. doi:10.1038/nature14539
29. David L, Thakkar A, Mercado R, Engkvist O. Molecular representations in AI-driven drug discovery: a review and practical guide. *J Cheminform*. 2020;12(1):56. doi:10.1186/s13321-020-00460-5
30. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci*. 1988;28(1):31–36. doi:10.1021/ci00057a005
31. Krenn M, Häse F, Nigam A, Friederich P, Aspuru-Guzik A. Self-referencing embedded strings (SELFIES): a 100% robust molecular string representation. *Mach Learn Sci Technol*. 2020;1(4):045024. doi:10.1088/2632-2153/aba947
32. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. 2010;50(5):742–754. doi:10.1021/ci100050t
33. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci*. 2002;42(6):1273–1280. doi:10.1021/ci010132r

34. Zhang J, Li H, Zhang Y, Huang J, Ren L, Zhang C, et al. Computational toxicology in drug discovery: applications of artificial intelligence in ADMET and toxicity prediction. *Brief Bioinform.* 2025;26(5):bbaf533. doi:10.1093/bib/bbaf533
35. Bai C, Wu L, Li R, Cao Y, He S, Bo X. Machine learning-enabled drug-induced toxicity prediction. *Adv Sci (Weinh).* 2025;12(16):e2413405. doi:10.1002/advs.202413405
36. Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: toxicity prediction using deep learning. *Front Environ Sci.* 2016;3:80. doi:10.3389/fenvs.2015.00080
37. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–589. doi:10.1038/s41586-021-03819-2
38. Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* 2024;52(D1):D368–D375. doi:10.1093/nar/gkad1011
39. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630(8016):493–500. doi:10.1038/s41586-024-07487-w
40. Huang K, Chandak P, Wang Q, Havaladar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. *Nat Med.* 2024;30(12):3601–3613. doi:10.1038/s41591-024-03233-x
41. Gangwal A, Ansari A, Ahmad I, Azad AK, Kumarasamy V, Subramaniyan V, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. *Front Pharmacol.* 2024;15:1331062. doi:10.3389/fphar.2024.1331062
42. Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: generative models for matter engineering. *Science.* 2018;361(6400):360–365. doi:10.1126/science.aat2663
43. Sun Y, Chen L, Deng Z, et al. De novo drug design by iterative multiobjective deep reinforcement learning with graph-based molecular quality assessment. *Bioinformatics.* 2023;39(4):btad157. doi:10.1093/bioinformatics/btad157
44. Pun FW, Podolskiy D, Izumchenko E, Mortlock A, Oprea TI, Scheibye-Knudsen M, Fortney K, Morgen E, Ren F, Zhavoronkov A. Target identification and assessment in the era of AI. *Nat Rev Drug Discov.* 2026. Published online 20 April 2026. doi:10.1038/s41573-026-01412-8
45. Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. *Sci Data.* 2023;10:67. doi:10.1038/s41597-023-01960-3
46. Kim H, Shim H, Ranganath A, He S, Stevenson G, Allen JE. Protein-ligand binding affinity prediction using multi-instance learning with docking structures. *Front Pharmacol.* 2025;15:1518875. doi:10.3389/fphar.2024.1518875
47. Brown BP. A generalizable deep learning framework for structure-based protein–ligand affinity ranking. *Proc Natl Acad Sci U S A.* 2025;122(42):e2508998122. doi:10.1073/pnas.2508998122
48. Wang DD, Wu W, Wang R. Structure-based, deep-learning models for protein-ligand binding affinity prediction. *J Cheminform.* 2024;16:2. doi:10.1186/s13321-023-00795-9
49. Kaneriy A, Samudrala M, Ganesh H, Moran J, Dandibhotla S, Dakshanamurthy S. StructureNet: physics-informed hybridized deep learning framework for protein–ligand binding affinity prediction. *Bioengineering (Basel).* 2025;12(5):505. doi:10.3390/bioengineering12050505
50. Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci Rep.* 2017;7:46710. doi:10.1038/srep46710
51. Loeffler HH, He J, Tibo A, Janet JP, Voronov A, Mervin LH, Engkvist O. Reinvent 4: modern AI-driven generative molecule design. *J Cheminform.* 2024;16:20. doi:10.1186/s13321-024-00812-5
52. Ren Y, Pieper AA, Cheng F. Utilization of precision medicine digital twins for drug discovery in Alzheimer’s disease. *Neurotherapeutics.* 2025:e00553. Epub 2025 Feb 17. doi:10.1016/j.neurot.2025.e00553
53. Bordukova M, Makarov N, Rodriguez-Esteban R, Schmich F, Menden MP. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opin Drug Discov.* 2024;19(1):33–42. doi:10.1080/17460441.2023.2273839
54. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci.* 2019;40(8):577–591. doi:10.1016/j.tips.2019.05.005
55. Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, et al. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials.* 2021;22:537. doi:10.1186/s13063-021-05489-x
56. Ong JCL, Chen MH, Ng N, Elangovan K, Tan NYT, Jin L, et al. A scoping review on generative AI and large language models in mitigating medication related harm. *NPJ Digit Med.* 2025;8(1):157. doi:10.1038/s41746-025-01565-7
57. Hu Q, Chen Y, Zou D, He Z, Xu T. Predicting adverse drug event using machine learning based on electronic health records: a systematic review and meta-analysis. *Front Pharmacol.* 2024;15:1497397. doi:10.3389/fphar.2024.1497397

58. Lyu J, Wang S, Balias TE, Singh I, Levit A, Moroz YS, et al. Ultra-large library docking for discovering new chemotypes. *Nature*. 2019;566(7743):224–229. doi:10.1038/s41586-019-0917-9
59. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol*. 2011;29(11):1046–1051. doi:10.1038/nbt.1990
60. Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminform*. 2009;1:8. doi:10.1186/1758-2946-1-8
61. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. *Nat Chem*. 2012;4(2):90–98. doi:10.1038/nchem.1243
62. Gao W, Coley CW. The synthesizability of molecules proposed by generative models. *J Chem Inf Model*. 2020;60(12):5714–5723. doi:10.1021/acs.jcim.0c00174
63. Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. De novo design of protein structure and function with RFdiffusion. *Nature*. 2023;620(7976):1089–1100. doi:10.1038/s41586-023-06415-8
64. Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol*. 2019;37(9):1038–1040. doi:10.1038/s41587-019-0224-x
65. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. *Cell*. 2020;180(4):688–702.e13. doi:10.1016/j.cell.2020.01.021
66. Wong F, Zheng EJ, Valeri JA, Donghia NM, Anahtar MN, Omori S, et al. Discovery of a structural class of antibiotics with explainable deep learning. *Nature*. 2024;626(7997):177–185. doi:10.1038/s41586-023-06887-8
67. International Council for Harmonisation. ICH E14/S7B: Clinical and nonclinical evaluation of QT/QTc interval prolongation and proarrhythmic potential, questions and answers. Geneva: ICH; 2022.
68. Cai C, Guo P, Zhou Y, Zhou J, Wang Q, Zhang F, et al. Deep learning-based prediction of drug-induced cardiotoxicity. *J Chem Inf Model*. 2019;59(3):1073–1084. doi:10.1021/acs.jcim.8b00769
69. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. *J Chem Inf Model*. 2019;59(8):3370–3388. doi:10.1021/acs.jcim.9b00237
70. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Applying and improving AlphaFold at CASP14. *Proteins*. 2021;89(12):1711–1721. doi:10.1002/prot.26257
71. Jayatunga MKP, 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;29(6):104009. doi:10.1016/j.drudis.2024.104009
72. 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(8):2602–2610. doi:10.1038/s41591-025-03743-2
73. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. 2018;9(2):513–530. doi:10.1039/C7SC02664A
74. Sheridan RP. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J Chem Inf Model*. 2013;53(4):783–790. doi:10.1021/ci400084k
75. Terwilliger TC, Liebschner D, Croll TI, Williams CJ, McCoy AJ, Poon BK, et al. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat Methods*. 2024;21(1):110–116. doi:10.1038/s41592-023-02087-4
76. Ren F, Aliper A, Chen J, et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat Biotechnol*. 2025;43(1):63–75. doi:10.1038/s41587-024-02143-0
77. ClinicalTrials.gov. Study evaluation rentosertib (INS018_055) administered orally in patients with idiopathic pulmonary fibrosis (IPF) [GENESIS-IPF-3]. Identifier NCT07687459. Sponsor: InSilico Medicine Hong Kong Limited. First submitted 2026 Jun 22.
78. U.S. Food and Drug Administration. Using artificial intelligence and machine learning in the development of drug and biological products: discussion paper and request for feedback. Silver Spring (MD): FDA; 2023 May (revised 2025 Feb).
79. European Medicines Agency. Reflection paper on the use of artificial intelligence in the medicinal product lifecycle. Amsterdam: EMA; 2024 Sep.
80. U.S. Food and Drug Administration. Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: draft guidance for industry and other interested parties. Silver Spring (MD): FDA; 2025 Jan. Docket FDA-2024-D-4689. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological>
81. Niazi SK. A critical review of the FDA's draft guidance on artificial intelligence in drug and biological product regulation. *J Chem*. 2026;2026:5202999. doi:10.1155/joch/5202999

82. Liu Q, Zhu H, Liu C, Jean D, Huang SM, ElZarrad MK, et al. Landscape analysis of the application of artificial intelligence and machine learning in regulatory submissions for drug development from 2016 to 2021. *Clin Pharmacol Ther.* 2023;113(4):771–774. doi:10.1002/cpt.2668
83. International Council for Harmonisation. ICH M15: General principles for model-informed drug development (final guideline, Step 4). Geneva: ICH; 2026 Jan 29. Available from: <https://www.fda.gov/media/184747/download>
84. Terranova N, Renard D, Shahin MH, Menon S, Cao Y, Hop CE, et al. Artificial intelligence for quantitative modeling in drug discovery and development: an Innovation and Quality Consortium perspective on use cases and best practices. *Clin Pharmacol Ther.* 2024;115(4):658–672. doi:10.1002/cpt.3053
85. Shahin MH, Desai P, Terranova N, Guan Y, Helikar T, Lobentanzer S, et al. AI-driven applications in clinical pharmacology and translational science: insights from the ASCPT 2024 AI preconference. *Clin Transl Sci.* 2025 Apr. doi:10.1111/cts.70203
86. Folguera-Blasco N, Boshier FAT, Uatay A, Pichardo-Almaraz C, Lai M, Biasetti J, et al. Coupling quantitative systems pharmacology modelling to machine learning and artificial intelligence for drug development: its pAIns and gAIns. *Front Syst Biol.* 2024;4:1380685. doi:10.3389/fsysb.2024.1380685
87. Ho D, Wang P, Kee T. Artificial intelligence in nanomedicine. *Nanoscale Horiz.* 2019;4(2):365–377. doi:10.1039/C8NH00233A
88. Adir O, Poley M, Chen G, Froim S, Krinsky N, Shklover J, et al. Integrating artificial intelligence and nanotechnology for precision cancer medicine. *Adv Mater.* 2020;32(13):e1901989. doi:10.1002/adma.201901989
89. Yang Y, Diao Y, Jiang L, Li F, Chen L, Ni M, Wang Z, et al. A computational medicine framework integrating multi-omics, systems biology, and artificial neural networks for Alzheimer's disease therapeutic discovery. *Acta Pharm Sin B.* 2025;15(9):4411–4426. doi:10.1016/j.apsb.2025.07.018
90. Oisakede EO, Ayo Daniel RI, Olawuyi OF, Alabi JO, Analikwu CC, Olawade DB. Translational gaps in immuno-AI: from algorithmic accuracy to clinical trust. *Hum Immunol.* 2026;87(8):111774. doi:10.1016/j.humimm.2026.111774
91. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447–453. doi:10.1126/science.aax2342
92. Behnam MAM, Cavalli A, Lousa D, Soares CM, Klein CD. Decoding of inconsistent biological data: a critical step toward enhanced AI predictivity in drug discovery. *ACS Pharmacol Transl Sci.* 2025. Published online 14 December 2025. doi:10.1021/acspsci.5c00677
93. Alzoubi YI, Mishra A. Green artificial intelligence in health applications. *Artif Intell Med.* 2026;178:103442. doi:10.1016/j.artmed.2026.103442
94. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* 2019:3645–3650. doi:10.18653/v1/P19-1355
95. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378
96. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med.* 2020;26(9):1364–1374. doi:10.1038/s41591-020-1034-x
97. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7
98. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med.* 2019;25(1):24–29. doi:10.1038/s41591-018-0316-z
99. U.S. Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions (final guidance). Silver Spring (MD): FDA; 2024 Dec.
100. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell.* 2020;2(10):573–584. doi:10.1038/s42256-020-00236-4