AI-Enabled Drug Discovery in Pharmaceutical Sciences: A Systematic Review of Molecular Models, Target Prediction, Lead Optimization, and Translational Barriers

James Anderson[1], Maria Rossi[2]*, William Clark[1]

[1]Department of AI-Enabled Pharmacognosy and Natural Product Discovery, Faculty of Pharmacy, University of Manchester, Manchester, United Kingdom.
[2]Department of Computational Pharmacology and Molecular Docking, Faculty of Pharmacy, University of Milan, Milan, Italy.

ABSTRACT

Artificial intelligence has become a major methodological force in pharmaceutical drug discovery, supporting molecular modeling, compound prioritization, target prediction, virtual screening, de novo design, ADMET estimation, and lead optimization. This systematic review evaluates how AI-enabled methods are used across drug discovery tasks and how far their outputs can be interpreted as translationally meaningful evidence. The review was designed around focused questions concerning molecular representations, model families, target and drug–target prediction, lead optimization workflows, validation strategies, and barriers to pharmaceutical implementation. The reviewed evidence covers machine learning, deep learning, graph neural networks, transformer architectures, generative models, multitask learning, transfer learning, explainable AI, benchmark resources, and structure-based computational approaches. Particular attention is given to the distinction between retrospective benchmark performance, in silico prioritization, prospective experimental validation, preclinical relevance, clinical translation, and regulatory readiness. The synthesis indicates that AI can strengthen hypothesis generation, expand searchable chemical space, prioritize compounds and targets, and improve decision support in early discovery. However, reliable translation depends on dataset quality, leakage-aware benchmarking, assay context, biological plausibility, reproducibility, interpretability, uncertainty characterization, and experimental confirmation. The review contributes a systematic evidence framework linking molecular model design to validation needs and translational barriers. Future pharmaceutical AI research should move beyond isolated benchmark improvement toward robust, explainable, externally validated, and experimentally integrated discovery workflows.

Key Words: *Artificial intelligence, Drug discovery, Pharmaceutical sciences, Systematic review, Molecular modeling, Target prediction*

eIJPPR 2025; 15(1):12-23

HOW TO CITE THIS ARTICLE: Anderson J, Rossi M, Clark W. AI-Enabled Drug Discovery in Pharmaceutical Sciences: A Systematic Review of Molecular Models, Target Prediction, Lead Optimization, and Translational Barriers. Int J Pharm Phytopharmacol Res. 2025;15(1):12-23. https://doi.org/10.51847/6JTpI8nEIi

INTRODUCTION

AI-enabled drug discovery has moved from a specialized computational activity to a central component of pharmaceutical research, where machine learning is increasingly applied to compound screening, molecular property prediction, target prioritization, pharmacokinetic prediction, safety estimation, and development decision support [1]. This transformation has been driven by the availability of bioactivity databases, larger chemical libraries, high-throughput assays, protein sequence and structure resources, and algorithmic methods capable of

Corresponding author: Maria Rossi
Address: Department of Computational Pharmacology and Molecular Docking, Faculty of Pharmacy, University of Milan, Milan, Italy.
E-mail: ✉ maria.rossi@unimi.it
Received: 11 November 2024; **Revised:** 09 Janaury 2025; **Accepted:** 15 January 2025

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

modeling nonlinear relationships across chemical and biological data. However, the usefulness of AI in drug discovery depends less on algorithmic sophistication alone than on whether models are trained, validated, interpreted, and applied in ways that support credible pharmaceutical decisions.

AI has been applied across discovery and development workflows, including virtual screening, drug repurposing, molecular design, toxicity prediction, and trial-support functions [2]. In early discovery, these applications are often positioned as methods for accelerating prioritization rather than replacing experimental biology or medicinal chemistry. This distinction is essential because computational prediction can identify plausible compounds, targets, or mechanisms, but it cannot by itself establish potency, selectivity, pharmacological activity, safety, manufacturability, clinical benefit, or regulatory acceptability.

Drug design in the AI era increasingly depends on how molecular information is represented and connected to chemical reasoning [3]. Fingerprints, descriptors, SMILES strings, molecular graphs, protein sequences, predicted structures, docking-derived features, and multimodal biological profiles each encode different assumptions about chemical similarity, activity relationships, and biological mechanism. Consequently, the same model family may provide useful decision support in one discovery context but weak or misleading signals in another if representation, endpoint, assay context, or validation strategy is poorly aligned with the pharmaceutical question.

End-to-end AI workflows can connect target selection, hit finding, lead optimization, and development-relevant prediction, but such pipelines require experimental integration and disciplined interpretation to avoid overstating computational outputs [4]. Prospective examples, including deep-learning-enabled antibiotic screening, demonstrate that AI can prioritize candidates for experimental testing, yet they also show that biological validation remains the key step separating in silico prioritization from drug discovery evidence [5]. This systematic review therefore evaluates AI-enabled drug discovery through a pharmaceutical lens, focusing on molecular models, target prediction, lead optimization, validation quality, and translational barriers.

Review scope and research questions

This systematic review was designed to evaluate peer-reviewed evidence on AI-enabled drug discovery in pharmaceutical sciences using transparent review logic, predefined eligibility criteria, structured extraction fields, and cautious synthesis. The reporting framework follows the principle that systematic reviews should make their search, screening, eligibility, extraction, and synthesis decisions explicit enough for readers to judge reproducibility and interpretive reliability [6]. The review does not present a meta-analysis because the included evidence spans heterogeneous drug discovery tasks, model families, datasets, endpoints, validation designs, and reporting conventions.

The scope covers studies and reviews addressing machine learning, deep learning, graph neural networks, transformer models, generative models, molecular representation learning, drug–target interaction prediction, target prediction, virtual screening, ADMET prediction, toxicity prediction, drug repurposing, explainability, benchmark quality, validation, and translational implementation. Search reporting was planned to preserve database names, search concepts, eligibility decisions, extraction fields, and limitations, consistent with search-transparency principles for systematic reviews [7]. This scope allows the review to evaluate not only what AI models can predict, but also whether their evidence is sufficiently interpretable and validated for pharmaceutical decision-making.

The research questions were focused rather than purely mapping-oriented: how AI-enabled molecular models are used in pharmaceutical drug discovery; which molecular representations and model families dominate the evidence; how AI supports target prediction, drug–target interaction modeling, and pathway-level inference; how AI contributes to virtual screening, de novo generation, ADMET prediction, repurposing, and lead optimization; and what validation and translational barriers limit pharmaceutical use. The distinction between systematic and scoping review logic is important because this article seeks to answer defined evidence questions rather than merely catalogue a broad research landscape [8].

The review population is the AI-enabled drug discovery literature relevant to pharmaceutical sciences, medicinal chemistry, pharmacology, cheminformatics, bioinformatics, computational biology, translational medicine, and biomedical informatics. The synthesis treats model performance claims cautiously and separates retrospective benchmarking, external validation, prospective experimental testing, preclinical relevance, clinical translation, and regulatory readiness. This approach follows systematic review reporting expectations by making the review's scope, questions, and limitations explicit before presenting model- and task-level synthesis [6].

Search strategy and study selection

A multi-database search strategy was used in the planning stage to improve coverage across pharmaceutical sciences, bioinformatics, cheminformatics, medicinal chemistry,

computational biology, and biomedical AI. The searched platforms included PubMed/MEDLINE, Scopus, Web of Science Core Collection, ScienceDirect, SpringerLink, Wiley Online Library, ACS Publications, Royal Society of Chemistry, Oxford Academic, Nature Portfolio, Cell Press, and supplementary Google Scholar checking for DOI and bibliographic verification. Multi-database searching was selected because relying on a single source can reduce retrieval completeness in systematic reviews [9].

Search strings combined AI and drug discovery concepts with task-specific terms, including machine learning, deep learning, molecular modeling, graph neural network, transformer, generative model, de novo drug design, lead optimization, drug–target interaction prediction, target prediction, virtual screening, ADMET prediction, toxicity prediction, QSAR, drug repurposing, explainable AI, data leakage, validation, translational barriers, regulatory challenges, and pharmaceutical implementation. Records were planned for export in RIS or BibTeX format with title, authors, journal, DOI, year, abstract, and source database fields. Deduplication was planned by DOI first, followed by normalized title matching and author-year-title comparison.

Title and abstract screening evaluated whether each record addressed AI-enabled drug discovery with enough pharmaceutical relevance to support at least one review domain. Full-text assessment then considered whether the article provided extractable information on model type, data source, molecular representation, drug discovery task, validation approach, limitation, and translational relevance. Risk-of-bias and applicability judgments were planned as structured qualitative assessments rather than informal impressions, because transparent appraisal is necessary for interpreting heterogeneous prediction-model evidence [10].

Quality considerations emphasized dataset representativeness, assay heterogeneity, benchmark splitting, leakage control, external validation, prospective or experimental validation, interpretability, reproducibility, and applicability to pharmaceutical decision-making. Prediction-model evidence was interreted with attention to bias and applicability, because high apparent performance may not translate when datasets, target distributions, chemical scaffolds, or biological contexts differ from the training environment [11]. Review quality was also considered in relation to methodological transparency, eligibility justification, extraction consistency, and synthesis logic [12].

The search and selection process was documented as a candidate-reference screening workflow. Database and supplementary searching identified 58 potentially relevant records. After removal of 4 duplicates, 54 records were screened by title and abstract. At this stage, 8 records were excluded because they were outside the article scope, lacked direct AI-enabled drug discovery relevance, or did not meet the peer-reviewed journal article requirement. Forty-six full-text articles were then assessed for eligibility. Eight full-text articles were excluded because they lacked sufficient extractable evidence on model type, discovery task, validation approach, translational barrier, or pharmaceutical relevance. The final included evidence set comprised 38 peer-reviewed journal articles with DOI information, which were used for structured extraction, citation planning, and narrative systematic synthesis.

Table 1 summarises the systematic review search, eligibility, screening, extraction, and synthesis protocol used to evaluate AI-enabled drug discovery evidence.

Table 1. Systematic Review Protocol for AI-Enabled Drug Discovery in Pharmaceutical Sciences: Databases, Search Terms, Eligibility Criteria, Screening Logic, Data Extraction Fields, Quality Considerations, Evidence Synthesis Domains, and Reporting Safeguards

Protocol component	Operational definition	Implementation rule	Data field required	Quality-control action	Risk of bias or limitation	Reporting requirement
Database selection	Scholarly indexing and publisher platforms covering pharmaceutical sciences, cheminformatics, bioinformatics, medicinal chemistry, and biomedical AI	Search multiple databases rather than relying on one source	Database name and search date	Record database source during export	Incomplete retrieval if database coverage differs	List databases searched and state if exact counts are unavailable
Search concepts	AI methods combined with drug discovery tasks and translational terms	Combine broad AI terms with task-specific terms	Search string and concept group	Preserve search logic for reproducibility	Search terms may miss emerging terminology	Report core search strings without inventing hit counts

Eligibility criteria	Peer-reviewed journal evidence with DOI and direct relevance to AI-enabled drug discovery	Include only articles supporting molecular modeling, target prediction, lead optimization, validation, or translation	Publication type, DOI, topic relevance	Exclude non-peer-reviewed or weakly relevant records	Borderline relevance may require judgment	State inclusion and exclusion criteria clearly
Deduplication	Removal of duplicate records before screening	Use DOI first, then normalized title and author-year-title matching	DOI, title, author, year	Check duplicate clusters manually when ambiguous	Same article may appear under variant metadata	Describe deduplication method
Title and abstract screening	Initial relevance assessment	Screen for AI method, drug discovery task, and pharmaceutical relevance	Abstract, keywords, article type	Apply predefined eligibility logic	Abstracts may omit validation details	Describe screening basis without unsupported counts
Full-text assessment	Detailed eligibility and evidence extraction	Assess model type, representation, task, validation, limitation, and translational relevance	Full text and extraction fields	Exclude articles without extractable evidence	Full text may not report all relevant fields	Report reasons for exclusion at category level
Data extraction	Structured capture of methodological and translational information	Extract model family, input data, task, validation, benchmark, strength, limitation, and research gap	Extraction table fields	Use “not clearly reported” when unavailable	Missing reporting may limit synthesis	Avoid inventing model-performance values
Quality appraisal	Qualitative assessment of bias, applicability, reproducibility, and validation	Evaluate data leakage, splitting, external validation, assay context, interpretability, and reproducibility	Quality and limitation fields	Align appraisal with prediction-model concerns	Appraisal is qualitative, not a pooled risk score	Report appraisal logic transparently
Evidence synthesis	Narrative systematic synthesis organized by review questions and discovery stages	Synthesize by model family, representation, task, validation, and barrier	Synthesis domain	Distinguish prediction from validation and translation	Heterogeneous evidence prevents pooled inference	Do not claim meta-analysis
Reporting safeguards	Protection against overclaiming and unsupported claims	Avoid invented counts, performance values, assay results, clinical outcomes, or regulatory approvals	Claim-source mapping	Use approved citations only	Overinterpretation of benchmark studies	State limitations and interpret findings cautiously

Figure 1 presents the systematic review search, screening, eligibility, and inclusion process used to identify peer-reviewed studies on AI-enabled drug discovery.

ISSN (Online) 2249-6084 (Print) 2250-1029

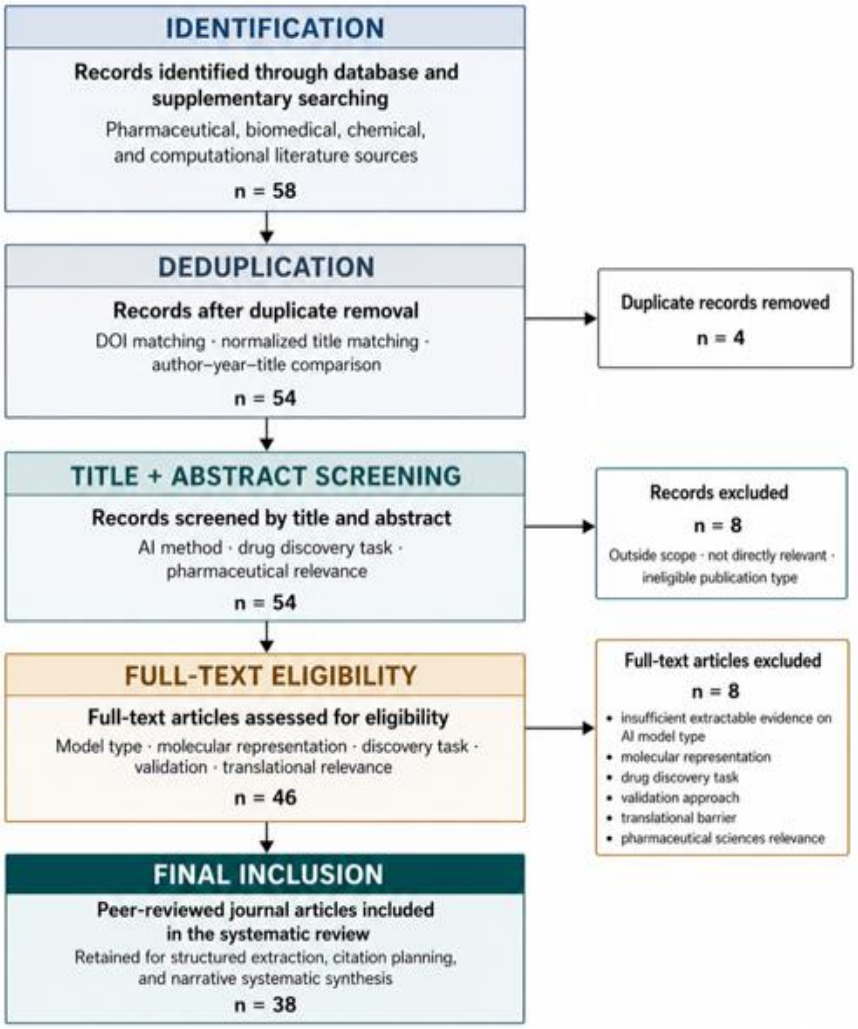

Figure 1. Systematic Review Search and Study Selection Flow for AI-Enabled Drug Discovery Studies

Alt-text description

A systematic review flow diagram showing records identified through database searching, records after duplicate removal, records screened, full-text articles assessed for eligibility, excluded articles with reasons, and final included studies on AI-enabled drug discovery. The diagram must use only verified screening counts from the review process; where verified counts are unavailable, numeric boxes remain unreported rather than estimated.

AI models in drug discovery

AI models in drug discovery range from classical machine learning systems using molecular descriptors or fingerprints to deep learning architectures that learn representations from molecular graphs, sequences, structures, and multimodal biological data. Benchmark resources have played a major role in standardizing molecular machine learning by defining datasets, endpoint categories, splitting strategies, and evaluation tasks for molecular property prediction [13]. These benchmarks are useful for comparing methods, but they should not be interpreted as direct evidence of clinical or pharmaceutical success unless supported by external and experimental validation.

Graph-based and learned molecular representations have become especially important because they allow models to encode atoms, bonds, neighborhoods, and higher-order molecular structure in ways that can support property prediction and lead-optimization tasks. Learned representations can improve predictive performance compared with fixed descriptors in some contexts, yet their value depends on endpoint quality, chemical-space coverage, data splitting, and the alignment between the prediction task and the downstream pharmaceutical decision [14]. For systematic synthesis, graph neural networks are therefore best interpreted as representation-learning tools rather than universally superior discovery engines.

Sequence-based molecular modeling provides another major route for AI-enabled drug discovery. SMILES-based methods can learn continuous molecular descriptors by translating equivalent chemical representations, supporting downstream prediction tasks while also raising questions

about chemical validity, representation invariance, and transferability across chemical spaces [15]. These methods are particularly relevant where large molecular string datasets are available, but they require careful validation because small syntactic changes in representation may not correspond cleanly to biological or pharmacological meaning.

Multitask learning, transformer models, and generative models extend AI-enabled drug discovery beyond single-endpoint prediction. Multitask deep featurization has been applied to ADMET prediction, supporting lead optimization by learning across related endpoints while remaining dependent on data quality and endpoint definition [16]. Transformer architectures have shown that sequence models can represent chemical reactions, which is relevant to synthesis-aware drug design but does not by itself establish therapeutic value [17]. Generative molecular design expands de novo compound proposal, but the reviewed evidence indicates that generated molecules require property constraints, synthesis feasibility, biological plausibility, and experimental confirmation before translational relevance can be inferred [18].

Table 2 compares major AI model families used in drug discovery by molecular representation, discovery task, validation approach, and interpretability challenge.

Table 2. AI Model Families in Drug Discovery: Molecular Representations, Learning Approaches, Discovery Tasks, Validation Strategies, Strengths, Limitations, Interpretability Challenges, and Pharmaceutical Use Cases

AI model family	Typical molecular representation	Drug discovery task	Learning approach	Validation strategy	Main strength	Main limitation	Interpretability concern	Pharmaceutical relevance
Classical machine learning	Molecular fingerprints, physicochemical descriptors, assay-derived features	QSAR, property prediction, virtual screening, toxicity estimation	Supervised classification or regression	Cross-validation, external test sets, scaffold-aware splits where available	Transparent baseline performance and lower computational complexity	Performance depends heavily on feature engineering and dataset quality	Feature importance may not map directly to mechanism	Useful for interpretable prioritization and benchmark comparison
Deep neural networks	Molecular descriptors, fingerprints, learned embeddings, multimodal inputs	Bioactivity prediction, ADMET prediction, repurposing, prioritization	Supervised deep learning	Retrospective benchmarks, external validation, prospective testing when available	Captures nonlinear relationships across large datasets	Requires sufficient, representative, well-curated data	Internal representations may be difficult to explain	Supports large-scale screening and multi-endpoint modeling
Graph neural networks	Atom-bond molecular graphs, graph embeddings, message-passing representations	Molecular property prediction, DTI prediction, lead optimization	Representation learning on molecular graphs	Benchmark datasets, scaffold splitting, external testing	Encodes molecular topology directly	Can overfit benchmark patterns if validation is weak	Learned substructures may not equal causal pharmacophores	Relevant to molecular modeling and structure-aware prioritization
SMILES-based sequence models	SMILES strings, tokenized molecular sequences, learned latent vectors	Property prediction, generative design, reaction prediction	Recurrent, convolutional, or transformer-based sequence learning	Retrospective sequence benchmarks and validity checks	Uses large text-like molecular corpora	Sensitive to syntax, augmentation, and representation choices	Token attention may be chemically ambiguous	Useful for scalable molecular learning and synthesis-aware modeling
Transformer models	Chemical sequences, reaction SMILES, protein sequences, multimodal tokens	Reaction prediction, molecular generation, protein-aware modeling	Attention-based sequence modeling	Benchmark evaluation, uncertainty estimation where available	Captures long-range dependencies in symbolic chemical data	Requires large training corpora and careful domain adaptation	Attention weights may not provide sufficient explanation	Relevant to synthesis planning and foundation-style molecular modeling

Generative models	Latent molecular vectors, graphs, SMILES, constrained design spaces	De novo molecular design and lead optimization	Variational, adversarial, reinforcement, or diffusion-like generation	Validity, novelty, property filtering, synthesis and assay testing where available	Expands chemical search and proposes new candidates	Generated molecules may be unrealistic, unsynthesizable, or biologically inactive	Optimization objectives may obscure medicinal chemistry rationale	Supports hypothesis generation, not standalone lead validation
Multitask and transfer learning models	Shared molecular embeddings across endpoints or datasets	ADMET, toxicity, bioactivity, data-scarce prediction	Shared representation learning and model adaptation	Multi-endpoint benchmarks and external validation	Can use related endpoints to improve sparse tasks	Negative transfer and endpoint inconsistency are possible	Shared features may be difficult to attribute to specific endpoints	Useful in lead optimization where many properties must be balanced
Explainable AI approaches	Model explanations, feature attributions, substructure importance, uncertainty outputs	Model interpretation and decision support	Post hoc or intrinsically interpretable methods	Expert review, plausibility checks, stability analysis	Improves trust and scientific interrogation	Explanations can be unstable or non-causal	Explanation may not equal mechanism	Important for medicinal chemistry, preclinical confidence, and regulatory communication

Target prediction and lead optimization

AI-enabled target prediction and drug–target interaction modeling occupy a central position between molecular modeling and pharmaceutical decision-making. DeepDTA demonstrated that convolutional neural networks can use drug and protein sequence representations to predict drug–target binding affinity, illustrating how AI can convert molecular and protein sequence inputs into prioritization signals for downstream discovery [19]. Such outputs are best understood as computational estimates of interaction likelihood or affinity rather than evidence of target validation, because biological relevance depends on assay confirmation, pathway context, selectivity, pharmacology, and disease plausibility.

Graph- and sequence-based compound–protein interaction models extend this logic by combining molecular graph representations with protein sequence information. End-to-end neural learning of compounds and proteins supports interaction prediction while reducing dependence on manually engineered features, but model performance remains shaped by dataset composition, chemical redundancy, and protein-family coverage [20]. GraphDTA further shows that graph neural networks can encode molecular topology for drug–target affinity prediction, strengthening representation-level modeling while still requiring external validation beyond retrospective benchmark performance [21].

Reusable drug–target interaction frameworks can improve reproducibility by making architectures, encoders, datasets, and evaluation workflows more explicit. DeepPurpose illustrates how standardized libraries may support systematic comparison of DTI models and drug repurposing hypotheses when input representations, training data, and evaluation protocols are reported transparently [22]. In lead optimization, however, DTI predictions are only one component of a broader medicinal chemistry decision process that also includes potency, selectivity, physicochemical properties, synthetic accessibility, ADMET behavior, formulation constraints, and biological mechanism.

Generative AI and experimental coupling provide a more translationally informative route when candidate molecules are not only proposed computationally but also synthesized and tested. Deep-learning-enabled identification of DDR1 kinase inhibitors shows that AI-generated or AI-prioritized designs can move into biochemical validation, but such evidence remains early-stage unless supported by broader pharmacological and preclinical assessment [23]. Combining generative design with on-chip synthesis further demonstrates how AI can support iterative lead discovery, although the resulting compounds still require assay validation, optimization, and biological interpretation before pharmaceutical relevance can be claimed [24].

Table 3 maps AI-enabled target prediction and lead optimization tasks across data sources, model outputs, validation needs, and translational relevance.

Table 3. AI-Enabled Target Prediction and Lead Optimization in Pharmaceutical Drug Discovery: Data Sources, Model Outputs, Biological Interpretation, ADMET Relevance, Validation Needs, Translational Utility, and Failure Risks

Discovery task	Input data source	AI model output	Biological or chemical interpretation	Lead optimization function	ADMET relevance	Validation requirement	Translational risk	Research implication
Drug–target interaction prediction	Compound structures, protein sequences, binding databases	Predicted interaction or affinity	Suggests possible compound–protein association	Prioritizes compounds and targets for testing	Indirect unless combined with safety and exposure data	External testing, assay confirmation, negative controls	Interaction prediction may be mistaken for target validation	Use DTI models as hypothesis generators
Target prediction	Bioactivity profiles, chemical similarity, omics data, target annotations	Candidate targets or target classes	Suggests plausible biological mechanisms	Supports target prioritization and mechanism exploration	Relevant when linked to safety liabilities or off-target effects	Orthogonal biological assays and pathway validation	Dataset bias may overrepresent well-studied targets	Expand target coverage and evaluate disease relevance
Virtual screening	Chemical libraries, molecular descriptors, docking features, target structures	Ranked candidate molecules	Identifies compounds predicted to bind or show activity	Narrows experimental screening sets	Limited unless ADMET filters are integrated	Prospective screening and hit confirmation	Ranking may reflect benchmark or docking artifacts	Combine structure, ligand, and assay validation
De novo molecular generation	Molecular graphs, SMILES, latent vectors, target constraints	Novel candidate structures	Proposes molecules satisfying design objectives	Expands chemical search space	Can include ADMET constraints if modeled explicitly	Validity, synthesis, property, and bioassay testing	Generated molecules may be invalid, inactive, or impractical	Integrate synthesis feasibility and biological assays
QSAR modeling	Chemical descriptors, fingerprints, bioactivity endpoints	Activity or property prediction	Relates structure to measured endpoint	Guides analogue prioritization	Can support toxicity or pharmacokinetic endpoints	Scaffold-aware validation and external test sets	Overfitting and chemical-series bias	Report applicability domain and uncertainty
ADMET prediction	Molecular structures, assay datasets, pharmacokinetic and toxicity endpoints	Predicted absorption, distribution, metabolism, excretion, or toxicity property	Estimates development-relevant liabilities	Filters or prioritizes leads with better property balance	Directly relevant	Endpoint-specific validation and assay context review	Poor assay harmonization may mislead optimization	Build multitask and externally validated ADMET models
AI-assisted docking	Target structures, ligand libraries, docking poses, scoring features	Docking-prioritized or ML-rescored compounds	Suggests potential binding hypotheses	Supports structure-based prioritization	Indirect unless linked to developability filters	Docking review, experimental binding assays, functional testing	Docking score may be overinterpreted as potency	Use AI docking as prioritization, not proof
Drug repurposing	DTI data, transcriptomics, disease signatures, known drug profiles	Candidate drug–disease or drug–target links	Suggests new therapeutic hypotheses	May accelerate repositioning decisions	Relevant for known drugs with existing safety data	Disease-context validation and clinical plausibility review	Association may not translate to efficacy	Combine computational signals with biological rationale
Lead optimization decision support	Multimodal chemical, biological, and ADMET data	Multi-objective prioritization recommendation	Balances potency, selectivity, safety, and developability	Supports medicinal chemistry iteration	Highly relevant when ADMET endpoints are included	Prospective optimization cycles and experimental feedback	Single-objective optimization can worsen other properties	Use multi-objective, uncertainty-aware workflows

ISSN (Online) 2249-6084 (Print) 2250-1029

Translational barriers and validation challenges

The most persistent translational barrier in AI-enabled drug discovery is that retrospective benchmarks can reward memorization rather than generalization. Ligand-based classification benchmarks may contain chemical redundancy or similarity patterns that allow models to perform well without learning robust, transferable structure–activity relationships [25]. This issue is especially important for pharmaceutical interpretation because a model that succeeds on a random split may fail when applied to new scaffolds, unfamiliar target families, different assay technologies, or disease-specific biological contexts.

Large-scale target-prediction comparisons show that model evaluation depends strongly on dataset curation, task construction, benchmark design, and endpoint definition. Machine learning comparisons on ChEMBL support systematic model assessment, but they also highlight that bioactivity data are heterogeneous and shaped by assay conditions, target annotation, and chemical-series composition [26]. Consequently, target prediction should be evaluated with attention to applicability domain, scaffold novelty, assay context, and whether predicted associations are biologically actionable.

Interpretability is another major barrier because black-box predictions can be difficult to translate into medicinal chemistry hypotheses, biological mechanisms, or regulatory explanations. Explainable AI can help connect model outputs to substructures, features, uncertainty, or biological rationale, but explanations must be stable, scientifically plausible, and useful for expert decision-making rather than merely visually persuasive [27]. In pharmaceutical settings, interpretability matters not only for trust but also for deciding which compounds to synthesize, which assays to run, and which liabilities require mitigation.

Translational caution is necessary because benchmark accuracy alone is not equivalent to discovery impact. Realistic assessment of AI in drug discovery requires measuring whether a model changes experimental decisions, improves prioritization, reduces unproductive testing, or identifies candidates that withstand biological validation [28]. Chemical and biological data limitations further constrain generalization when assay noise, missing negative data, target imbalance, or inconsistent endpoint definitions distort training signals [29]. Responsible implementation also requires attention to data governance, reproducibility, ethics, workflow integration, and regulatory expectations across the pharmaceutical lifecycle [30].

Evidence synthesis

The reviewed evidence indicates that AI-enabled drug discovery is most mature where computational tasks are well-defined, datasets are sufficiently structured, and validation endpoints are closely aligned with the prediction objective. Protein structure prediction has expanded the structural context available for target assessment and structure-informed discovery, but predicted structures should be viewed as enabling evidence rather than direct proof of ligand efficacy or therapeutic success [31]. This distinction is central to pharmaceutical interpretation because structural plausibility must still be connected to binding, function, selectivity, pharmacology, and disease relevance.

Structure and interaction prediction further strengthen the evidence base by allowing computational models to reason about protein conformations, complexes, and potential interaction contexts. Three-track neural network modeling supports structure and interaction prediction, but the translational meaning of such outputs depends on downstream experimental testing and biological plausibility [32]. In the synthesis framework, structure prediction is therefore positioned as a bridge between target biology and molecular modeling rather than as a replacement for target validation.

Bioactivity databases provide the empirical foundation for many AI-enabled drug discovery models, especially in target prediction, QSAR, DTI prediction, ADMET modeling, and virtual screening. ChEMBL illustrates the value of curated bioassay resources for model development, while also showing why assay metadata, endpoint definitions, target annotations, and deposition quality matter for interpretation [33]. The synthesis therefore treats data resources as active determinants of model reliability rather than neutral containers of training examples.

AI-enhanced virtual screening is a comparatively mature area because it offers a clear operational role: reducing large chemical spaces to prioritized subsets for further computational and experimental evaluation. Deep Docking shows how deep learning can augment structure-based screening at scale, but selected compounds still require docking review, assay confirmation, and medicinal chemistry evaluation before they can be considered meaningful hits [34]. This task maturity arises from the compatibility between AI ranking and established screening workflows, not from any ability of AI to bypass experimental validation.

Ultra-large virtual screening platforms demonstrate how computational infrastructure can expand chemical search far beyond conventional library sizes. Open-source screening workflows can make large-scale prioritization more accessible, yet they also increase the burden of triage, validation, and reproducibility because more candidates

can be generated than can realistically be synthesized or tested [35]. Across the evidence base, the strongest translational position for AI is therefore as a prioritization and hypothesis-generation layer embedded within iterative experimental drug discovery.

Figure 2 illustrates the evidence synthesis pathway linking AI molecular models, target prediction, lead optimization, validation requirements, and translational barriers in pharmaceutical drug discovery.

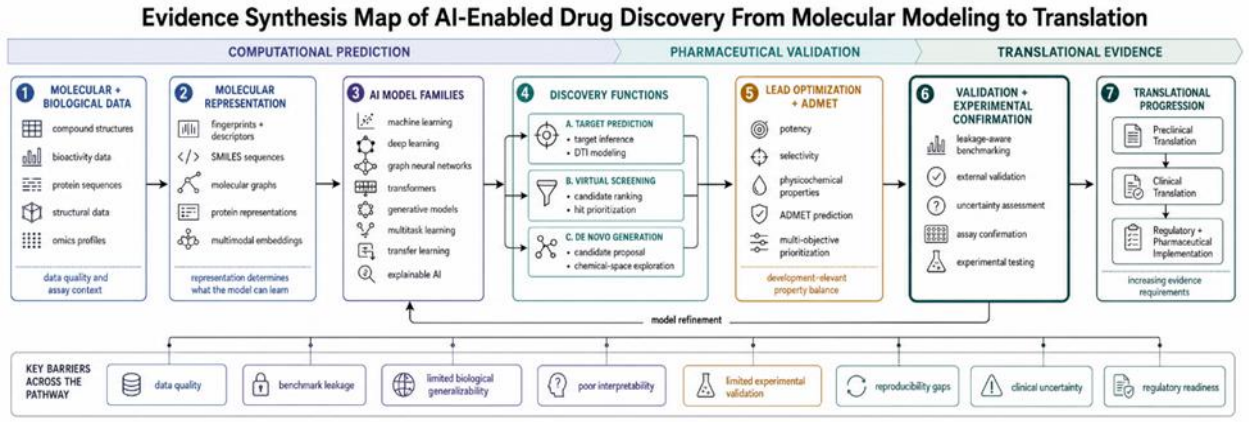

Figure 2. Evidence Synthesis Map of AI-Enabled Drug Discovery From Molecular Modeling to Translation

Alt-text description

A conceptual synthesis map showing AI-enabled drug discovery as a pipeline from molecular and biological data inputs to molecular representation, AI model selection, target prediction, virtual screening, de novo design, lead optimization, ADMET prediction, validation, experimental testing, preclinical translation, clinical translation, and regulatory or implementation barriers.

Research gaps and future directions

A major future direction is the development of chemistry-aware AI systems that support medicinal chemistry reasoning rather than simply optimizing numerical objectives. Future AI drug design needs uncertainty estimates, applicability-domain reporting, synthesis awareness, expert feedback, and decision-oriented validation that shows how model outputs improve real discovery choices [36]. This requires moving from model-centric benchmarking toward workflow-centric evidence in which algorithms are assessed by their contribution to compound prioritization, design iteration, assay planning, and portfolio decisions.

Self-supervised and contrastive molecular representation learning may help address data scarcity by learning transferable representations from large unlabeled molecular corpora. Molecular contrastive learning with graph neural networks supports this direction by improving learned representations, but its pharmaceutical value still depends on external validation, endpoint relevance, and performance in chemically novel or biologically difficult settings [37]. Future work should therefore evaluate pretrained and transfer-learning models under scaffold-aware, time-split, assay-aware, and prospective validation protocols.

Regulatory and implementation readiness will become increasingly important as AI systems move from exploratory modeling toward decisions that influence therapeutic development. Risk-based regulation of AI-enabled therapeutics requires transparent documentation, credible evidence, lifecycle monitoring, and clarity about how models are used within human decision-making systems [38]. For drug discovery, this does not mean every early-stage model requires the same evidentiary standard as a clinical tool, but it does mean that claims about translational relevance should match the model's actual role, validation level, and decision impact.

Future research should also prioritize open and reproducible workflows, benchmark audits, shared negative data, uncertainty-aware predictions, explainable model outputs, multimodal biological integration, and stronger links between computational predictions and wet-lab validation. These directions are especially important for generative design, DTI prediction, ADMET modeling, repurposing, and structure-based prioritization because these tasks can easily produce plausible outputs that remain biologically unconfirmed. The next stage of pharmaceutical AI should therefore be judged not by model novelty alone, but by reproducibility, interpretability, experimental utility, and responsible implementation.

CONCLUSION

This systematic review synthesizes AI-enabled drug discovery in pharmaceutical sciences across molecular

models, target prediction, lead optimization, validation strategies, and translational barriers. The evidence indicates that AI can support drug discovery by improving hypothesis generation, molecular prioritization, target inference, virtual screening, de novo design, ADMET estimation, and decision support. However, computational prediction is not equivalent to drug discovery success. Reliable pharmaceutical translation depends on data quality, leakage-aware validation, biological plausibility, interpretability, reproducibility, uncertainty assessment, experimental confirmation, medicinal chemistry integration, preclinical evidence, clinical relevance, and regulatory readiness. AI should therefore be understood as a powerful complement to pharmaceutical science rather than a substitute for experimental biology, pharmacology, toxicology, clinical development, or regulatory evaluation.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] 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):463-77. doi:10.1038/s41573-019-0024-5

[2] 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

[3] Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA Jr, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-64. doi:10.1038/s41573-019-0050-3

[4] Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, et al. Exploiting machine learning for end-to-end drug discovery and development. Nat Mater. 2019;18(5):435-41. doi:10.1038/s41563-019-0338-z

[5] 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

[6] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71

[7] Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10(1):39. doi:10.1186/s13643-020-01542-z

[8] Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol. 2018;18(1):143. doi:10.1186/s12874-018-0611-x

[9] Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev. 2017;6(1):245. doi:10.1186/s13643-017-0644-y

[10] McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods. 2021;12(1):55-61. doi:10.1002/jrsm.1411

[11] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-8. doi:10.7326/M18-1376

[12] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008. doi:10.1136/bmj.j4008

[13] 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):513-30. doi:10.1039/C7SC02664A

[14] 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):3370-88. doi:10.1021/acs.jcim.9b00237

[15] Winter R, Montanari F, Noé F, Clevert DA. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. Chem Sci. 2019;10(6):1692-701. doi:10.1039/C8SC04175J

[16] Feinberg EN, Joshi E, Pande VS, Cheng AC. Improvement in ADMET prediction with multitask deep featurization. J Med Chem. 2020;63(16):8835-48. doi:10.1021/acs.jmedchem.9b02187

[17] Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction.

ACS Cent Sci. 2019;5(9):1572-83. doi:10.1021/acscentsci.9b00576

[18] Zeng X, Wang F, Luo Y, Kang SG, Tang J, Lightstone FC, et al. Deep generative molecular design reshapes drug discovery. Cell Rep Med. 2022;3(12):100794. doi:10.1016/j.xcrm.2022.100794

[19] Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics. 2018;34(17). doi:10.1093/bioinformatics/bty593

[20] Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics. 2019;35(2):309-18. doi:10.1093/bioinformatics/bty535

[21] Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: predicting drug-target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140-7. doi:10.1093/bioinformatics/btaa921

[22] Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: a deep learning library for drug-target interaction prediction. Bioinformatics. 2020;36(22-23):5545-7. doi:10.1093/bioinformatics/btaa1005

[23] 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):1038-40. doi:10.1038/s41587-019-0224-x

[24] Grisoni F, Huisman BJH, Button AL, Moret M, Atz K, Merk D, et al. Combining generative artificial intelligence and on-chip synthesis for de novo drug design. Sci Adv. 2021;7(24). doi:10.1126/sciadv.abg3338

[25] Wallach I, Heifets A. Most ligand-based classification benchmarks reward memorization rather than generalization. J Chem Inf Model. 2018;58(5):916-32. doi:10.1021/acs.jcim.7b00403

[26] Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chem Sci. 2018;9(24):5441-51. doi:10.1039/C8SC00148K

[27] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4

[28] 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):511-24. doi:10.1016/j.drudis.2020.12.009

[29] 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. Drug Discov Today. 2021;26(4):1040-52. doi:10.1016/j.drudis.2020.11.037

[30] Blanco-González A, Cabezón A, Seco-González A, Conde-Torres D, Antelo-Riveiro P, Piñeiro Á, et al. The role of AI in drug discovery: Challenges, opportunities, and strategies. Pharmaceuticals (Basel). 2023;16(6):891. doi:10.3390/ph16060891

[31] 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):583-9. doi:10.1038/s41586-021-03819-2

[32] Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373(6557):871-6. doi:10.1126/science.abj8754

[33] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: Towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1). doi:10.1093/nar/gky1075

[34] Gentile F, Agrawal V, Hsing M, Ton AT, Ban F, Norinder U, et al. Deep Docking: a deep learning platform for augmentation of structure based drug discovery. ACS Cent Sci. 2020;6(6):939-49. doi:10.1021/acscentsci.0c00229

[35] Gorgulla C, Boeszoermenyi A, Wang ZF, Fischer PD, Coote PW, Das KMP, et al. An open-source drug discovery platform enables ultra-large virtual screens. Nature. 2020;580(7805):663-8. doi:10.1038/s41586-020-2117-z

[36] Brown N, Ertl P, Lewis R, Luksch T, Reker D, Schneider N. Artificial intelligence in chemistry and drug design. J Comput Aided Mol Des. 2020;34(7):709-15. doi:10.1007/s10822-020-00317-x

[37] Wang Y, Wang J, Cao Z, Barati Farimani A. Molecular contrastive learning of representations via graph neural networks. Nat Mach Intell. 2022;4(3):279-87. doi:10.1038/s42256-022-00447-x

[38] Singh R, Paxton M, Auclair J. Regulating the AI-enabled ecosystem for human therapeutics. Commun Med (Lond). 2025;5(1):181. doi:10.1038/s43856-025-00910-x