

# The AI Compass: Guiding Drug Discovery and Development from Molecular Design to Patient Trials

Yash Prajapati<sup>1</sup>, Tisha Patel<sup>2</sup>, Bisajeet Khandai<sup>3</sup>

<sup>1</sup>Pharm.D intern, Department of Pharmacy, Sumandeep Vidyapeeth Deemed to be University, Vadodara

<sup>2,3</sup>Pharm.D student, Department of Pharmacy Practice, Maliba Pharmacy College, Bardoli

Corresponding Author: Yash Prajapati

Date of Submission: 25-01-2026

Date of Acceptance: 05-02-2026

## ABSTRACT

**Objective:** To provide a comprehensive review of the applications of AI in drug discovery and development, highlighting its impact on target identification, lead optimization, preclinical and clinical evaluation, and post-marketing surveillance, and to assess its potential to improve efficiency, reduce costs, and enhance translational success.

**Data Sources:** Relevant literature from peer-reviewed journals, public databases and case studies focusing in AI application in drug discovery.

**Study Selection:** Research describing AI methodologies, including machine learning, deep learning, graph neural networks, and transformer-based architectures, was included. Studies demonstrating AI in target prioritization, chemical space exploration, ADMET prediction, clinical trial support, and manufacturing optimization were prioritized. Case studies illustrating successful real-world translation were also considered.

**Summary of Contents:** AI supports the drug discovery continuum by improving target identification, structure-activity relationship modelling, and early prediction of ADMET and toxicity profiles. In development, AI enhances formulation design, continuous manufacturing, and process analytical technologies. Clinically, applications such as synthetic control arms, digital twins, and AI-assisted regulatory documentation streamline trials and decision-making. Selected case studies demonstrate AI-driven antibiotic discovery, rapid drug repurposing during health emergencies, and first-in-class clinical candidate identification.

**Conclusion:** AI represents a paradigm shift in pharmaceutical research, offering predictive, efficient, and cost-effective strategies. Its integration across discovery, development, and clinical stages accelerates therapeutic innovation, reduces attrition, and improves translational success.

**Keywords:** AI, Drug Discovery, ADMET prediction, Clinical Development

## I. INTRODUCTION

Drug discovery and development is a process that involves the identification, optimization, pre-clinical and clinical studies to extensively test and characterize the new drug molecule for its pharmacological properties and toxicity profile(1). During the process of new drug discovery ventures in different pharmaceutical industries, researchers are more focused towards the development of new molecular entities (NMEs) and novel dosage forms(2). The increasing biological complexity of diseases, coupled with rising research and development (R&D) costs, has placed significant pressure on conventional drug discovery paradigms to deliver safer and more effective therapeutics within feasible timelines(3).

Recent advances in computational sciences have highlighted artificial intelligence (AI) as a transformative approach capable of reshaping the drug discovery pipeline. By enabling data-driven decision-making across molecular, biological, and clinical domains, AI offers new opportunities to overcome longstanding inefficiencies inherent in traditional research-based drug discovery (RBDD)(4).

### 1.1. Challenges of Traditional Research-Based Drug Discovery:

Traditional research-based drug discovery is largely sequential, labor-intensive, and dependent on empirical experimentation(5). The idea for a target can come from a variety of sources including academic and clinical research and from the commercial sector. It may take many years to build up a body of supporting evidence before selecting a target for a costly drug discovery programme(6).

Traditionally this process has demanded significant time and resources, requiring an average of 14-16 years (figure 1) and a cost of US\$2.6 billion to advance a single molecule from

conception to FDA approval. Despite these efforts, the process is marked by high attrition rates, significant adverse effects of modern drugs, and persistent challenges in addressing chronic diseases

such as diabetes and cancers(7),(8). Small biotech companies often struggle to bring innovative therapies to market due to the high cost of clinical trials(9).

### DRUG DISCOVERY AND DEVELOPMENT



Figure 1 The process of Drug Discovery and Development

#### 1.2. The AI Paradigm as a Solution to Drug Discovery Challenges:

The challenges mentioned above demand more efficient methods, where artificial intelligence (AI) and machine learning (ML) offer a promising path towards increased efficiency and success rates in drug development, providing the pharmaceutical industry with a solution with AI/ML implementation to correct limitations while also opening up novel opportunities using new model implementations based on AI parameters(10). Certain modern methodologies such as Graph Neural Network (GNN) and Transformer Architectures have demonstrated remarkable success in learning complex molecular representations and achieving state-of-the-art performance on a variety of crucial drug discovery tasks(11).

#### II. AI IN TARGET IDENTIFICATION AND VALIDATION

The process of identifying the right biological molecules or cellular pathways that can be modulated by drugs to achieve therapeutic benefits, is increasingly important in modern drug discovery. Although innovations in experimental and omic technologies have been growing over the past few decades identifying actionable therapeutic targets remains challenging. The integration of multi-omic data with AI algorithms has recently emerged as a promising approach for target identification (12).

AI prioritizes targets for specific indications by using multi-models that utilize a diverse range of publicly available omic and text data. Omic data encompass genomics, transcriptomics, proteomics, epigenomics, and metabolomics. These data provide information about altered signaling pathways, molecular

interactions, and protein–protein interactions that can serve as additional inputs for target prioritization. Text-based data are retrieved from funding reports, patents, publications, and clinical trials. During target prioritization, multiple target selection criteria such as protein family class, development status, druggability, toxicity, and novelty can be applied to refine the list of AI-driven targets to align with specific research objectives (13).

GNNs are also being employed in drug target discovery. One such approach is EMOGI, a graph convolutional network (GCN) that predicts cancer drug targets(14).

### **2.1 The use of AI-generated synthetic data for target identification:**

'Synthetic data' refers to artificially generated data that mimic real-world patterns and characteristics. In rare diseases or conditions where patient data are limited, AI can generate synthetic data based on existing knowledge and patterns. These synthetic data can then be used to train AI models and identify potential therapeutic targets that may have been overlooked (15).

In some therapeutic areas, particular patient populations may be under-represented in the available datasets, leading to challenges in target identification. AI can generate synthetic data representing these under-represented populations, allowing more comprehensive and inclusive analysis (16).

### **2.2 Target Selection Criteria:**

Apart from experimental methods, a common computational approach to infer causal relationships between targets and diseases is network-based analysis, which involves the construction of biological networks that capture the relationships between different genes, proteins, drugs, and other molecular entities (17). These networks can be used to identify potential targets that might have a causal involvement in a disease based on their centrality and connectivity within the network.

### **2.3 AI-identified targets validated in experiments:**

Target validation using both cell and animal models is crucial to confirm the modulatory effects of the proposed target on disease development. Organoids – 3D cell models derived from either induced pluripotent stem cells (iPSCs) or adult stem cells (ASCs) – have arisen as a promising technique for both disease research and

drug testing by allowing the capture of tissue architecture and cellular microenvironment in vitro (18).

An increasing number of AI-identified targets are being successfully validated. For example, 28 AI-proposed targets for ALS treatment were validated in an ALS-mimicking *Drosophila* model, revealing eight unreported targets whose suppression strongly rescues eye neurodegeneration (19).

Furthermore, the integration of AI with fully automated robotic laboratories offers the potential for high-throughput target validation and screening. Automated experiments, coupled with AI-driven data analysis, can expedite the validation of predicted targets, enabling researchers to assess their therapeutic potential quickly. This combination of AI and automation has the potential to revolutionize the drug discovery process and significantly reduce the time and cost required for target identification and validation(13).

## **III. AI IN LEAD DISCOVERY AND OPTIMIZATION**

A chemical lead is defined as a synthetically stable, feasible, and drug like molecule active in primary and secondary assays with acceptable specificity, affinity and selectivity for the target receptor (20). Lead optimization is the process by which a drug candidate is designed after an initial lead compound is identified. Potential leads are evaluated for a range of properties, including selectivity and binding mechanisms during lead optimization, as the final step in early stage drug discovery. The purpose of lead optimization is to maintain favorable properties in lead compounds, while improving on deficiencies in lead structure.

In order to find potential candidates, virtual screening (VS), which includes both ligand-based (such as QSAR and pharmacophore modeling) and structure-based techniques, is essential. Lead optimization still faces several obstacles despite notable progress, including as precisely taking into consideration receptor flexibility, desolvation effects, and the intrinsic intricacy of ligand-receptor interactions. This stage is being completely transformed by the advent of AI/MLdriven techniques like Generative Therapeutics Design (GTD) and Query-based molecular optimization (QMO), which allow for more effective chemical space exploration, more accurate molecular property prediction, and faster creation of new chemical entities (21).

### 3.1 Virtual Screening:

It is a computational technique for finding new drug candidates that makes use of model-based computer programs. Identifying compounds with characteristics such as high potency and low toxicity that could be utilized to create new drugs is feasible with virtual screening. To predict the properties of biological pharmacological activities, machine learning (ML) algorithms are trained on large datasets of well-known substances and their attributes (22).

AutoDock's Lamarckian genetic algorithm uses pharmacy grids to identify beneficial ligand-binding types (23). DeepDock uses deep learning to improve structure-based drug discovery by optimizing the identification of prospective therapeutic candidates (24),(25). CarsiDock, a deep learning algorithm, uses large-scale pretraining to enable precise docking and screening (26). KarmaDock, allows the high-precision docking of many ligand libraries (24). DynamicBind advances this field by predicting ligand-specific protein-ligand complex structures with a deep equivariant generative model (27).

### 3.2 De Novo Drug Design:

De novo design is a set of computational methods that can be used to design a compound without using a previously known one as a starting point (28). Deep generative models enable de novo molecular design by learning statistical patterns in chemical datasets to generate novel, valid compounds (29).

Recurrent Neural Networks (RNNs) excel in generating new molecular structures by identifying patterns in training data. Notable developments include the first model by Olivecrona et al.(30) and subsequent advancements like DrugEx(31),(32), which emphasizes multi-objective optimization, including toxicity considerations.

Variational Autoencoders (VAEs) function through a dual neural network architecture, comprised of an encoder that maps molecular structures into a latent space, and a decoder that reverses this transformation. Models have been designed to accommodate both 2D and 3D molecular representations (33). The PASITHEA model introduced "deep dreaming" to molecular design (34).

Generative Adversarial Networks (GANs) involves two competing neural networks: a generator and a discriminator. The generator's primary objective is to create molecular structures that the discriminator cannot distinguish from

genuine molecules (28). The first example in generative drug design was ORGANIC (35) and other ones were rapidly developed, such as ATNC (36) and MolGAN(37) which improved the rate of valid molecules generated.

Transformer-based models, originally developed for NLP, are also applied to molecular science. These models typically employ a string based molecular representation. For example, ChemBERTa(38), trained on millions of SMILES strings using masked token prediction, produces rich, label-free molecular embeddings suitable for downstream tasks such as property prediction or generative modelling (29).

### 3.3 Structure-Activity Relationship (SAR) Modeling:

SAR modeling is a critical element of lead optimization that seeks to establish systematic relationships between chemical structure and biological activity. Early SAR approaches were primarily based on incremental structural modifications guided by medicinal chemistry expertise and simple statistical correlations, which limited scalability and predictive power(39). The integration of artificial intelligence (AI) and machine learning (ML) has significantly enhanced SAR modeling by enabling the analysis of large, complex datasets containing chemical, biological, and pharmacological information. ML-based SAR models use molecular descriptors, fingerprints, and physicochemical properties to predict biological activity more accurately than classical quantitative methods(40).

Deep learning approaches have further advanced SAR analysis by capturing nonlinear and hierarchical relationships between molecular features and pharmacological responses. Neural network-based SAR models have shown superior performance in predicting compound potency, selectivity, and off-target effects across diverse chemical spaces(41). Among these, graph neural networks (GNNs) represent a major breakthrough, as they encode molecules as graphs and learn directly from atomic connectivity, eliminating the need for handcrafted descriptors(42).

AI-enabled quantitative SAR (QSAR) modeling also supports multi-objective optimization by simultaneously predicting efficacy, toxicity, solubility, and metabolic stability. This capability is essential in modern drug discovery, where late-stage failures are frequently caused by unfavorable pharmacokinetic or safety profiles rather than insufficient target activity(43). Recent progress has emphasized improving the

interpretability of AI-based SAR models through explainable AI (XAI) techniques. These methods help identify key molecular substructures driving biological activity, thereby supporting rational structure refinement and increasing confidence in AI-guided decision-making(44). Consequently, AI-driven SAR modeling plays a pivotal role in accelerating lead optimization and improving translational success.

#### IV. AI IN PRECLINICAL ASSESSMENT

Adverse pharmacokinetic properties pose a significant threat to human health and environmental safety, representing one of the leading causes of drug development failure. ~40% of preclinical candidate drugs fail due to insufficient ADMET profiles, while nearly 30% of marketed drugs are withdrawn due to unforeseen toxic reactions(45). Early integration of ADMET factors into the evaluation of new chemical entities has been shown to significantly reduce attrition rates in drug discovery(46). Therefore, it is crucial to predict and optimize the ADMET properties of candidate compounds in advance. ADMET

evaluation encompasses the absorption, distribution, metabolism, excretion, and toxicity of drugs, providing a comprehensive assessment of their in vivo behavior and predicting their clinical efficacy and safety.

#### 4.1 Predictive ADMET-Learning Model to Predict ADMET:

The rise of ML/AI has stimulated a new generation of ADMET prediction platforms that can process large libraries of candidate molecules, capturing molecular structure and physicochemical descriptors at scale. These approaches markedly reduce experimental cost and time by screening out compounds with poor developability profiles before labor-intensive assays. By exploiting large historical ADMET datasets, AI models refine prediction accuracy and permit multi-perspective analysis of drug performance under various physiological and environmental conditions, thereby supporting more evidence-based decision-making. Conceptually, an ADMET prediction platform is a multilayered workflow spanning input, computational methods and predictive output (figure 2).

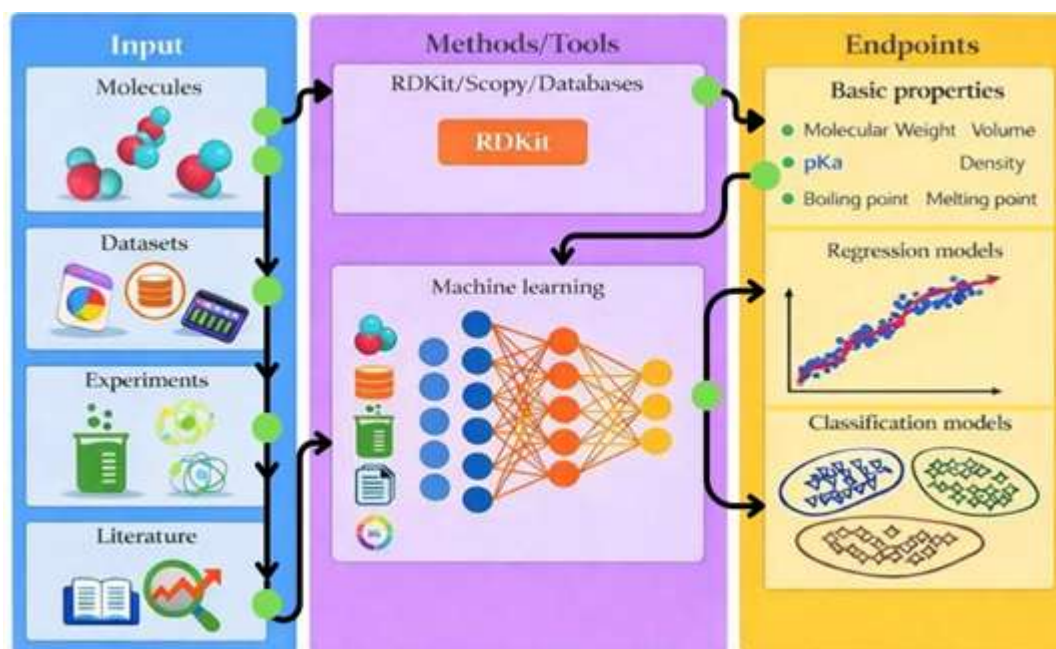


Figure 2 The Basic Framework of ADMET Prediction Platforms.

The input layer aggregates rich chemical structure information (formulas, weights, 2D/3D structures) together with diverse experimental ADMET data, including bioavailability, metabolic stability and clearance under multiple conditions.

Curated literature data further broadens coverage and improves generalizability within the tools/methods layer, cheminformatics modules such as RDKit and Scopy compute core physicochemical properties (molecular weight,

pKa, log P, TPSA, hydrogen-bond donors/acceptors), which then serve as features for ML models. On top of these descriptors, algorithms including SVM, RF, neural networks and gradient-boosting trees learn to predict continuous endpoints (for example  $t_{1/2}$ , VDss, CL, MRT) and classification outcomes such as BBB permeability and HLM stability. Iterative feature selection and hyperparameter optimization enhance performance across both in vitro and in vivo tasks.

**Output component:** the output component represents the final form of the platform's predictions, generally referred to as "endpoints" indicating the predictive values or classification results corresponding to each ADMET characteristic. The number and type of these endpoints vary among platforms. However, comprehensive ADMET prediction platforms can evaluate over 100 different endpoints, encompassing both in vitro properties (such as HLM stability and PPB) and in vivo pharmacokinetic parameters (such as bioavailability (F), half-life ( $t_{1/2}$ ), and renal clearance (CL<sub>r</sub>)). These results are presented through intuitive data visualization and reporting tools, enabling researchers to swiftly acquire both in vitro and in vivo drug ADMET characteristics to support informed decision-making. The platform outputs extend beyond single predictive values and incorporate physicochemical properties with predictive outcomes to offer deeper analyses, such as directions for drug optimization, potential side effects, and optimal routes of administration(47). Additionally, some platforms facilitate comparisons between candidate drugs and other compounds, assisting developers in drug screening and optimization processes(48).

#### 4.2 Prediction of Toxicity Endpoints:

Nowadays, numerous computational tools have been developed to predict various ADMET-related properties, ranging from broad spectrum platforms to tools specialized in specific aspects. Broad-spectrum platforms, such as admetSAR 3.0(49), ADMETlab 3.0(50), etc provide comprehensive coverage across all five ADMET dimensions. By integrating multiple predictive models, these platforms offer systematic assessments of compounds in terms of ADMET. In contrast, platforms like Swiss ADME(48), and FAF-Drugs4.0(51), focus specifically on pharmacokinetic properties, emphasizing predictions related to absorption, distribution, metabolism, and excretion. Additionally, certain platforms concentrate solely on toxicity prediction;

e.g. ProTox 3.0(52) and VenomPred 2.0(53) are tailored for evaluating toxicity endpoints such as hepatotoxicity, carcinogenicity, and mutagenicity. Furthermore, tools like BioTransformer 3.0(54), XenoSite(55), etc. specifically address interactions with cytochrome P450 enzymes (CYPs), playing critical roles in drug metabolism studies.

### V. AI IN FORMULATION, MANUFACTURING AND DELIVERY

Pharmaceutical sciences feature advanced formulations like solid dispersions, extrudates, pellets, nanoparticles, and liposomes alongside standard dosage forms. These formulation techniques enhance functionality in tablets and address API challenges such as low solubility, stability, bioavailability, and manufacturability. AI applications in these techniques warrant investigation for next-generation drug products with optimized critical quality attributes (CQAs) and superior efficacy (56).

#### 5.1 Formulation Optimization-Use of AI to Predict Critical Quality Attributes:

AI streamlines formulation optimization by predicting critical quality attributes (CQAs) like solubility, stability, and dissolution through drug-excipient interaction modeling(57),(58). ML models (random forest, SVMs, DNNs) screen excipient-API combinations, predict compatibility, and optimize parameters pre-lab, replacing trial-and-error DoE. Systematic use of molecular/historical data via RF regressors/SVMs predicts solubility, dissolution, hygroscopicity; enables rapid prototyping, cuts costs(57). Reinforcement learning algorithms are being explored for optimal decisions to dynamically update developmental strategies in response to real-time data, further supporting the concept of adaptive formulation design(59). Another hopeful application is in silico-DoE, where AI-driven tools help reduce the need for high-throughput experimental runs by generating virtual screening environments(60).

AI in formulation development workflows improves speed, precision, and accuracy, aligning with the regulatory principles such as quality by design (QbD)(61). AI also helps in the determination of the encapsulation efficiency and the drug release kinetics in advanced formulation platforms such as liposomes, lipid nanoparticles, solid colloidal dispersions, and other controlled-release drug delivery systems(62). By analyzing large datasets, AI models can provide information

on key physicochemical and biopharmaceutical characteristics such as solubility, rate of dissolution, permeability, and chemical stability at various temperatures and humidity. These datasets are not limited to molecular descriptors, but also formulation parameters and biological responses. Modeling of these properties plays a vital role in determining bioavailability, absorption kinetics, and therapeutic efficacy(63).

### 5.2 AI's Role in Continuous Manufacturing and Process Analytical Technology:

Smart manufacturing transforms pharmaceutical production lines into AI-driven smart factories, where machines autonomously adjust processes using live sensor data, robotics, and IoT devices for enhanced efficiency, precision, and scalability. AI-driven robotics automates repetitive, risky tasks like packaging, labeling, and material handling; these fatigue-free robots operate 24/7 with consistent output and minimal errors, safely managing toxic substances without contamination risks(64). In smart manufacturing, AI enables real-time adjustments to parameters such as temperature, pressure, or mixing speed based on sensor feedback, ensuring batch consistency. A Novartis case study from 2020 showed AI-enabled automation at its site yielding 15% higher production efficiency and 20% lower operating costs. AI integrates with emerging technologies like 3D printing by optimizing designs, predicting parameters, and enabling personalized drug release profiles(65),(66),(67).

AI/ML models control parameters via real-time sensor data, detecting variances for continuous adjustments. This ensures quality consistency, cuts recalls, batch failures, and human errors, the biggest cost driver in pharmaceutical production(68). AI predicts equipment failures for maintenance, preventing downtime and product loss. Quality systems detect early defects; AI-based methods can actually reduce up to 25% waste compared to traditional methods if subpar batches are not allowed to proceed to production. AI optimizes raw materials/labor by calculating exact usage and forecasting demand, avoiding waste/overproduction. This minimizes supply chain costs efficiently(69).

The collaboration of PAT, process data science, and AI establishes robust monitoring for continuous manufacturing lines, yielding benefits like decreased production expenses through reduced final product testing and application of PAT principles. This enables quicker product release, lower inventory, and faster cost recovery-

most valuable for companies manufacturing products long-term, with leadership collaborating regulators for shareholder and patient gains. Beneficial AI tools include expert systems, fuzzy logic, neural networks (ANNs), genetic algorithms, and Model Predictive Control (MPC) for emulating human decision-making in process control, optimization, monitoring, prediction, chemometrics, scale-up/down, and soft sensors. Though few real-time PAT apps exist, ANNs optimize API synthesis(70),(71),(72). Information from data-rich PAT tools such as in-line microscopic images has been extracted using ANNs. When used for contamination classification, a ResNet CNN demonstrated >98% accuracy in classifying crystals found in PVM images(73). Additionally, CNN-based in-line image analysis could be used to measure the particle size distribution and predict the crystal growth rate.

## VI. AI IN CLINICAL DEVELOPMENT AND POST-MARKETING

The clinical phase represents the most capital-intensive and high-risk stage of drug development. AI is shifting this paradigm from empirical execution to predictive precision, addressing critical bottlenecks in trial design, regulatory documentation, and safety surveillance(74).

### 6.1 Clinical Trial Optimization: Synthetic Arms and Digital Twins:

The traditional randomized controlled trial (RCT) model is often hindered by the ethical and logistical challenges of recruiting control groups, particularly for rare diseases. AI is addressing this through Synthetic Control Arms (SCAs).

**Synthetic Data Generation:** Recent studies demonstrate that generative AI can create synthetic external control arms using historical trial data and real-world evidence. A 2025 study validated a "reversible data generalization" technique that allows synthetic data to serve as a statistical proxy for empirical control arms in single-arm trials, effectively reducing the need for placebo cohorts while preserving patient privacy(75).

**Digital Twins:** Beyond static controls, AI is enabling Digital Twins—virtual replicas of patients or organs that integrate multi-omics and real-time sensor data. These models allow for in silico clinical trials where dosage regimens and physiological responses are simulated before human administration. This approach allows researchers to predict individual treatment

outcomes and adverse events with high granularity, optimizing inclusion/exclusion criteria to prevent late-stage trial failures(76).

### 6.2 Generative AI in Regulatory Documentation:

The regulatory landscape requires the synthesis of massive datasets into coherent Clinical Study Reports (CSRs). Generative AI (GenAI) has moved from proof-of-concept to practical application in automating these expert-intensive tasks.

**Automated Drafting:** Large Language Models (LLMs) are now deployed to draft patient safety narratives and regulatory documents by synthesizing raw clinical data (Tables, Listings, and Figures). A 2025 mini-review highlights that these tools can reduce drafting time for CSRs by 30–85% while maintaining terminological consistency(77).

**Human-in-the-Loop Verification:** Despite the efficiency gains, current reviews emphasize the necessity of a "human-in-the-loop" framework to mitigate "hallucinations" (factually incorrect generation) and ensure that AI-generated summaries align with rigorous medical standards before submission(77),(78).

### 6.3 Post-Marketing Surveillance and Pharmacovigilance:

Once a drug enters the market, AI transforms pharmacovigilance (PV) from reactive reporting to proactive signal detection.

**Automated Signal Detection:** Modern PV systems utilize Natural Language Processing (NLP) to automate the extraction of adverse drug reactions (ADRs) from unstructured text, such as medical literature and electronic health records.

**Causality Assessment:** Advanced implementations now include expert-defined Bayesian networks that assist in causality assessment. These probabilistic models can process complex safety signals to distinguish between drug-induced events and background disease progression, significantly reducing the manual burden on safety physicians and enabling faster regulatory decision-making(79).

## VII. PRACTICAL CASE STUDIES OF AI IN DRUG DISCOVERY

### 7.1 Case Study: AI-Driven Discovery and Repurposing of Halicin

El Belghiti et al. (2025) conducted a study of halicin, an antibiotic discovered by deep learning-based artificial intelligence screening, as a

proof-of-concept for AI-driven drug repurposing in the post-antibiotic era. This compound was first identified using a deep neural network, trained on large chemical–biological datasets, to predict antibacterial activity beyond conventional structure–activity relationships. To expand the power of this AI discovery, the authors experimentally validated halicin’s antibacterial activity against multidrug-resistant (MDR) ESKAPE pathogens through CLSI-guided in vitro assays. Halicin shows sustained inhibition of *Escherichia coli*, *Staphylococcus aureus*, *Acinetobacter baumannii*, *Enterobacter cloacae*, and *Klebsiella pneumoniae* with MIC values ranging from 16 to 64 µg/mL, while *Pseudomonas aeruginosa* remained intrinsically resistant. Scanning electron microscopy corroborated the AI-predicted mechanism by demonstrating membrane disruption consistent with proton motive force collapse. This study exemplifies how AI-enabled deep learning platforms can successfully identify non-obvious antimicrobial candidates and accelerate translational validation, highlighting artificial intelligence as a transformative tool for combating antimicrobial resistance and revitalizing antibiotic discovery pipelines(80).

### 7.2 Case Study: AI-Assisted Repurposing of Baricitinib for COVID-19

Richardson et al. (2022) outlined the AI-assisted repurposing of baricitinib as a landmark case of artificial intelligence powering rapid therapeutic discovery during a global health emergency. In their paper, the authors used BenevolentAI’s knowledge graph-based AI platform to consolidate the large-scale biomedical literature, in addition to protein–protein interaction networks, and host–virus biology, to identify approved drugs capable of simultaneously modulating hyperinflammation and viral entry in COVID-19. AI-based analysis predicted baricitinib as a dual-action candidate on the basis of its inhibition of Janus kinase (JAK1/2)–mediated cytokine signaling and numb-associated kinases (AAK1 and GAK) involved in clathrin-mediated endocytosis (a key route of SARS-CoV-2 cell entry). These AI-based hypotheses were then validated using mechanistic studies and large randomized clinical trials and led to substantial reductions in mortality in hospitalized patients with COVID-19. The results of the current study provide robust clinical evidence of the ability of AI-mediated knowledge graph models to convert in silico predictions into regimens that could transform guidelines, and positions AI as a

powerful and rapidly scalable tool for both drug repurposing and pandemic preparedness(81).

### 7.3 Case Study: AI-Facilitated Discovery of Ebola Virus Inhibitors

Kwofie et al. (2023) presents an integrated view of how artificial intelligence, machine learning, and big data analytics can be applied towards the drug discovery of Ebola virus, focusing on AI as a key facilitator of data-based antiviral drug development. This work systematically examined the use of supervised learning algorithms, such as Bayesian classifiers, support vector machines, random forest models, and artificial neural networks to predict small molecules for inhibitors against important Ebola viral proteins (VP24, VP35, VP40, glycoproteins). With respect to bioassay datasets (selected from PubChem, ChEMBL, DrugRepV, BindingDB), AI models performed very well at prediction, which facilitated the prioritization of candidate compounds that were further validated using in vitro assays. Of greatest importance, the authors stressed that deep neural networks had a better ability to accommodate high-dimensional biological data and to describe nonlinear drug–target interactions than traditional machine-learning methods. This research highlights AI-based predictive modeling and big-data integration as key in the acceleration of antiviral drug discovery in a cost, time, and attrition efficient manner with the implication for reinforcing AI as a leading technology for future infectious disease therapeutics(82).

### 7.4 Case Study: AI-Driven Discovery of a 5HT<sub>1A</sub> Agonist for OCD

Imai et al. In 2021, a study reported a novel end-to-end artificial intelligence–based drug discovery platform, based on its comprehensive mechanism, that allowed DSP-1181, a novel 5-HT<sub>1A</sub> receptor agonist for obsessive–compulsive disorder (OCD), to be quickly identified. The study successfully tied AI-backed de novo molecular design, optogenetic technology, and translational biomarkers to enhance central nervous system (CNS) drug discovery. An AI platform developed by Exscientia was used to automate the generation and optimization of virtual compounds reflecting the target affinity, pharmacokinetic properties, and brain permeability expected, supporting cycles of two-week design–synthesis–test iterations. This AI-driven process minimized the discovery time frame to less than 12 months, well short of the industry average of four to five years. The hypothesized

compounds were validated by optogenetic manipulation of OCD-relevant neural circuits in animal models and translational biomarker analysis, utilizing pupillary response, which represents a proxy for central 5-HT<sub>1A</sub> receptor involvement. This research led to the successful implementation of first-in-human clinical trials, showcasing how the integration of AI-powered molecular generation with novel neurobiological validation may vastly improve the efficiency, translational applicability, and the probability of success in CNS drug discovery(83).

### 7.5 Case Study: TNIK-Targeted Pulmonary Therapy via Generative AI

This Nature Medicine (2025) study highlights the successful clinical translation of artificial intelligence (AI) in drug discovery through the development of rentosertib, an oral TNIK inhibitor for idiopathic pulmonary fibrosis (IPF). Using an AI-driven platform (PandaOmics), TNIK was rapidly identified as a novel antifibrotic target, and rentosertib was designed and advanced into human trials. In this phase 2a randomized, placebo-controlled study (n = 71), rentosertib demonstrated good safety and tolerability, with dose-dependent stabilization and improvement in lung function (FVC) and reductions in fibrosis-associated serum biomarkers over 12 weeks. The study provides strong evidence that AI can accelerate target identification, drug design, and early clinical validation, marking a significant milestone for AI-enabled precision medicine in fibrotic lung disease(84).

## REFERENCES

- [1]. Sleire AL, Førde HE, Anne I, Leiss L, Enger PØ, Sleire L, et al. Drug repurposing in cancer. *Pharmacol Res* [Internet]. 2017;124:74–91. Available from: <http://dx.doi.org/10.1016/j.phrs.2017.07.013>
- [2]. Shareef U, Altaf A, Ahmed M, Akhtar N, Almuhayawi MS, Al SK, et al. A comprehensive review of discovery and development of drugs discovered from 2020 – 2022. *Saudi Pharm J* [Internet]. 2024;32(1):101913. Available from: <https://doi.org/10.1016/j.jsps.2023.101913>
- [3]. Roope LSJ. The economic challenges of new drug development. *J Control Release* [Internet]. 2022;345(March):275–7. Available from:

- <https://doi.org/10.1016/j.jconrel.2022.03.023>
- [4]. Mishra A. A Comprehensive Review of Artificial Intelligence and Machine Learning: Concepts, Trends, and Applications. *Int J Sci Res Sci Technology* [Internet]. 2024;126–42. Available from: <https://doi.org/10.32628/IJSRST2411587>
- [5]. Paul D, Sanap G, Shenoy S. Artificial intelligence in drug discovery and development. *Elsevier Drug Discov Today*. 2021;26(4):80–93.
- [6]. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug. *Br J Pharmacol*. 2011;1239–49.
- [7]. Khan SR, Rijjal D Al, Piro A, Wheeler MB. Integration of AI and traditional medicine in drug discovery. *Drug Discov Today* [Internet]. 2021;26(4):982–92. Available from: <https://doi.org/10.1016/j.drudis.2021.01.008>
- [8]. Abbas MKG, Rassam A, Karamshahi F, Abunora R. The Role of AI in Drug Discovery. *ChemBioChem*. 2024;25(14).
- [9]. Stevens M, El-tayeb A, Johnson E. Modern Approaches to Drug Development: From Discovery to Market. *Acad J Pharm Sciencenes Drug Researrh*. 2024;01(01):6–12.
- [10]. Sharma D, Anabala M, Jain VV, Shyam M, Prince SE, Muniyan R. Computational Landscape in Drug Discovery: From AI / ML Models to Translational Application. Wiley. 2025;
- [11]. Ferreira FJN, Carneiro AS. AI-Driven Drug Discovery: A Comprehensive Review. *Am Chem Soc*. 2025;23889–903.
- [12]. Savage N. Tapping into the drug discovery potential of AI. *Nature* [Internet]. 2021;(June):37–9. Available from: [www.nature.com/articles/d43747-021-00045-7](http://www.nature.com/articles/d43747-021-00045-7)
- [13]. Pun FW, Ozerov I V, Zhavoronkov A. AI-powered therapeutic target discovery. *Trends Pharmacol Sci* [Internet]. 2023;44(9):561–72. Available from: <https://doi.org/10.1016/j.tips.2023.06.010>
- [14]. Wenteler A, Cabrera CP, Wei W, Neduva V, Barnes MR. AI approaches for the discovery and validation of drug targets. Cambridge Univ Press. 2024;
- [15]. Song Jiali et.al. The Discovery of New Drug-Target Interactions for Breast Cancer Treatment. 2021;26:7474. Available from: <https://doi.org/10.3390/molecules26247474>
- [16]. Chawla N V, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res* [Internet]. 2002;16:321–57. Available from: <https://doi.org/10.1613/jair.953>
- [17]. Buphamalai P, Kokotovic T, Nagy V, Menche J. Network analysis reveals rare disease signatures across multiple levels of biological organization. *Nat Commun*. 2021;
- [18]. Jensen C, Teng Y. Is It Time to Start Transitioning From 2D to 3D Cell Culture? *Front Mol Biosci*. 2020;7(March):1–15.
- [19]. Zhavoronkov A. Identification of Therapeutic Targets for Amyotrophic Lateral Sclerosis Using PandaOmics – An AI-Enabled Biological Target Discovery Platform. *Front Aging Neurosci*. 2022;14(June):1–16.
- [20]. Deore Amol et. al. The Stages of Drug Discovery and Development Process. *Asian J Phramaceutical Res Dev* [Internet]. 2019;7(6):62–7. Available from: <http://dx.doi.org/10.22270/ajprd.v7i6.616>
- [21]. Chavan SD, Chede DD, Shitole AR. Lead Optimization: Integrating Experimental, Computational, and AI / ML Approaches. *Int J Pharm Sci*. 2025;3(8):309–19.
- [22]. Slater O, Kontoyianni M, Slater O, Kontoyianni M. The compromise of virtual screening and its impact on drug discovery The compromise of virtual screening and its impact on drug discovery. *Expert Opin Drug Discov* [Internet]. 2019;14(7):619–37. Available from: <https://doi.org/10.1080/17460441.2019.1604677>
- [23]. Graff David et.al. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem Sci*. 2021;7866–81.
- [24]. Abadi HE. et. al. AI-driven covalent drug design strategies targeting main protease (mpro) against SARS-CoV-2: structural insights and molecular mechanisms. *J Biomol Struct Dyn*. 2024;
- [25]. Shao Q, Zhu W. Exploring the Ligand Binding/Unbinding Pathway by

- Selectively Enhanced Sampling of Ligand in a Protein–Ligand Complex. *J Phys Chem B*. 2019;123:7974–83.
- [26]. Sala D, Batebi H, Ledwith K, Hildebrand PW, Meiler J. Targeting in-silico GPCR Conformations with Ultra-Large Library Screening for Hit Discovery. *Trends Pharmacol Sci*. 2024;44(3):150–61.
- [27]. Tsou LK, Yeh SH, Ueng SH, Chang CP, Song JS. Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery. *Sci Rep*. 2020;16771.
- [28]. Crucitti D, Míguez CP, Arias AD, Beltra D, Orgueira AM. De novo drug design through artificial intelligence: an introduction. *Front Hematol*. 2024;3(January):1–14.
- [29]. Das U. Generative AI for drug discovery and protein design: the next frontier in AI-driven molecular science. *Med Drug Discov [Internet]*. 2025;27(June):100213. Available from: <https://doi.org/10.1016/j.medidd.2025.100213>
- [30]. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de novo design through deep reinforcement learning. *J Cheminform*. 2017;1–14.
- [31]. Martin S, Luukkonen S, Maagdenberg HW Van Den, Schoenmaker L, Olivier JMB, Westen GJP Van. DrugEx: Deep Learning Models and Tools for Exploration of Drug-Like Chemical Space. *J Chem Inf Model*. 2023;3629–36.
- [32]. Cheminform J, Liu X, Ye K, Vlijmen HWT Van, Emmerich MTM, Ijzerman AP, et al. DrugEx v2: de novo design of drug molecules by Pareto-based multi-objective reinforcement learning in polypharmacology. *J Cheminform [Internet]*. 2021;1–15. Available from: <https://doi.org/10.1186/s13321-021-00561-9>
- [33]. Drotár P, Liò P. Structure-aware generation of drug-like molecules. 2021;
- [34]. Shen C, Krenn M, Eppel S. Deep molecular dreaming: inverse machine learning for de-novo molecular design and interpretability with surjective representations OPEN ACCESS Deep molecular dreaming: inverse machine learning for de-novo molecular design and interpretability with surj. *Mach Learn Sci Technol*. 2021;
- [35]. Sanchez-lengeling B, Outeir C, Guimaraes GL. Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC). 2017;
- [36]. Putin E, Asadulaev A, Vanhaelen Q, Ivanenkov Y, Aladinskaya A V. Adversarial Threshold Neural Computer for Molecular de Novo Design. *Mol Pharm*. 2018;4386–97.
- [37]. Cao N De, Kipf T. MolGAN: An implicit generative model for small molecular graphs. 2019;
- [38]. Walid A, Simon E, Berkeley UC. ChemBERTa-2: Towards Chemical Foundation Models. 2021;
- [39]. Kubinyi H. QSAR and 3D QSAR in drug design. Part 1: Methodology. *Drug Discov Today*. 1997;2(11):457–67.
- [40]. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR Modeling- Where have you been? Where are you going to? *J Med Chem*. 2013;1–139.
- [41]. Gawehn E, Hiss JA, Schneider G. Deep Learning in Drug Discovery. *Mol Inform*. 2016;35(1):3–14.
- [42]. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. 34th Int Conf Mach Learn ICML 2017. 2017;3:2053–70.
- [43]. Ekins S. The Next Era: Deep Learning in Pharmaceutical Research. *Pharm Res [Internet]*. 2016;33(11):2594–603. Available from: <http://dx.doi.org/10.1007/s11095-016-2029-7>
- [44]. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell [Internet]*. 2020;2(10):573–84. Available from: <http://dx.doi.org/10.1038/s42256-020-00236-4>
- [45]. Wu P, Lin S, Cao G, Wu J, Jin H, Wang C, et al. Absorption, distribution, metabolism, excretion and toxicity of microplastics in the human body and health implications. *J Hazard Mater [Internet]*. 2022;437:129361. Available from: <https://doi.org/10.1016/j.jhazmat.2022.129361>

- [46]. 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* [Internet]. 2015;14(7):475–86. Available from: <http://dx.doi.org/10.1038/nrd4609>
- [47]. Yi JC, Yang ZY, Zhao WT, Yang ZJ, Zhang XC, Wu CK, et al. ChemMORT: an automatic ADMET optimization platform using deep learning and multi-objective particle swarm optimization. *Brief Bioinform*. 2024;25(2):1–10.
- [48]. Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* [Internet]. 2017;7(March):1–13. Available from: <http://dx.doi.org/10.1038/srep42717>
- [49]. Gu Y, Yu Z, Wang Y, Chen L, Lou C, Yang C, et al. admetSAR3.0: a comprehensive platform for exploration, prediction and optimization of chemical ADMET properties. *Nucleic Acids Res*. 2024;52(W1):W432–8.
- [50]. Fu L, Shi S, Yi J, Wang N, He Y, Wu Z, et al. ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Res*. 2024;52(W1):W422–31.
- [51]. Lagorce D, Bouslama L, Becot J, Miteva MA, Villoutreix BO. FAF-Drugs4: Free ADME-tox filtering computations for chemical biology and early stages drug discovery. *Bioinformatics*. 2017;33(22):3658–60.
- [52]. Banerjee P, Kemmler E, Dunkel M, Preissner R. Pro Tox 3.0: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res* [Internet]. 2024;52:W513–20. Available from: <https://doi.org/10.1093/nar/gkac303>.
- [53]. Di Stefano M, Galati S, Piazza L, Granchi C, Mancini S, Fratini F, et al. VenomPred 2.0: A Novel In Silico Platform for an Extended and Human Interpretable Toxicological Profiling of Small Molecules. *J Chem Inf Model*. 2024;64(7):2275–89.
- [54]. Wishart DS, Tian S, Allen D, Oler E, Peters H, Lui VW, et al. BioTransformer 3.0 - a web server for accurately predicting metabolic transformation products. *Nucleic Acids Res*. 2022;50(W1):W115–23.
- [55]. Matlock MK, Hughes TB, Swamidass SJ. XenoSite server: A web-available site of metabolism prediction tool. *Bioinformatics*. 2015;31(7):1136–7.
- [56]. Kuwatii AA, Nazer K, AI-Reedy A, AI-Shehri S, AI-Muhanna A, Subbarayalu AV, et al. The Role of Artificial Intelligence in Healthcare: A Review. *J Pers Med*. 2023;13(1):951.
- [57]. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today* [Internet]. 2018;23(6):1241–50. Available from: <https://doi.org/10.1016/j.drudis.2018.01.039>
- [58]. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery [Internet]. Vol. 25, *Molecular Diversity*. Springer International Publishing; 2021. 1315–1360 p. Available from: <https://doi.org/10.1007/s11030-021-10217-3>
- [59]. Schwaller P, Vaucher AC, Laplaza R, Bunne C, Krause A, Corminboeuf C, et al. Machine intelligence for chemical reaction space. *WIREs Comput Mol Sci*. 2022;12(e1604).
- [60]. MacArron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* [Internet]. 2011;10(3):188–95. Available from: <http://dx.doi.org/10.1038/nrd3368>
- [61]. Yu LX, Amidon G, Khan MA, Hoag SW, Polli J, Raju GK, et al. Understanding pharmaceutical quality by design. *AAPS J*. 2014;16(4):771–83.
- [62]. Matalqah Sina; Lafi Zainab; Mhaidat Qasim; Asha Nisreen; Asha Yousef Sara. Applications of machine learning in liposomal formulation and development. *Pharm Dev Technol* [Internet]. 2025;126–36. Available from: <https://doi.org/10.1080/10837450.2024.2448777>
- [63]. Bhatarai B, Walters WP, Hop CECA, Lanza G, Ekins S. Opportunities and

- challenges using artificial intelligence in ADME/Tox. *Nat Mater.* 2019;18(5):418–22.
- [64]. Hassanien, A., Azar, A., Gaber, T., Bhatnagar, R., F. Tolba M (eds. Role of Robotic Process Automation in Pharmaceutical Industries. *Int Conf Adv Mach Learn Technol Appl.* 2019;921.
- [65]. Fatma M. Talaat EH. Artificial Intelligence in 3D Bio Printing. *Enabling Mach Learn Appl Data Sci Algorithms Intell Syst.* 2021;(Springer Nature):77–86.
- [66]. Zhu Z, Ng DWH, Park HS, McAlpine MC. 3D-printed multifunctional materials enabled by artificial-intelligence-assisted fabrication technologies. *Nat Rev Mater* [Internet]. 2021;6(1):27–47. Available from: <http://dx.doi.org/10.1038/s41578-020-00235-2>
- [67]. Rojek I, Mikołajewski D, Dostatni E, Macko M. Ai-optimized technological aspects of the material used in 3d printing processes for selected medical applications. *Materials (Basel).* 2020;13(23):1–19.
- [68]. Vaghela MC, Rathi S, Shirole RL, Verma J, Shaheen, Panigrahi S, et al. Leveraging AI and Machine Learning in Six-Sigma Documentation for Pharmaceutical Quality Assurance. *Chinese J Appl Physiol.* 2024;40:e20240005.
- [69]. Shah N. Pharmaceutical supply chains: Key issues and strategies for optimisation. *Comput Chem Eng.* 2004;28(6–7):929–41.
- [70]. Zhou Z, Li X, Zare RN. Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Cent Sci.* 2017;3(12):1337–44.
- [71]. Gbadago DQ, Moon J, Kim M, Hwang S. A unified framework for the mathematical modelling, predictive analysis, and optimization of reaction systems using computational fluid dynamics, deep neural network and genetic algorithm: A case of butadiene synthesis. *Chem Eng J* [Internet]. 2021;409(December 2020):128163. Available from: <https://doi.org/10.1016/j.cej.2020.128163>
- [72]. Wong WC, Chee E, Li J, Wang X. Recurrent neural network-based model predictive control for continuous pharmaceutical manufacturing. *Mathematics.* 2018;6(11):242.
- [73]. McDonald MA, Salami H, Harris PR, Lagerman CE, Yang X, Bommarium AS, et al. Reactive crystallization: A review. *React Chem Eng.* 2021;6(3):364–400.
- [74]. Olawade DB, Fidelis SC, Marinze S, Egbon E, Osunmakinde A, Osborne A. Artificial intelligence in clinical trials: A comprehensive review of opportunities, challenges, and future directions. *Int J Med Inform* [Internet]. 2026;206(September 2025):106141. Available from: <https://doi.org/10.1016/j.ijmedinf.2025.106141>
- [75]. Elvatun S, Knoors D, Brant S, Jonasson C, Nygård JF. Synthetic data as external control arms in scarce single-arm clinical trials. *PLOS Digit Heal.* 2025;4(1):1–13.
- [76]. Tortora M, Pacchiano F, Ferraciolli SF, Criscuolo S, Gagliardo C, Jaber K, et al. Medical Digital Twin: A Review on Technical Principles and Clinical Applications. *J Clin Med.* 2025;14(2):1–22.
- [77]. Morone G, De Angelis L, Martino Cinnera A, Carbonetti R, Bisirri A, Ciancarelli I, et al. Artificial intelligence in clinical medicine: a state-of-the-art overview of systematic reviews with methodological recommendations for improved reporting. *Front Digit Heal* [Internet]. 2025;7(March):1–13. Available from: <https://doi.org/10.3389/fdgh.2025.1550731>
- [78]. Bhuyan SS, Sateesh V, Mukul N, Galvankar A, Mahmood A, Nauman M, et al. Generative Artificial Intelligence Use in Healthcare: Opportunities for Clinical Excellence and Administrative Efficiency. *J Med Syst.* 2025;49(1):1–11.
- [79]. Algarvio RC, Conceição J, Rodrigues PP, Ribeiro I, Ferreira-da-Silva R. Artificial intelligence in pharmacovigilance: a narrative review and practical experience with an expert-defined Bayesian network tool. *Int J Clin Pharm* [Internet]. 2025;47(4):932–44. Available from: <https://doi.org/10.1007/s11096-025-01975-3>
- [80]. El Belghiti I, Hammani O, Moustoufi F, Aghrouh M, Lemkhente Z, Boubrik F, et al. Halicin: A New Approach to Antibacterial Therapy, a Promising Avenue for the Post-Antibiotic Era. *Antibiotics.* 2025;14(7):1–12.
- [81]. Richardson PJ, Robinson BWS, Smith DP, Stebbing J. The AI-Assisted Identification



- and Clinical Efficacy of Baricitinib in the Treatment of COVID-19. *Vaccines*. 2022;10(6):1–17.
- [82]. Kwofie SK, Adams J, Broni E, Enniful KS, Agoni C, Soliman MES, et al. Artificial Intelligence, Machine Learning, and Big Data for Ebola Virus Drug Discovery. *Pharmaceuticals*. 2023;16(3):1–12.
- [83]. Hideaki Imai, Hidefumi Yoshunaga TI. An Innovative Approach to the Discovery of DSP-1181: Contributions of Artificial Intelligence, Optogenetic Technology, and Translational Biomarkers to CNS Drug Discovery. *SUMITOMO KAGAKU*; 2021.
- [84]. Xu Z, Ren F, Wang P, Cao J, Tan C, Ma D, et al. A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial. *Nat Med*. 2025;31(8):2602–10.