ISSN: 2584-2384; SJPS
Scienxt Journal of Pharmaceutical Sciences

Artificial intelligence in pharmaceutical sciences: recent advances in drug discovery, drug design, and personalized medicine – A comprehensive review

Dr. Vandana Soneja[*1], Ms. Mansi Negi[2], Ms. Parul Srivastava[3], Nooman Ansari[4]

*[*1]Principal, Subramania Barathi College of Science and Technology, Approved by A.I.C.T.E. And P.C.I. (Govt. of India) Affiliated to B.T.E. (Govt. of NCT of Delhi)*
[2]H.O.D., Subramania Barathi College of Science and Technology, Approved by A.I.C.T.E. And P.C.I. (Govt. of India) Affiliated to B.T.E. (Govt. of NCT of Delhi)
[3]Associate Professor, Delhi College of Pharmacy, Approved by Govt. of NCT of Delhi, Affiliated to Guru Gobind Singh Indraprastha University, Dwarka, Delhi
[4]Assistant Professor, Delhi College of Pharmacy-DIRD (nooman8377@gmail.com)
ORCID: 0009-0006-1021-4290

DOI: 10.5281/zenodo.24362338

Abstract:

Artificial intelligence (AI) is becoming an important analytical capability in pharmaceutical sciences because modern drug research generates large, diverse and rapidly expanding datasets. Machine learning (ML), deep learning, natural language processing, knowledge graphs and generative models can support the identification of disease mechanisms, prioritization of therapeutic targets, screening of chemical libraries, prediction of molecular properties and interpretation of patient-level information. These methods do not replace laboratory experimentation, clinical judgement or regulatory assessment; rather, they can help researchers decide which hypotheses and candidates deserve further investigation. This review examines recent advances across drug discovery, computer-aided drug design and personalized medicine, with attention to the evidence required for translation into practice.

AI-assisted target discovery combines genomics, transcriptomics, proteomics, metabolomics, imaging and biomedical literature. Predictive models can identify patterns associated with disease and rank candidate targets for experimental testing. In molecular design, structure-based prediction, ligand-based modelling and generative algorithms can propose or prioritize compounds with selected properties. Protein-structure prediction systems, including alphaFold, have expanded access to structural hypotheses, while newer interaction-prediction systems extend modelling to molecular complexes (Jumper et al., 2021; Abramson et al., 2024). Nevertheless, a predicted structure or high model score is not equivalent to demonstrated binding, biological activity, clinical benefit or safety.

Personalized medicine applies computational methods to understand differences between patients. AI may integrate genomic variants, clinical characteristics, laboratory results, medication histories and real-world outcomes to support patient stratification, biomarker discovery and individualized treatment decisions. Pharmacogenomic information can be clinically useful for selected gene–drug pairs, but recommendations must be based on validated evidence and implemented with professional oversight (Relling and Evans, 2015; Caudle et al., 2020). Major barriers include incomplete and unrepresentative datasets, poor interoperability, privacy risks, model opacity, bias, reproducibility problems and uncertainty about regulatory responsibilities.

The central conclusion is that AI is most valuable when it is embedded in a transparent, experimentally validated and clinically governed workflow. Future progress requires prospective evaluation, external validation, high-quality data, explainable outputs, cybersecurity, multidisciplinary collaboration and meaningful patient involvement. The review therefore considers both the technical promise and the practical conditions needed for responsible adoption across the pharmaceutical life cycle.

Keywords:
Artificial intelligence; machine learning; deep learning; drug discovery; computer-aided drug design; generative models; ADMET; pharmacogenomics; personalized medicine; precision therapeutics.

1. Introduction and review approach:

Pharmaceutical research is a sequence of interdependent decisions: understanding disease biology, selecting a target, finding an active molecule, optimizing its properties, establishing safety, demonstrating clinical benefit and ensuring consistent manufacture. Conventional approaches rely on experimental biology, medicinal chemistry, pharmacology and clinical investigation. These remain indispensable, but the scale of modern biological and chemical data creates opportunities for computational methods that can identify relationships not easily recognized through manual analysis. AI describes a broad family of computational approaches that learn patterns from data or use encoded knowledge to generate predictions, classifications, rankings or text. In pharmaceutical sciences, these methods are increasingly used as decision-support tools rather than autonomous authorities (Topol, 2019; Hasselgren and Oprea, 2024).

This narrative review focuses on three connected domains: AI-supported drug discovery, AI-supported drug design and AI-enabled personalized medicine. It also discusses adjacent applications in preclinical assessment, clinical development, manufacturing and pharmacovigilance because the value of a candidate molecule depends on evidence across its full life cycle. The literature base emphasizes peer-reviewed studies, major methodological papers, high-impact reviews and regulatory guidance available up to November 2026. Foundational studies are included where they explain methods still used in current research. Recent reviews and policy documents are used to describe emerging practice, but their claims are interpreted in light of the underlying validation evidence.

The review is organized thematically rather than as a formal systematic review. It does not claim exhaustive database retrieval, quantitative meta-analysis or a formal risk-of-bias assessment. Priority is given to studies that describe a defined method, dataset, validation strategy or translational application. In interpreting evidence, a distinction is maintained between computational performances, experimental confirmation and demonstrated patient benefit. A model may perform well on a retrospective test set but fail when applied to a different laboratory, population, disease subtype or clinical workflow. This distinction is essential when evaluating claims that AI shortens development or improves success rates (Vamathevan et al., 2019; Zhang et al., 2025).

The review uses “AI” as an umbrella term. ML refers to algorithms that learn relationships from examples; deep learning uses multilayer neural networks; natural language processing extracts or generates information from text; and generative AI creates new candidate data such as molecular structures or sequences. “Personalized medicine” is used broadly for care informed by individual biological and clinical characteristics, while “precision medicine” emphasizes the selection of interventions for groups or individuals based on relevant variation. These terms overlap, but neither implies that every patient requires a unique medicine. The practical goal is to improve the match between a patient, a treatment and a clinically meaningful outcome.

2. Foundations: data, algorithms and validation:

The performance of an AI system depends on the quality, relevance and structure of its input data. Pharmaceutical datasets may include chemical structures, assay results, protein sequences, three-dimensional structures, gene-expression profiles, pathology images, electronic health records, adverse-event reports and scientific publications. These data differ in scale, format, measurement error and biological meaning. Before modelling, investigators must define the prediction task, identify the intended population and outcome, assess missingness, standardize units and labels, and document the origin of each dataset. Data leakage occurs when information from the test set or future outcome is inadvertently available during model training, producing an overly optimistic estimate of performance.

Supervised learning uses labelled examples to predict an outcome, such as biological activity or toxicity. Unsupervised learning identifies structure in data without predefined labels and may be used for clustering or dimensionality reduction. Reinforcement learning optimizes a sequence of actions through feedback and has been explored for molecular generation and laboratory automation. Deep neural networks can learn complex representations from images, sequences and molecular graphs, but they often require substantial data and computational resources. Graph neural networks represent atoms and bonds as nodes and edges; transformers use attention mechanisms to model relationships across sequences; and generative models learn distributions from which new structures can be sampled (Gangwal et al., 2024).

A valid model assessment requires more than reporting accuracy. For classification, sensitivity, specificity, precision, recall, calibration and area under the receiver operating characteristic curve may be relevant. For regression, investigators may report mean absolute error, root mean square error and correlation, while ranking systems require measures such as enrichment or precision among top-ranked candidates. Metrics should reflect the cost of false positives and false negatives in the intended use. A toxicity model used to exclude a candidate may require a different threshold from a model used to prioritize compounds for further testing.

Validation should be separated into internal, temporal, external and prospective stages. Internal validation tests performance on data withheld from model fitting; temporal validation examines later observations; external validation evaluates independent datasets, institutions or populations; and prospective validation tests performance in the workflow where the model will be used. For a pharmaceutical decision, validation should also examine domain applicability, uncertainty, subgroup performance and robustness to changes in assay conditions. Model cards, data documentation, version control and pre-specified acceptance criteria help make the evidence auditable. The FDA's 2025 draft guidance recommends a risk-based credibility assessment tied to a clearly defined context of use, rather than treating model credibility as a universal property (US FDA, 2025).

Reproducibility requires reporting data sources, preprocessing, and model architecture, training procedures, hyperparameters, software versions, evaluation design and limitations. Independent replication is particularly important when the model is proprietary or trained on non-public data. AI should be viewed as one component of a scientific evidence chain: a prediction proposes a testable hypothesis, while experimental and clinical studies determine whether the hypothesis is correct and useful.

3. AI in drug discovery: target identification:

Target identification seeks biological molecules or pathways whose modulation may alter disease processes. AI can support this task by integrating genomic association studies, gene expression, protein interaction networks, perturbation screens, disease phenotypes, chemical biology and biomedical literature. Network-based methods can identify genes that occupy influential positions in disease-associated pathways, while representation-learning methods can compare molecular profiles across tissues, cell types and disease states. Natural language processing can extract relationships among genes, proteins, compounds and diseases from publications, helping researchers organize large bodies of evidence (Vamathevan et al., 2019; Zhang et al., 2025).

Multi-omics integration is attractive because disease biology is rarely explained by a single measurement. Genomic variants may indicate inherited susceptibility; transcriptomic changes show altered gene activity; proteomic data describe protein abundance and modification; and metabolomic profiles provide information about biochemical activity. AI can combine these layers to identify convergent signals and propose candidate biomarkers. However, multi-omics datasets are often small relative to the number of measured features. Batch effects, differences in sample handling, population structure and inconsistent clinical phenotyping can produce associations that do not reproduce. Model development should therefore include biological plausibility, independent replication and carefully designed perturbation experiments.

Knowledge graphs provide another framework. They represent entities such as genes, proteins, diseases, drugs, pathways and adverse effects as nodes connected by typed relationships. Graph algorithms can identify indirect links, prioritize repurposing hypotheses or reveal gaps in evidence. These outputs are useful for hypothesis generation, but an edge in a knowledge graph may reflect a weak or indirect association rather than a causal mechanism. Each proposed relationship should be traced to its source and assessed for evidence strength, directionality, experimental context and recency.

Target prioritization should combine computational scores with druggability, tissue expression, disease relevance, safety biology, genetic support and feasibility of experimental testing. A target with a strong disease association may be unsuitable if its inhibition causes unacceptable toxicity in essential tissues. Conversely, a modest statistical association may become compelling when supported by human genetics, functional experiments and a tractable binding site. AI can help rank these factors, but the weighting of criteria should be explicit and open to review.

A translational target workflow therefore proceeds from data integration to candidate ranking, mechanistic interpretation, perturbation testing and independent validation. Researchers may use CRISPR-based experiments, RNA interference, organoids, cellular assays or animal models to determine whether manipulating the proposed target produces the expected phenotype. The most credible use of AI is to reduce the search space and guide experiments, not to declare a target validated solely from computational evidence. Human genetic evidence and well-designed functional studies remain important safeguards against pursuing attractive but biologically irrelevant correlations.

4. Virtual screening and hit identification:

Virtual screening applies computational methods to prioritize compounds for experimental testing. Ligand-based approaches use information from known active and inactive molecules, including chemical fingerprints, descriptors and learned molecular representations. Structure-based approaches use a three-dimensional model of a biological target to estimate how candidate molecules may fit within a binding site. Docking algorithms generate possible binding poses and scoring functions estimate interactions, including shape complementarity, hydrogen bonding, electrostatics and hydrophobic contacts. AI models can augment these methods by learning from large collections of molecular structures and assay results, improving prioritization when training data are relevant to the target and assay.

A typical screening workflow begins with a curated compound library. Structures are standardized for salts, tautomers, stereochemistry and protonation states; duplicates and invalid molecules are removed; and compounds may be filtered using physicochemical properties. Candidate molecules are then ranked through ligand-based prediction, docking, pharmacophore matching or ensemble methods. The top-ranked molecules are reviewed for chemical plausibility, novelty, availability, synthetic accessibility and potential liabilities before experimental testing. Screening performance should be evaluated using known actives and appropriate decoys, while avoiding unrealistic assumptions about the chemical space represented in the benchmark.

Protein structure prediction has strengthened structure-guided discovery. AlphaFold demonstrated highly accurate prediction of many protein structures from sequence information, expanding structural coverage for biological targets (Jumper et al., 2021; Tunyasuvunakool et al., 2021). AlphaFold 3 extended prediction to complexes involving proteins, nucleic acids, small molecules, ions and modified residues (Abramson et al., 2024). These developments provide useful structural hypotheses, but predicted structures have uncertainty and may not represent all conformational states, ligand-induced changes, membrane environments or post-translational modifications. Predicted protein–ligand interactions should be tested through biochemical, biophysical and cellular experiments.

AI-based screening may improve enrichment of active compounds, but retrospective performance can be inflated by dataset bias, analogue leakage or differences between training and test assays. Random splitting of highly similar molecules across training and test sets can overestimate generalization. Scaffold-based, temporal and external validation provide more demanding tests. Models should also be assessed for applicability domain: a prediction for a chemical series unlike anything in the training data should carry greater uncertainty. Human review is needed to identify reactive groups, unstable structures, synthetic difficulties and assay-interference risks.

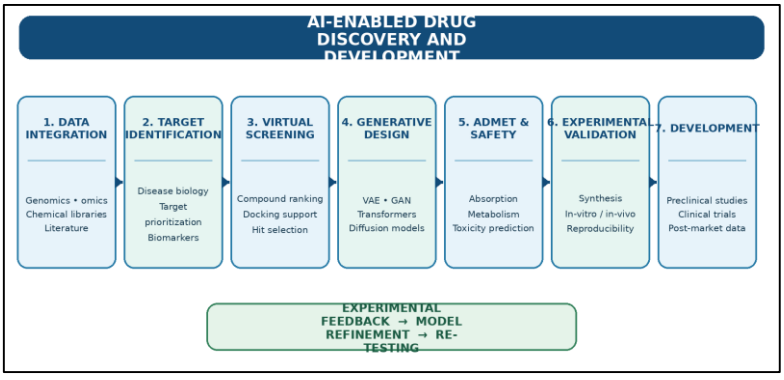


Figure. 1: AI-enabled drug discovery and development pipeline

5. AI in drug design and lead optimization:

Drug design transforms a biological hypothesis and an initial hit into a candidate with an appropriate balance of potency, selectivity, exposure, safety and developability. AI can support structure-based design by analysing target structures and ligand interactions, and ligand-based design by learning relationships between molecular features and measured activity. Quantitative structure–activity relationship models estimate how changes in chemical structure may alter biological activity. Modern approaches use molecular fingerprints, graph neural networks, three-dimensional descriptors and pretrained representations to learn patterns across chemical space (Schneider et al., 2020; Singh et al., 2024).

Lead optimization is a multi-objective problem. Improving potency may reduce solubility; increasing lipophilicity may improve membrane permeability but worsen metabolic stability or off-target binding; and structural changes may affect synthetic accessibility. Multi-task and multi-objective models can help explore these trade-offs by predicting several properties together. Optimization should not focus on a single score. Instead, medicinal chemists consider a property profile that includes target activity, selectivity, solubility, permeability, metabolic stability, plasma protein binding, clearance, toxicity alerts and the feasibility of synthesis.

Molecular docking and molecular dynamics can provide complementary information about binding. Docking samples poses and ranks them using approximate scoring functions, whereas molecular dynamics simulates molecular motion under a force field. AI can accelerate parts of these calculations, learn surrogate models or help analyze trajectories. However, predicted affinity is sensitive to protein conformation, protonation, solvent treatment, ligand strain and scoring assumptions. A low predicted binding energy should not be interpreted as proof of affinity or efficacy. Orthogonal assays are required to confirm binding and functional activity.

Protein and antibody design are expanding areas. Generative systems can propose sequences or structures with selected properties, while structure-prediction tools can help assess whether a proposed sequence may adopt a plausible fold. These approaches may support enzyme engineering, biologics optimization and the design of binders. Yet protein expression, folding, aggregation, immunogenicity, stability and manufacturability are experimentally determined properties. In silico design is a starting point for iterative design–build–test cycles, not a substitute for biochemical characterization.

Medicinal chemists remain central because chemical decisions require tacit knowledge of reaction mechanisms, reactivity, stereochemistry, protecting groups, route design, intellectual property and laboratory constraints. An effective human–AI workflow presents ranked suggestions with uncertainty and interpretable supporting features. Chemists can reject proposals that are chemically implausible or strategically unsuitable and can use experimental results to update the model. This partnership is more credible than fully automated optimization without expert review.

6. Generative AI and de novo molecular design:

Generative AI learns patterns in existing data and uses them to propose new examples. In drug discovery, a generative model may produce molecular strings, molecular graphs, three-dimensional conformations, peptides or protein sequences. Variational autoencoders learn a compressed latent representation from which new structures can be sampled. Generative adversarial networks train a generator against a discriminator. Autoregressive transformers predict the next token or structural element based on previous context, while diffusion models iteratively transform noise into structured outputs. Each family has different strengths, limitations and requirements for molecular validity (Gangwal et al., 2024; Gangwal and Lavecchia, 2024). A generative workflow begins by defining the design objective and constraints. These may include a target-associated activity profile, molecular weight range, solubility, selectivity, synthetic accessibility, novelty or avoidance of toxicophores. The model proposes candidate structures, which are filtered for valid chemistry, duplicates, reactive groups, stability and feasibility. Candidates are then evaluated using predictive models, docking, pharmacophore methods or physics-based calculations. Shortlisted compounds are synthesized and tested, and the results are fed back into subsequent model iterations. This design–make–test–analyse cycle is central to credible generative discovery.A key challenge is reward hacking: a model may optimize a numerical objective while producing compounds that are chemically unstable, synthetically inaccessible or active only in a flawed assay? Models may also reproduce biases in training libraries, generate familiar scaffolds rather than truly novel chemistry, or exploit weaknesses in scoring functions. Validity, uniqueness and novelty metrics are necessary but insufficient. A candidate should also be assessed for route feasibility, intellectual-property considerations, assay interference, structural alerts and the reliability of predicted properties.Generative models are increasingly combined with reinforcement learning, transfer learning and active learning. Reinforcement learning can adjust a generator toward a specified objective, but the reward function must represent meaningful pharmaceutical priorities. Active learning selects the next compounds for

testing to maximize information gain or improve a model in regions of uncertainty. Transfer learning can adapt a model trained on broad chemical data to a smaller target-specific dataset, although overfitting remains possible. Human review of proposed structures and experimental validation are essential throughout.The most realistic near-term contribution of generative AI is not the automatic creation of a finished medicine. It is the ability to explore chemical hypotheses, suggest focused analogues, prioritize synthesis and reduce unproductive experimental cycles. A compound generated by AI must still pass the same standards for identity, purity, potency, selectivity, pharmacokinetics, toxicology, clinical efficacy and manufacturing quality as any other candidate. Claims of accelerated discovery should be supported by transparent comparisons with conventional workflows and by reporting both successful and unsuccessful projects.

7. Admet, toxicology and preclinical decision support:

ADMET describes absorption, distribution, metabolism, excretion and toxicity. These properties determine whether a compound can reach its target at a sufficient concentration without causing unacceptable harm. AI models are used to predict solubility, permeability, plasma protein binding, metabolic stability, transporter interactions, cytochrome P450 inhibition, clearance, cardiac ion-channel liability, mutagenicity and organ toxicity. Models may use molecular descriptors, fingerprints, graph representations, assay results and biological context. Early prediction can help prioritize candidates and identify liabilities before extensive synthesis or animal studies (Vamathevan et al., 2019; Hasselgren and Oprea, 2024).

The reliability of an ADMET model depends on how closely the training data match the intended application. Different laboratories may use different assay protocols, cell systems, species, concentrations and endpoints. A model trained on one assay may not predict another assay measuring a related but non-identical property. In vitro endpoints do not always translate directly to in vivo exposure or clinical toxicity. Investigators should record assay provenance, define the endpoint precisely, harmonize units and distinguish measured values from imputed or literature-derived labels.

Toxicity prediction requires particular caution because false negatives may expose patients or trial participants to harm, while false positives may eliminate useful compounds. A model should provide calibrated uncertainty and should be evaluated at clinically or scientifically meaningful thresholds. It should not be used as the sole basis for declaring a candidate safe. Safety assessment integrates in vitro assays, in vivo studies where appropriate, mechanistic understanding, exposure margins, pathology and clinical monitoring. AI can help organize evidence and identify patterns, but toxicologists must interpret biological plausibility and study limitations.

Physiologically based pharmacokinetic modelling and quantitative systems pharmacology can be combined with ML to predict exposure and response. These models may integrate absorption, distribution, metabolism, excretion, organ function, drug interactions and disease physiology. Hybrid models can combine mechanistic equations with data-driven components, potentially improving extrapolation when data are sparse. However, a hybrid model still requires verification of assumptions, sensitivity analysis and validation against observed concentrations or outcomes.

Preclinical decisions should be documented in a traceable evidence package. This includes the model's intended use, training data, and performance by chemical class, applicability domain, uncertainty, experimental confirmation and known failure modes. The objective is not to replace established safety testing, but to allocate experiments more intelligently and identify risks earlier. A candidate should advance only when the totality of evidence supports further development, not because a computational dashboard displays a favorable composite score.

8. AI in personalized medicine:

Personalized medicine recognizes that patients with the same diagnosis may differ in disease biology, drug response, adverse-effect susceptibility, organ function, coexisting conditions and treatment preferences. AI can integrate information from genomic testing, clinical records, laboratory measurements, imaging, medication histories, lifestyle factors and real-world outcomes to identify clinically relevant subgroups. Predictive models may estimate the probability of response, adverse events or disease progression and may support selection among available treatment options. The goal is to improve the match between patient characteristics and evidence-based treatment, not to automate decisions without clinical oversight (Topol, 2019; Rajkomar et al., 2019).

Multi-omics analysis is a major opportunity. Genomics identifies inherited variants; transcriptomics measures RNA expression; proteomics examines protein abundance and modification; and metabolomics reflects biochemical activity. When combined with clinical phenotypes, these layers may reveal disease subtypes or biomarkers associated with treatment response. AI methods can reduce dimensionality, cluster patients and

identify nonlinear relationships. Yet high-dimensional data can produce unstable patterns, particularly when sample sizes are small. Validation in independent cohorts and assessment across demographic and clinical subgroups are essential before using a model to guide care.

Clinical decision support can combine predicted risk with guidelines, contraindications, laboratory values and patient preferences. A system may flag a potential interaction, suggest a dose review for impaired renal function, identify a relevant pharmacogenomic result or remind a clinician to monitor a biomarker. Such tools must display the basis and limitations of recommendations and allow clinicians to override them with documented reasoning. A model should not recommend a treatment outside the evidence base simply because it predicts a high response probability.

AI-based stratification may be useful in oncology, rare diseases, cardiovascular medicine, infectious diseases and chronic conditions. In oncology, models can integrate tumour genomics, histopathology and clinical variables to support molecular subgrouping and treatment selection. In rare diseases, phenotype matching and genomic prioritization may shorten the search for a diagnosis. In chronic disease, longitudinal data may help identify trajectories and predict deterioration. These applications differ in maturity, and evidence from one setting should not be generalized to another without validation.

Fig. 2 presents a conceptual framework for personalized medicine. Genomic and pharmacogenomic information, clinical and laboratory data, medication exposure, lifestyle and real-world outcomes are integrated to inform a patient-centred decision. The final decision remains a shared process involving the clinician and patient, with consideration of treatment goals, contraindications, uncertainty, affordability and monitoring needs. Personalized medicine is therefore a sociotechnical process, not merely an algorithmic prediction.

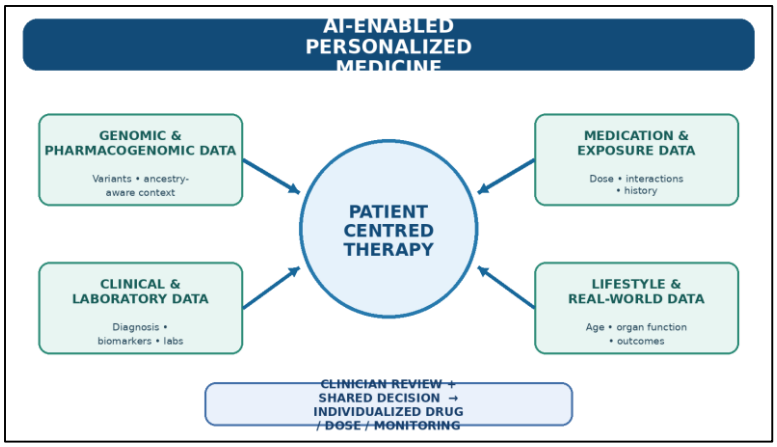


Figure. 2: AI-enabled personalized medicine

9. Pharmacogenomics, dose individualization and equity:

Pharmacogenomics studies how genetic variation influences medication response. Variants in drug-metabolizing enzymes, transporters, receptors and immune-related genes may alter drug exposure, efficacy or risk of adverse reactions. AI can help analyze complex genotype–phenotype relationships, combine genetic information with clinical covariates and identify patients who may benefit from testing. However, clinical implementation should rely on validated gene–drug evidence and established interpretation standards. CPIC guidelines translate selected pharmacogenomic findings into prescribing recommendations, while emphasizing that the availability and interpretation of a genetic result are distinct from the decision to order a test (Relling and Evans, 2015; Caudle et al., 2020).

Dose individualization may incorporate age, body size, renal and hepatic function, interacting medicines, therapeutic drug monitoring, adherence and pharmacogenomic phenotype. ML models can estimate exposure or predict dose requirements from historical data, but they must be validated for the intended population and dosing context. A model trained in adults may not apply to children; a model trained in one ancestry group may perform poorly in another; and a model derived from a specialist hospital may not transfer to primary care. Dose recommendations require clinical review, monitoring and clear responsibility for action.

Personalized medicine also raises questions about equity. Genomic databases and clinical datasets have historically overrepresented some populations and underrepresented others. A model trained on unbalanced data may produce less accurate predictions for groups with limited representation. Differences in access to testing, specialist care, digital devices and high-quality records may widen disparities even when the technology performs well in a research setting. Fairness assessment should therefore include subgroup performance, calibration, access barriers and the consequences of errors. Simply removing demographic variables does not necessarily remove bias because other variables may act as proxies.

Patient privacy and consent are central. Genomic data are inherently identifying and may reveal information about biological relatives. Pharmaceutical and healthcare organizations should use data minimization, secure storage, access controls, encryption, audit trails and clear governance for secondary use. Patients should understand what data are collected, why they are used, who may access them and whether results will influence care. Where possible, privacy-preserving approaches such as federated learning may allow models to learn across institutions without centralizing all raw records, although these approaches do not eliminate privacy risks and still require governance.

For responsible use, an AI-supported pharmacogenomic recommendation should state the relevant variant or phenotype, evidence level, guideline source, intended medication, clinical caveats and monitoring plan. The clinician should confirm the patient identity, medication list, laboratory status and clinical context before acting. AI can make complex information more accessible, but it cannot resolve uncertainty in the underlying evidence or substitute for a patient-specific benefit–risk assessment.

10. AI in clinical development and evidence generation:

Clinical development is expensive and complex because trials must recruit appropriate participants, use valid endpoints, protect participants and generate reliable evidence. AI may support protocol feasibility, site selection, participant identification, recruitment forecasting, data cleaning, imaging analysis, endpoint assessment and safety monitoring. Natural language processing can assist with screening clinical narratives for eligibility criteria, while predictive models may estimate recruitment rates or identify sites with relevant patient populations. These applications should be evaluated for accuracy, transparency and effects on participant inclusion (Zhang et al., 2025).

Trial enrichment uses biomarkers or clinical characteristics to select participants more likely to experience the outcome of interest or respond to a treatment. AI may help identify candidate subgroups from high-dimensional data. However, enrichment can reduce generalizability if the resulting trial excludes patients who will receive the medicine after approval. Investigators should justify the biological rationale, pre-specify subgroup analyses and report how selection affects the intended population. Model-driven eligibility should not introduce unjustified exclusion based on socioeconomic status, disability, ethnicity or data availability.

Adaptive trial designs may use accumulating data to modify randomization, sample size or treatment allocation under pre-specified rules. AI can support simulation and monitoring, but trial adaptation must preserve statistical validity and participant safety. Any changes should be governed by the protocol, independent oversight and appropriate regulatory review. AI predictions must not be used to bypass informed consent or to make unreviewed changes to participant care.

Clinical data management is another area of potential benefit. Algorithms can flag inconsistent values, missing data, unusual patterns or possible protocol deviations. Automated checks may reduce manual workload, but false alerts can create additional work and missed errors can compromise data integrity. Source data verification, audit trails, validation of electronic systems and clear accountability remain necessary. AI-generated summaries should be traceable to the original data and should not obscure uncertainty or contradictory findings (Mangi L, Chawla A, et al., 2026).

The use of AI in regulatory evidence generation is receiving increasing attention. The US FDA's January 2025 draft guidance proposes a risk-based credibility framework for AI models used to support regulatory decisions about drug and biological product safety, effectiveness or quality (US FDA, 2025). The framework emphasizes context of use and the credibility of the model for that use. This is important because a model used for exploratory research has a different risk profile from one used to support a pivotal efficacy claim or a manufacturing release decision. Sponsors should document intended use, data provenance, validation, uncertainty, change control and human oversight.

11. Pharmaceutical manufacturing and pharmacovigilance:

AI applications extend beyond discovery and clinical trials into pharmaceutical manufacturing. Process analytical technology, sensor networks and manufacturing execution systems generate data on temperature, pressure, mixing, granulation, drying, compression, coating and packaging. ML models can identify process patterns associated with product quality, predict equipment maintenance needs and support real-time process monitoring. Digital twins may combine process models and live data to simulate manufacturing conditions and explore the effect of changes before they are introduced into production.

Quality-by-design principles require manufacturers to understand critical quality attributes, critical material attributes and critical process parameters. AI may help analyze multivariate relationships and detect process drift, but the model must be validated within the manufacturing context. Changes in raw-material suppliers, equipment, scale, environmental conditions or analytical methods can affect performance. A model used in a regulated environment requires documented lifecycle management, version control, change assessment, data integrity and clear procedures for human intervention. AI does not replace good manufacturing practice, validated analytical methods or batch-release requirements.

Pharmacovigilance involves the detection, assessment, understanding and prevention of adverse effects and other medicine-related problems. Spontaneous reports, electronic health records, literature, claims data and patient-generated information can be large and heterogeneous. Natural language processing can extract suspected products, reactions, timing and outcomes from narrative reports. ML can assist with duplicate detection, case triage, coding support and signal prioritization. These tools may help safety teams focus attention, but a statistical association is not proof that a medicine caused an event.

Signal detection requires clinical review of temporal relationships, biological plausibility, alternative explanations, background incidence, reporting patterns and the totality of evidence. Under-reporting, stimulated reporting, missing information and confounding by indication can distort observed associations. AI systems should therefore support, not replace, expert causality assessment. Their outputs should preserve links to source reports, explain relevant features and make uncertainty visible. Human reviewers should be able to challenge and correct model outputs, with feedback recorded for future improvement.

Across manufacturing and pharmacovigilance, governance should cover data integrity, cybersecurity, access permissions, model performance monitoring, incident reporting and business continuity. A model that performs well at deployment may degrade as products, populations, assays or manufacturing processes change. Periodic revalidation and predefined triggers for review are necessary. Responsible adoption is achieved when AI strengthens quality systems and patient safety rather than creating an opaque additional layer of risk.

12. Challenges, ethics and regulatory governance:

Data quality is a foundational challenge. Pharmaceutical datasets may contain missing values, inconsistent terminology, duplicate records, measurement error, selection bias and non-representative populations. Labels may be derived from assays that differ across laboratories or from clinical records that were not designed for research. A large dataset is not necessarily a good dataset. Data curation, provenance, harmonization and documentation require substantial scientific effort and should be treated as core research activities rather than administrative tasks. Interpretability is important when a model influences a high-consequence decision. Some systems can provide feature importance, attention maps, counterfactual examples or local explanations, but these tools do not automatically reveal the true causal mechanism. Explanations can be unstable or misleading if the underlying model is poorly specified. A transparent workflow should communicate what the model predicts, the data on which it relies, the level of uncertainty, the situations in which it should not be used and the actions that require human review. Bias can enter through sampling, measurement, labels, historical treatment patterns and deployment conditions. If past clinical practice reflects unequal access or differential treatment, a model trained on those records may reproduce the same inequalities. Fairness assessment should be tied to the intended use and should examine who benefits, who is excluded and who bears the consequences of error. Stakeholders from affected communities should be involved in defining acceptable use and monitoring outcomes. Equity is not achieved merely by reporting a single overall performance metric (Topol, 2019; Rajkomar et al., 2019).Privacy and cybersecurity are especially important because pharmaceutical AI may use genomic, health, proprietary chemical and manufacturing data. Governance should define lawful access, consent or other appropriate legal basis, data retention, secure transfer, de-identification limits, vendor responsibilities and breach response. Federated learning and privacy-enhancing computation may reduce the need to centralize raw data, but they do not remove the need for threat modelling and independent security review. Intellectual property, data ownership and licensing must also be addressed when models are trained on proprietary or third-party datasets.Regulatory expectations are evolving. The FDA's 2025 draft guidance frames credibility assessment around the model's context of use and the risk of the decision it supports (US FDA, 2025). The European Medicines Agency and other regulators have also emphasized lifecycle oversight,

human accountability and the need for reliable evidence. Organizations should establish model inventories, intended-use statements, validation plans, change-control procedures and post-deployment monitoring. The same model may require different evidence depending on whether it is used for exploratory analysis, clinical decision support, and regulatory submission or product quality decisions.

A final challenge is organizational. Successful implementation requires collaboration among medicinal chemists, pharmacologists, bioinformaticians, data scientists, clinicians, statisticians, quality professionals, regulators and patients. Training should address both technical literacy and the limits of prediction. Procurement decisions should consider interoperability, vendor transparency, auditability and exit strategies. The aim is not to adopt AI for its novelty, but to demonstrate that a defined application improves the quality, efficiency or safety of a pharmaceutical decision.

13. Future directions:

The next phase of AI in pharmaceutical sciences is likely to involve more integrated models that combine molecular, biological, clinical and real-world information. Foundation models trained on large chemical, protein, image or text datasets may be adapted to specialized tasks through fine-tuning or retrieval-based methods. Multimodal systems could connect molecular structure with assay results, microscopy, omics and clinical outcomes. Knowledge graphs may improve traceability across complex evidence, while digital twins and laboratory automation may support more iterative experimentation. These developments will be valuable only if data provenance, uncertainty and model limitations remain visible.

A promising direction is the closer integration of computational prediction with automated experimentation. In a closed-loop workflow, a model proposes candidate experiments, robotic platforms perform selected assays, results are analyzed and the model is updated. Such systems may improve experimental efficiency by selecting informative experiments rather than testing large numbers of compounds indiscriminately. However, automation introduces its own risks: equipment calibration, assay drift, sample tracking, contamination, software faults and poorly specified optimization objectives. Quality assurance and independent replication remain essential.

Research priorities include prospective evaluations that compare AI-supported workflows with conventional practice; external validation across institutions and populations; transparent reporting of negative results; benchmark datasets that reflect real pharmaceutical tasks; and methods for estimating uncertainty and detecting out-of-distribution inputs. Studies should report not only predictive metrics but also decision impact, cost, time, safety, reproducibility and effects on equity. Regulatory science should continue to develop proportionate frameworks that encourage innovation while preserving standards for quality, safety and efficacy.

Personalized medicine will require careful integration of pharmacogenomics, longitudinal clinical data, and medication exposure and patient preferences. AI may help identify which patients need closer monitoring, which biomarkers are relevant and where treatment response is likely to differ. Yet a recommendation must be clinically actionable, understandable and consistent with validated evidence. Patients should retain a meaningful role in decisions, and clinicians should remain accountable for interpreting predictions in context. The success of precision therapeutics depends as much on equitable access, communication and follow-up as on algorithmic performance.

14. Conclusion:

Artificial intelligence is reshaping pharmaceutical sciences by supporting target discovery, virtual screening, molecular design, ADMET prediction, clinical development, manufacturing, pharmacovigilance and personalized medicine. Its strongest contribution is the ability to integrate complex information and prioritize decisions for further scientific evaluation. It cannot independently establish biological causality, prove clinical benefit or guarantee safety. Sustainable progress will depend on high-quality data, rigorous validation, human expertise, transparent governance, patient-centred implementation and continuous monitoring. AI should be regarded as an enabling scientific tool whose value is demonstrated through better medicines and safer, more equitable care—not through computational novelty alone.

15. References:

(1) Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7.

(2) Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-477. doi:10.1038/s41573-019-0024-5.

(3) Hasselgren C, Oprea TI. Artificial intelligence for drug discovery: are we there yet? Annu Rev Pharmacol Toxicol. 2024;64:527-550. doi:10.1146/annurev-pharmtox-040323-040828.
(4) Zhang K, Yang X, Wang Y, Yu Y, Huang N, Li G, et al. Artificial intelligence in drug development. Nat Med. 2025;31:45-59. doi:10.1038/s41591-024-03434-4.
(5) Gangwal A, Ansari AA, Ahmad I, Azad AK, Kumarasamy VK, Subramaniyan VS, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. Front Pharmacol. 2024;15:1331062. doi:10.3389/fphar.2024.1331062.
(6) Gangwal A, Lavecchia A. Unleashing the power of generative AI in drug discovery. Drug Discov Today. 2024;29(6):103992. doi:10.1016/j.drudis.2024.103992.
(7) Singh S, Kaur N, Gehlot A. Application of artificial intelligence in drug design: a review. Comput Biol Med. 2024;108810. doi:10.1016/j.compbiomed.2024.108810.
(8) 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):353-364. doi:10.1038/s41573-019-0050-3.
(9) Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2.
(10) Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596(7873):590-596. doi:10.1038/s41586-021-03828-1.
(11) Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500. doi:10.1038/s41586-024-07487-w.
(12) 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.
(13) Zhavoronkov A, Ivanenkov YA, Aliper A, Vesíkov S, Aladinskiy V, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37(9):1038-1040. doi:10.1038/s41587-019-0224-x.
(14) Walters WP, Murcko MA. Assessing the impact of generative AI on medicinal chemistry. Nat Biotechnol. 2020;38(2):143-145. doi:10.1038/s41587-020-0418-2
(15) Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature. 2015;526(7573):343-350. doi:10.1038/nature15817.
(16) Caudle KE, Gammal RS, Whirl-Carrillo M, Hoffman JM, Relling MV, Klein TE. Evidence and resources to implement pharmacogenetic knowledge for precision medicine. Clin Pharmacol Ther. 2020;108(4):681-692. doi:10.1002/cpt.1914.
(17) Mangi L, Chawla A, Kukkar S, Sachdeva S, Nirmal, Kumar M. Glioblastoma multiforme: an updated review of pathophysiology, diagnosis, and temozolomide-based therapy. Scienxt J Pharm Sci. 2026;4(1):1-16. doi:10.5281/zenodo.18334201.
(18) Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347-1358. doi:10.1056/NEJMra1814259.
(19) Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29. doi:10.1038/s41591-018-0316-z.
(20) Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372(9):793-795. doi:10.1056/NEJMp1500523.
(21) US Food and Drug Administration. Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: draft guidance for industry and other interested parties. Silver Spring (MD): FDA; 2025 Jan.
(22) World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.
(23) Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-1250. doi:10.1016/j.drudis.2018.01.039.