International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 13; Issue: 6; June 2026
Website: www.ijrrjournal.com
Review Paper *E-ISSN: 2349-9788; P-ISSN: 2454-2237*

Artificial Intelligence and Large Language Models in Drug Discovery and Pharmacogenomics: A Comprehensive Pharmacological Review of AlphaFold, Generative AI for *De Novo* Design, Clinical Decision Support, and the Translational Pipeline

Adiba Begum[1], R. L. Manisha[2], Muvvala Sudhakar[3]

[1]Student, Department of Pharmacology, Malla Reddy College of Pharmacy, Maisammaguda, Dhulapally, Secunderabad–500100, Telangana, India.
[2]Head of Department, Department of Pharmacology, Malla Reddy College of Pharmacy, Maisammaguda, Dhulapally, Secunderabad–500100, Telangana, India.
[3]Principal, Malla Reddy College of Pharmacy, Maisammaguda, Dhulapally, Secunderabad–500100, Telangana, India.

Corresponding Author: Dr. R. L. Manisha

DOI: https://doi.org/10.52403/ijrr.20260610

ABSTRACT

Artificial intelligence (AI) — particularly deep learning, generative models, and large language models (LLMs) — has emerged as a transformative force across pharmacology, drug discovery, clinical pharmacy, and pharmacogenomics. The 2020 release of AlphaFold by DeepMind, recognized with the 2024 Nobel Prize in Chemistry (jointly to Demis Hassabis, John Jumper, and David Baker), produced near-experimental-accuracy three-dimensional structures for nearly all known proteins, fundamentally transforming structural biology and structure-based drug design. The 2022–2025 wave of generative AI and large language models (GPT-4, Claude, Gemini, Llama) has demonstrated capabilities relevant to medical decision support, clinical documentation, drug information, pharmacovigilance, and patient education. AI-native biotechnology companies including Insilico Medicine, Recursion Pharmaceuticals, BenevolentAI, Atomwise, Exscientia, Isomorphic Labs, and Iambic Therapeutics have advanced AI-discovered drug candidates into clinical trials, with INS018_055 (Insilico, idiopathic pulmonary fibrosis) progressing to Phase 2 as a notable early example. The pharmacogenomics field is increasingly leveraging machine learning for genome-wide variant interpretation, polygenic risk scoring, drug-gene interaction prediction, and dose individualization. This review provides a comprehensive analysis of AI applications across the pharmaceutical value chain, including the technical foundations, target identification and validation, hit identification and lead optimization, ADMET prediction, AI-driven clinical trial design and patient stratification, post-market pharmacovigilance, LLM-based clinical decision support, the evolving regulatory landscape (FDA, EMA, and CDSCO frameworks), pharmacogenomic applications, the Indian AI and biotechnology context, and future directions for pharmacology education and translational research.

Keywords: Artificial intelligence; Large language models; AlphaFold; Drug discovery; Pharmacogenomics; Generative AI; Deep learning; Transformer; Clinical decision support; Foundation models; Precision medicine.

1. INTRODUCTION

Artificial intelligence (AI) has transitioned from speculative future possibility to operational reality across pharmacology, drug discovery, clinical medicine, and pharmacogenomics. The transformation has been driven by convergent advances in deep learning architectures (transformers, diffusion models, graph neural networks), exponentially increasing computational power (specialized GPU clusters, TPUs), unprecedented data scale (genomic databases, electronic health records, chemical libraries, structural biology resources), and large-scale investment from both established pharmaceutical companies and AI-native biotechnology firms. The pharmaceutical industry — long characterized by 10–15-year development timelines, attrition rates exceeding 90% across phase 1 to approval, and per-approval costs estimated at USD 2.6 billion — represents both a major beneficiary and a critical proving ground for AI capabilities. [1,2,3,4,5,6]

The 2020 release of AlphaFold 2 by DeepMind marked a watershed moment. After decades of incremental progress in protein structure prediction, AlphaFold achieved near-experimental accuracy (median backbone GDT_TS of 92.4 in CASP14), producing the AlphaFold Protein Structure Database with predicted structures for over 200 million proteins from species across the tree of life. The 2024 Nobel Prize in Chemistry recognized the achievement, jointly awarded to Demis Hassabis and John Jumper of DeepMind (for AlphaFold) and David Baker (for computational protein design). AlphaFold 3 (2024) extended capabilities to protein-ligand, protein-protein, and protein-DNA/RNA complex prediction, further expanding utility for drug discovery. [7,8]

Parallel to AlphaFold's structural revolution, the 2022-2025 wave of generative AI and large language models has demonstrated transformative capabilities in natural language understanding, medical knowledge synthesis, and creative generation. GPT-4 (OpenAI, 2023), Claude (Anthropic, 2023-2025), Gemini (Google DeepMind, 2023-2025), and open-source alternatives including Llama (Meta) and Mistral, achieved human-expert-level performance on medical knowledge tests including USMLE (90%+) and demonstrated practical applications in clinical documentation, drug information, pharmacovigilance, patient education, and clinical decision support. Specialized biomedical LLMs (Med-PaLM 2, GatorTron, BioMedLM, ChemLLM) extend foundation model capabilities to biomedical domains. [9,10,11]

Several AI-native biotechnology companies have advanced beyond proof-of-concept to clinical-stage assets. Insilico Medicine's INS018_055, an AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis, progressed to Phase 2 trials with a notable accelerated discovery timeline (approximately 18 months from target identification to phase 1, versus typical 4-6 years). Recursion Pharmaceuticals has built a phenotypic AI platform with multiple clinical candidates including REC-994 (cerebral cavernous malformation) and REC-2282 (neurofibromatosis 2). BenevolentAI applied AI in COVID-19 drug repurposing identifying baricitinib in early pandemic phases. Exscientia advanced AI-designed candidates in oncology and immunology. Isomorphic Labs (DeepMind subsidiary) has established multibillion-dollar partnerships with Eli Lilly and Novartis. The translation from AI hype to demonstrated clinical impact is occurring. [12]

This review provides a comprehensive analysis of AI and LLM applications across pharmacology and the pharmaceutical value chain. We examine the technical foundations of modern AI (transformer architectures,

generative models, geometric deep learning, foundation models), the impact of AlphaFold and structural AI, target identification and validation, hit identification and lead optimization, ADMET prediction, AI-driven clinical trial design and patient stratification, real-world evidence generation, post-market pharmacovigilance, LLM-based clinical decision support, the regulatory landscape, pharmacogenomic applications, the Indian AI/biotech context, ethical considerations, and future directions. The review aims to provide a foundational understanding for pharmacy and medical educators preparing for the AI-augmented future of clinical and pharmaceutical practice.

2. TECHNICAL FOUNDATIONS OF MODERN AI

2.1 Deep Learning and Transformers

Modern AI in biomedicine builds on deep neural networks — multilayered architectures that learn hierarchical representations from data. The transformer architecture, introduced in 2017 (Vaswani et al., 'Attention Is All You Need'), revolutionized natural language processing through self-attention mechanisms that capture long-range dependencies efficiently. Transformers underlie the major LLMs (GPT, Claude, Gemini), protein structure prediction (AlphaFold employs attention-based modules called Evoformer and structure module), and increasingly chemistry (chemical language models, molecular property prediction). The scaling of transformer-based models — with parameters from millions to hundreds of billions, and training datasets from millions to trillions of tokens — has produced emergent capabilities not present in smaller models. [13,14]

2.2 Generative AI: Diffusion Models and Variational Autoencoders

Generative AI extends machine learning from discriminative tasks (classification, regression) to creative generation — producing novel data resembling the training distribution. Diffusion models (e.g., RFdiffusion for protein design, DALL-E for images) have achieved state-of-the-art performance in many generative tasks. Variational autoencoders (VAEs) and generative adversarial networks (GANs) have been extensively applied to molecular generation, producing novel chemical structures with desired properties. In drug discovery, generative models enable de novo design of small molecules, peptides, antibodies, and even gene therapy constructs, expanding chemical space exploration beyond traditional library screening.

2.3 Geometric Deep Learning and Graph Neural Networks

Biological molecules — proteins, nucleic acids, small molecules — are fundamentally geometric and relational objects, with structure (atomic positions in three-dimensional space) and graph topology (covalent and non-covalent connections). Geometric deep learning, particularly graph neural networks (GNNs), provides architectures naturally suited to molecular and biological data. GNNs operate directly on molecular graphs (atoms as nodes, bonds as edges) with learned representations that respect chemical and physical symmetries. Equivariant neural networks (e.g., E3NN, SE (3)-Transformers, EquiFormer) extend the framework to rotation and translation symmetries critical for molecular modeling.

2.4 Foundation Models for Biology

Foundation models — large pre-trained models that can be adapted to downstream tasks — are emerging in biology and chemistry. Protein language models (ESM-2, ProtT5, ProtBERT) trained on millions of protein sequences learn generalizable representations of protein structure and function. Chemical language models (ChemBERTa, MolBERT, ChemGPT) treat molecules as strings (SMILES, SELFIES) and learn molecular representations. Multi-modal models integrate protein sequences, structures, chemistry, and biological assay data. The foundation model paradigm — pre-training at scale followed by fine-tuning for specific tasks — is rapidly transforming biological AI and accelerating drug discovery applications. [15,16,17,18]

3. ALPHAFOLD AND STRUCTURAL AI

3.1 AlphaFold 2: The 2020 Breakthrough

AlphaFold 2, developed by DeepMind and reported in Nature 2021, achieved near-experimental accuracy in protein structure prediction. The Critical Assessment of Structure Prediction (CASP14, 2020) demonstrated AlphaFold's transformative performance: median GDT_TS of 92.4 for free modeling targets (where the closest structural template has been hidden), substantially exceeding all other methods and approaching experimental accuracy. AlphaFold's architecture — multiple sequence alignment processing through the Evoformer module, pair representations, and the structure module producing 3D coordinates — represents a fundamental advance in computational structural biology. The AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/), launched in 2021 and freely accessible, contains predicted structures for over 200 million proteins, including the entire human proteome and most model organisms.

3.2 AlphaFold 3 and Multi-Component Prediction

AlphaFold 3, reported in Nature May 2024, extends capabilities to predict protein-ligand, protein-protein, protein-DNA, protein-RNA, and protein-ion complexes. The unified architecture employs diffusion-based generation of structure coordinates and substantially improved performance for the molecular complexes critical for drug discovery. AlphaFold 3 enables structure-based virtual screening, structural understanding of antibody-antigen interactions, RNA structure prediction (relevant for RNA therapeutics), and detailed protein-cofactor modeling. Isomorphic Labs (DeepMind subsidiary) has commercialized AlphaFold 3 for proprietary drug discovery applications alongside the AlphaFold Server providing public research access. [19]

3.3 RFdiffusion and Generative Protein Design

RFdiffusion, developed by David Baker's laboratory at the University of Washington, applies diffusion models to de novo protein design — generating novel protein backbones with specified shapes, binding sites, or symmetries. Combined with sequence design tools (ProteinMPNN), RFdiffusion enables the design of proteins for specific functions including therapeutic binders, vaccine immunogens, enzymes, and biosensors. The Baker laboratory's contributions including RFdiffusion were recognized in the 2024 Nobel Prize in Chemistry alongside DeepMind's AlphaFold work. Therapeutically designed proteins (e.g., engineered miniproteins as antagonists or agonists of disease targets) are advancing through preclinical and early clinical development. [20,21]

3.4 Structural AI in Drug Design

AlphaFold predictions and related structural AI have transformed structure-based drug design (SBDD). Virtual screening can now target proteins lacking experimental structures, expanding druggable target space substantially. Structural elucidation of complex protein-protein interfaces, cryptic binding pockets, and conformational ensembles enables design strategies previously impossible. AI-augmented free energy perturbation (FEP) calculations accelerate hit-to-lead and lead optimization. Combined approaches integrating AlphaFold predictions with experimental cryo-EM, NMR, and crystallography produce hybrid structural understanding.

4. AI ACROSS THE DRUG DISCOVERY PIPELINE

4.1 Target Identification and Validation

AI-driven target identification leverages multi-omics integration (genomics, transcriptomics, proteomics, metabolomics, phenomics), disease network analysis, and causal inference. Mendelian randomization combined with machine learning identifies druggable targets with human genetic support. Knowledge graph approaches (BenevolentAI, Cyclica) integrate diverse biomedical data sources for target prioritization. Phenotypic AI (Recursion's image-based platform, Insitro's machine

learning-driven biology) identifies disease-modifying targets through cellular phenotype screening. The combination of AI-driven target identification with experimental validation has accelerated the early stages of drug discovery, with some companies reporting 50%+ reductions in target identification timelines. [22]

4.2 Hit Identification and Lead Optimization

AI methods including machine learning-based virtual screening, generative molecular design, and active learning have transformed hit identification and lead optimization. Generative models (REINVENT, MolGAN, ChemVAE) produce novel molecular structures with desired properties. AI-guided design-make-test-analyze (DMTA) cycles compress lead optimization timelines from years to months. Examples include Insilico's INS018_055 (TNIK inhibitor for IPF, identified and optimized in approximately 18 months), Exscientia's DSP-1181 (5-HT1A agonist for OCD, the first AI-designed drug to enter clinical trials in 2020), and multiple other early-stage candidates from BenevolentAI, Atomwise, Iambic, Schrödinger, and others. [23,24,25,26]

4.3 ADMET Prediction

ADMET (absorption, distribution, metabolism, excretion, toxicity) prediction is a long-standing AI application area. Machine learning models predict molecular properties including solubility, permeability, metabolic stability, CYP inhibition, hERG channel binding (cardiotoxicity), and various toxicity endpoints. Modern deep learning approaches (graph neural networks, transformer-based models) provide superior performance versus traditional QSAR methods. ADMET prediction enables prioritization of candidates with favorable drug-like properties early in discovery, reducing late-stage attrition. Commercial platforms (ADMETlab, ADMET Predictor, OpenADMET) integrate AI-predicted ADMET with experimental workflows.

4.4 Clinical Trial Design and Patient Stratification

AI applications in clinical trials include patient identification and recruitment optimization (matching electronic health records to eligibility criteria), digital biomarker development (wearables, smartphone-based assessments), synthetic control arms (using real-world data to reduce placebo arm requirements), adaptive trial designs (AI-guided dose escalation and arm allocation), and post-hoc subgroup analysis. Patient stratification using machine learning on multi-modal data (clinical, imaging, genomic, biomarker) increasingly identifies responder populations and informs precision medicine approaches. Notable examples include the use of AI for cancer genomics-guided clinical trials, the Karuna/BMS AI-aided trial design for KarXT, and AI-augmented stratification in oncology and immunology.

5. LARGE LANGUAGE MODELS IN CLINICAL MEDICINE

5.1 Medical Knowledge and Clinical Reasoning

Large language models have demonstrated remarkable medical knowledge and clinical reasoning capabilities. GPT-4 achieved approximately 86-90% on the USMLE Step examinations without medical-specific training; specialized medical models (Med-PaLM 2 by Google, achieving 86.5% on MedQA in 2023) exceed clinician baseline performance on standardized examinations. Performance on differential diagnosis, treatment recommendation, and clinical reasoning approaches and sometimes exceeds physician baseline, particularly for less experienced clinicians or in cognitive bias scenarios. Limitations include hallucinations (generating plausible but false information), poor calibration of uncertainty, training data cutoff dates, and lack of access to patient-specific context absent from the prompt. [27,28,29,30,31,32,33,34,35,36]

5.2 Clinical Documentation and Workflow

Clinical documentation — long a major time burden for clinicians — has been a major early LLM application. Ambient clinical AI scribes (Nuance DAX Copilot, Abridge,

Suki, Augmedix, Doximity GPT) generate clinical notes from patient-clinician conversations, with reported physician time savings of 30-50% and substantial reductions in after-hours documentation burden. Clinical workflow applications include automated coding (ICD-10, CPT), prior authorization assistance, patient communication drafting, discharge summary generation, and chart summarization. Adoption has accelerated substantially in 2023-2025 as accuracy and clinical integration have improved.

5.3 Drug Information and Pharmacovigilance

LLMs are increasingly applied to drug information services, pharmacovigilance signal detection, and adverse event analysis. Applications include automated drug interaction checking, literature synthesis for drug reviews, pharmacovigilance database analysis (FAERS, EudraVigilance), case report processing, and patient-facing medication information. Clinical pharmacist workflows benefit from LLM-augmented drug consultation, formulary decision support, and medication therapy management. Specialized platforms (e.g., Atropos Health, Tempus, Genospace) combine LLMs with specialized clinical and pharmaceutical databases for higher accuracy and traceability. [37]

5.4 Patient Education and Engagement

Patient-facing LLM applications include health information chatbots, medication adherence support, lifestyle and behavior change coaching, and accessible information for low-literacy populations. Challenges include accuracy verification, appropriate handoff to human clinicians, regulatory framework (medical device vs information service), and equity (digital divide affecting elderly, rural, and economically disadvantaged populations). Multilingual capabilities of modern LLMs (GPT-4, Claude, Gemini operate across dozens of languages including major Indian languages) offer particular potential for global health applications. [38]

6. AI-DRIVEN BIOTECHNOLOGY COMPANIES

6.1 Insilico Medicine and INS018_055

Insilico Medicine, founded in 2014 by Alex Zhavoronkov, has built one of the most prominent AI-native drug discovery platforms (Pharma.AI suite). INS018_055, identified as a TNIK kinase inhibitor for idiopathic pulmonary fibrosis using AI-driven target identification (PandaOmics) and generative chemistry (Chemistry42), entered phase 1 trials in 2023 with positive results and is advancing through phase 2. The compound's development timeline (target identification to phase 1 in approximately 18 months) represents a notable demonstration of AI-accelerated discovery. Insilico's broader pipeline includes additional AI-discovered candidates across fibrotic, oncologic, and immunologic indications. [39]

6.2 Recursion Pharmaceuticals and Phenotypic AI

Recursion Pharmaceuticals applies phenotypic screening at massive scale using cellular imaging combined with deep learning to identify therapeutic candidates. The Recursion Operating System integrates wet-lab automation, deep learning image analysis, and machine learning prioritization. Recursion's clinical pipeline includes REC-994 (cerebral cavernous malformation, phase 2/3), REC-2282 (NF2-associated meningiomas), REC-4881 (familial adenomatous polyposis), and partnerships with multiple pharmaceutical companies. The August 2023 NVIDIA $50M investment validated the platform's computational requirements. Subsequent Recursion-Exscientia merger (December 2024) created a larger AI biotech entity. [40]

6.3 Isomorphic Labs and DeepMind

Isomorphic Labs, founded in 2021 as a DeepMind subsidiary, applies AlphaFold and related AI to drug discovery. Multibillion-dollar partnerships with Eli Lilly (2024) and Novartis (2024) provide commercial validation. Isomorphic Labs' integrated platform combines AlphaFold protein structure prediction, generative molecular design, and computational

chemistry, with early-stage internal pipeline development complementing the partnership programs. The Isomorphic-Eli Lilly partnership specifically targets multiple therapeutic areas with AI-accelerated discovery.

6.4 Other Notable Players

BenevolentAI (knowledge graph-driven target identification, COVID-19 baricitinib repurposing success); Atomwise (AI-driven virtual screening, AtomNet platform); Exscientia (AI-designed candidates including DSP-1181, GTAEXS617, EXS-21546; merged with Recursion late 2024); Iambic Therapeutics (transformer-based de novo design); Insitro (machine learning-driven biology, founded by Daphne Koller); Schrödinger (computational chemistry pioneer with growing AI capabilities); NVIDIA BioNeMo (foundation model platform for biology). The breadth reflects intense investment and competitive activity in AI-driven biotechnology.

7. PHARMACOGENOMICS AND PRECISION MEDICINE

7.1 Pharmacogenomic Variant Interpretation

Pharmacogenomic variant interpretation — assigning functional significance to genetic variants in drug-metabolizing enzymes, transporters, and drug targets — increasingly leverages machine learning. The Clinical Pharmacogenetics Implementation Consortium (CPIC) maintains guidelines for major drug-gene pairs (CYP2C19-clopidogrel, CYP2D6-codeine/tamoxifen, TPMT-thiopurines, HLA-B*5701-abacavir, DPYD-fluoropyrimidines, and others). Machine learning extends interpretation to novel variants without functional data, predicts variant effects on enzyme activity, and integrates polygenic effects. Implementation in clinical practice requires integration with electronic health records, clinical decision support, and provider education. [41]

7.2 Polygenic Risk Scores and AI Integration

Polygenic risk scores (PRS) aggregate effects of many genetic variants to estimate disease risk. AI/ML methods improve PRS accuracy through integration of genomic, clinical, demographic, and environmental data. Applications include cardiovascular risk stratification, breast cancer screening optimization, statin treatment intensity selection, and emerging applications in psychiatric, neurological, and metabolic disorders. Major equity considerations include PRS validity across diverse populations — most discovery cohorts have been European-ancestry-biased, with substantially reduced PRS performance in non-European populations including South Asians. Indian-population-specific PRS development is an active research priority. [42,43,44,45,46,47,48,49]

7.3 Drug-Drug-Gene Interactions

Beyond single drug-gene interactions, complex polypharmacy in elderly and chronic disease populations involves drug-drug-gene interactions that exceed manual analysis capability. AI/ML platforms (e.g., Genoa, Translational Software, Sema4) integrate pharmacogenomic data with patient-specific medication lists to identify clinically relevant interactions and provide alternative recommendations. Implementation in clinical practice expands access to personalized medicine guidance, particularly when integrated with electronic health records and clinical decision support. [50,51]

8. INDIAN AI/BIOTECH CONTEXT

India has substantial and growing AI/biotech capabilities. Major academic centers including IIT Delhi, IIT Bombay, IIT Madras, IISc Bangalore, IIIT Hyderabad, and AIIMS Delhi have established AI research and translational programs. The Centre for Cellular and Molecular Platforms (CCAMP, Bangalore), the Regional Centre for Biotechnology (RCB, Faridabad), and the Centre for Cellular and Molecular Biology (CCMB, Hyderabad) have advanced computational biology and AI biology programs. Indian AI/biotech startups include

Strand Life Sciences (genomic medicine), MedGenome (clinical genomics with AI), HealthifyMe (AI-driven health platform), Niramai (AI-based breast cancer screening), Qure.ai (radiology AI), and emerging AI drug discovery firms. [52]

National initiatives supporting AI/biotech include the Atal Innovation Mission, the National Mission on Quantum Technologies and Applications, the Genome India Programme (sequencing 10,000 Indian genomes for population-specific data), the Indian Genome Project, and the Department of Biotechnology's National Biopharma Mission. The IndiaAI mission and the National Digital Health Mission provide infrastructure relevant to clinical AI deployment. The CDSCO (Central Drugs Standard Control Organization) has begun developing regulatory frameworks for AI-based medical devices, with the IndAIDM (Indian AI for Drug and Medical Devices) initiative. [53,54,55]

Indian pharmaceutical companies — Sun Pharma, Dr. Reddy's, Lupin, Cipla, Aurobindo, Biocon, Glenmark, and others — are increasingly investing in AI capabilities for generics development, biosimilar engineering, drug repurposing, and discovery. Biocon's pioneering biosimilars program, Dr. Reddy's research collaborations with AI startups, and Lupin's AI initiatives represent industry adoption. The Indian generic and biosimilar industry — globally the largest generic drug producer — has unique opportunities to apply AI for cost-effective drug development serving Indian and global low-resource populations. [56,57]

Indian AI/pharma context priorities include: development of indigenous foundation models trained on Indian-population data (genomic diversity, disease prevalence, drug response); expansion of clinical AI deployment in primary, secondary, and tertiary care settings; AI-driven drug repurposing for neglected diseases and tropical infections (tuberculosis, leishmaniasis, dengue, leprosy) where market incentives are limited; pharmacogenomic implementation in clinical pharmacy practice; expansion of biomedical AI education in pharmacy, medical, and biotechnology curricula; ethical frameworks for AI in healthcare addressing privacy, consent, algorithmic fairness, and equity; and global collaboration leveraging Indian computational, biomedical, and clinical research capabilities.

9. REGULATORY AND ETHICAL FRAMEWORKS

Regulatory frameworks for AI in healthcare and pharmaceuticals are evolving rapidly. The FDA has approved over 800 AI/ML-based medical devices (predominantly in radiology, cardiology, and ophthalmology) and published draft guidance on AI/ML software as medical device (SaMD). The agency's 2024 framework for predetermined change control plans addresses the unique challenge of continuously learning AI systems. EU regulations under the EU AI Act (2024) classify medical AI as high-risk with substantial compliance requirements. The Indian CDSCO is developing parallel frameworks. Critical issues include validation requirements for AI/ML in regulatory submissions, transparency of AI-discovered candidates' design rationale, and post-market surveillance for adaptive AI systems. [58,59,60,61]

Ethical considerations include algorithmic fairness (avoiding discrimination across demographic groups), training data representativeness (ensuring diverse populations are represented), transparency and explainability (the 'black box' challenge), privacy and data governance (genomic, clinical, behavioral data protection), accountability for AI-driven decisions, equity in AI access, and societal effects on the practice of pharmacy and medicine. Professional guidelines from the AMA, AHA, ACP, FIP, and Indian medical and pharmacy associations increasingly address AI integration into practice.

SUMMARY TABLES

Table 1. Major AI Foundation Models in Biomedicine

Model	Developer	Domain	Application	Year/Status
AlphaFold 2	DeepMind	Protein structure	Near-experimental accuracy structure prediction	2020 (Nobel 2024)
AlphaFold 3	DeepMind/Isomorphic Labs	Protein-ligand/protein-protein/RNA complexes	Drug discovery, structural biology	May 2024
ESM-2	Meta AI	Protein language model	Sequence-based protein representation	2022
ProtT5/ProtBERT	RostLab	Protein language model	Function prediction, classification	2021
RFdiffusion	Baker Lab (UW)	Protein de novo design	Generative protein design (Nobel 2024)	2023
ProteinMPNN	Baker Lab (UW)	Protein sequence design	Sequence design for given backbone	2022
ChemBERTa/MolBERT	Various	Molecular property prediction	QSAR, ADMET	2020-2022
GPT-4	OpenAI	General LLM	Medical knowledge, documentation, decision support	2023
Med-PaLM 2	Google	Medical LLM	Medical question answering, clinical reasoning	2023
GatorTron / BioMedLM / ChemLLM	Various	Specialized biomedical LLMs	Various biomedical NLP	2022-2024

Table 2. AI-Discovered Drug Candidates in Clinical Development

Drug	Company	Target/Indication	Phase	Key Notes
INS018_055	Insilico Medicine	TNIK / IPF	Phase 2	AI-discovery to phase 1 in ~18 months
DSP-1181	Exscientia/Sumitomo	5-HT1A agonist / OCD	Discontinued	First AI-designed drug in clinic (2020); did not advance
REC-994	Recursion	CCM	Phase 2/3	Phenotypic AI discovery
REC-2282	Recursion	NF2-associated meningiomas	Phase 2/3	
REC-4881	Recursion	FAP	Phase 2	
GTAEXS617	Exscientia	CDK7 / oncology	Phase 1/2	
EXS-21546	Exscientia	A2A receptor / oncology	Phase 1	
ISM3091	Insilico Medicine	USP1 / oncology	Phase 1	
BEN-2293	BenevolentAI	Pan-Trk / atopic dermatitis	Phase 2	
RNAi from Atomwise hits	Various	Various oncology, antivirals	Various	Multiple programs
Isomorphic Labs internal pipeline	Isomorphic/DeepMind	Various	Preclinical	Major Lilly + Novartis partnerships

Table 3. LLM Capabilities and Limitations in Clinical Medicine

Capability	Performance	Strength	Limitation/Caveat
USMLE Step	GPT-4: ~86-90%	High knowledge baseline	Standardized exam, not real-world
Med-PaLM 2 MedQA	86.5%	Specialized medical training	Single-best-answer format limited
Clinical documentation	30-50% time savings	High utility, growing adoption	Accuracy verification still required
Differential diagnosis	Performance comparable to senior residents	Comprehensive coverage	Hallucination, missing rare conditions
Drug information	Generally accurate basic info	Wide coverage	May miss recent updates, dosing nuance
Pharmacovigilance signal	Emerging applications	Scalable text analysis	Validation in regulatory context limited
Patient education	Multilingual, accessible	Equity benefit potential	Verification, handoff design important
Calibration of uncertainty	Generally poor	Improving with techniques	Important for clinical reliability
Multimodal (images + text)	Rapidly advancing	Radiology, pathology applications	Domain-specific training needed

Table 4. AI Applications by Drug Discovery Stage

Stage	Key AI Applications	Lead Companies/Platforms
Target identification	Multi-omics integration, knowledge graphs, phenotypic screening	Recursion, BenevolentAI, Insilico, Insitro
Hit identification	Virtual screening, generative chemistry, active learning	Atomwise, Schrödinger, Iambic, Exscientia
Lead optimization	ML-guided SAR, generative design, ADMET prediction	Exscientia, Insilico Chemistry42, Schrödinger LiveDesign
Structural biology	Protein structure prediction, ligand-protein modeling	DeepMind AlphaFold, Isomorphic Labs, Baker Lab
ADMET prediction	Permeability, metabolism, toxicity prediction	OpenADMET, ADMETlab, commercial platforms
Clinical trial design	Patient recruitment, synthetic controls, adaptive designs	Tempus, ConcertAI, Medidata, Saama
Patient stratification	Genomic, multimodal biomarker classification	Tempus, Genospace, Foundation Medicine, Sema4
Pharmacovigilance	Signal detection, case processing, literature analysis	Saama, Cortellis, Inspirata, ArisGlobal
Real-world evidence	Electronic health record analysis, claims analysis	Aetion, Verana, Komodo, Flatiron
Manufacturing	Process optimization, quality control	NVIDIA BioNeMo, Synthace, Atomic Industries

Table 5. Indian AI/Biotech Ecosystem

Element	Status India	Notes
Major academic AI/biotech research centers	IITs, IISc, AIIMS, NCBS, CCMB, CCAMP, RCB	Growing programs, increasing publication output
Indian AI biotech startups	Strand, MedGenome, Qure.ai, Niramai, HealthifyMe, AarogyaAI, others	Growing ecosystem; substantial investment 2023-2025
Pharmaceutical company AI initiatives	Sun, DRL, Lupin, Cipla, Biocon, Aurobindo, Glenmark	Generics-focused; biosimilar engineering
Indian genome programs	Genome India (10,000 genomes), IndiGen, MedGenome IIN	Indian-population-specific data essential
Regulatory framework	CDSCO IndAIDM emerging	Following FDA/EU evolution
National Digital Health Mission	Active deployment	Foundation for health data ecosystem

Pharmacogenomic implementation	Limited; emerging at academic centers	Major implementation opportunity
Foundation model development	Limited indigenous capability	Major opportunity
AI in pharmacy/medical education	Emerging in select curricula	Significant gap; needs strengthening
Talent base	Strong CS/AI graduates; growing biomedical AI integration	Major strategic advantage

11. FUTURE DIRECTIONS

Several developmental trajectories will shape the next decade of AI in pharmacology and drug discovery. First, foundation models for biology — large pre-trained models analogous to GPT for language — will continue to advance, with models trained on multimodal biological data (sequences, structures, expression data, chemistry, clinical outcomes) providing more comprehensive biological reasoning capabilities. NVIDIA BioNeMo, ESM-3, and emerging multimodal models are early examples.

Second, the integration of AI with experimental robotics — closed-loop AI-experimental cycles in self-driving laboratories — will substantially accelerate drug discovery and biology research. Notable examples include the Recursion Operating System, Insitro's experimental platforms, and academic developments at Argonne National Laboratory and elsewhere. Third, agentic AI systems capable of multi-step reasoning, tool use, and autonomous task completion will enable more sophisticated drug discovery workflows, with claims of substantial productivity gains from Insilico and others.

Fourth, AI-enabled precision medicine will increasingly integrate genomic, clinical, biomarker, and lifestyle data for individualized treatment recommendations. The integration of pharmacogenomics into routine clinical practice — long discussed but slow to implement — may finally occur with AI-driven clinical decision support reducing the cognitive burden on clinicians. Fifth, AI applications in regulatory affairs, including AI-augmented regulatory submissions, automated label generation, and pharmacovigilance signal detection, will accelerate.

Sixth, the AI/healthcare regulatory landscape will continue evolving with new frameworks for adaptive AI systems, foundation models in medicine, and AI-discovered drug candidate review. Seventh, equity considerations will increasingly drive policy, with attention to ensuring AI benefits reach diverse global populations including India. Eighth, biomedical AI education in pharmacy and medical curricula will expand substantially, preparing future clinicians and researchers for AI-augmented practice. [62]

12. CONCLUSION

Artificial intelligence and large language models have transitioned from speculative possibility to operational reality across pharmacology, drug discovery, clinical pharmacy, and pharmacogenomics. The 2020 AlphaFold breakthrough and the 2022-2025 wave of generative AI and LLMs represent watershed moments in computational biomedicine. The 2024 Nobel Prize in Chemistry recognizing AlphaFold and computational protein design confirms the transformative significance. AI-native biotechnology companies — Insilico Medicine, Recursion, Isomorphic Labs, BenevolentAI, Exscientia, Atomwise, Iambic, Insitro, and others — have moved beyond proof-of-concept to clinical-stage assets with notable acceleration of discovery timelines.

LLM applications in clinical practice — clinical documentation, drug information, pharmacovigilance, decision support, patient education — are rapidly adopted with substantial productivity benefits and quality-of-care implications. Pharmacogenomic applications are advancing with machine learning-enhanced variant interpretation, polygenic risk scoring, and clinical decision support integration. The Indian AI/biotech

ecosystem is rapidly developing with strong academic foundations, growing startup ecosystem, generic and biosimilar industry adoption, and national initiatives supporting genomics and digital health infrastructure.

Challenges include validation of AI tools in real-world clinical settings, regulatory frameworks for adaptive AI systems and foundation models, equity and fairness considerations, training data representativeness for diverse populations including South Asians, ethical considerations around privacy and accountability, and integration with established clinical workflows. The Indian context offers both substantial opportunities (large talent base, generics industry, demographic diversity, growing health data infrastructure) and challenges (regulatory frameworks development, foundation model gaps, education curriculum updating).

For Indian pharmacology and pharmacy education, AI represents perhaps the most important paradigm shift in pharmaceutical practice since the introduction of randomized clinical trials and the molecular biology revolution. Pharmacology educators should integrate AI and LLM topics into the curriculum addressing the technical foundations, the major applications in drug discovery and clinical practice, the ethical and regulatory considerations, and the practical skills required for the AI-augmented future of pharmacy. Indian pharmacy and medical professionals — many of whom will be working in an AI-saturated practice environment within 5-10 years — must be prepared not just to use AI tools but to critically evaluate them, contribute to their development, and ensure they serve diverse Indian patient populations equitably. The AI/LLM era represents not just a new set of tools but a fundamental transformation of pharmaceutical practice, and Indian pharmacology education must rise to this transformative moment.

Author Contributions

Adiba Begum conceived the review topic, conducted the literature search, drafted the manuscript, and prepared the tables. R. L. Manisha (corresponding author) supervised the work, critically reviewed and revised the manuscript, and finalized the scientific content. Muvvala Sudhakar provided institutional oversight, critically reviewed the manuscript, and approved the final version. All authors read and approved the final manuscript.

Declaration by Authors

Ethical Approval: Not applicable

Acknowledgement: The authors thank the Department of Pharmacology and the Management of Malla Reddy College of Pharmacy, Secunderabad, for institutional support, and the library and information resource center for facilitating literature access.

Source of Funding: No external funding was received for the preparation of this review.

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

REFERENCES

1. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-477.
2. Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. JAMA. 2020;323(9):844-853.
3. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ. 2016; 47:20-33.
4. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9(3):203-214.
5. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019;20(2):273-286.
6. Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Drug Discov Today. 2019;24(3):773-780.
7. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.
8. Varadi M, Anyango S, Deshpande M, et al. AlphaFold Protein Structure Database:

massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50(D1): D439-D444.
9. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172-180.
10. Lee P, Goldberg C, Kohane I. The AI Revolution in Medicine: GPT-4 and Beyond. Pearson; 2023.
11. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
12. Ren F, Aliper A, Chen J, et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. Nat Biotechnol. 2025;43(1):63-75.
13. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017; 30:5998-6008.
14. Sapoval N, Aghazadeh A, Nute MG, et al. Current progress and open challenges for applying deep learning across the biosciences. Nat Commun. 2022;13(1):1728.
15. Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023;379(6637):1123-1130.
16. Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. arXiv preprint. 2021.
17. Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. Nat Commun. 2022;13(1):4348.
18. Madani A, Krause B, Greene ER, et al. Large language models generate functional protein sequences across diverse families. Nat Biotechnol. 2023;41(8):1099-1106.
19. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500.
20. Watson JL, Juergens D, Bennett NR, et al. De novo design of protein structure and function with RFdiffusion. Nature. 2023;620(7976):1089-1100.
21. Dauparas J, Anishchenko I, Bennett N, et al. Robust deep learning-based protein sequence design using ProteinMPNN. Science. 2022;378(6615):49-56.
22. Pun FW, Ozerov IV, Zhavoronkov A. AI-powered therapeutic target discovery. Trends Pharmacol Sci. 2023;44(9):561-572.
23. Schneider P, Walters WP, Plowright AT, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-364.
24. Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018;17(2):97-113.
25. Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688-702.
26. Wong F, de la Fuente-Nunez C, Collins JJ. Leveraging artificial intelligence in the fight against infectious diseases. Science. 2023;381(6654):164-170.
27. Singhal K, Tu T, Gottweis J, et al. Toward expert-level medical question answering with large language models. Nat Med. 2025;31(3):943-950.
28. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347-1358.
29. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317-1318.
30. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29.
31. Liu PR, Lu L, Zhang JY, et al. Application of artificial intelligence in medicine: an overview. Curr Med Sci. 2021;41(6):1105-1115.
32. Tu T, Schaekermann M, Palepu A, et al. Towards conversational diagnostic AI. arXiv preprint. 2024.
33. Saab K, Tu T, Weng WH, et al. Capabilities of Gemini models in medicine. arXiv preprint. 2024.
34. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA. 2023;330(1):78-80.
35. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open. 2024;7(10): e2440969.
36. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nat Med. 2022;28(1):31-38.
37. Pant K, Sharma A. Artificial intelligence in pharmacovigilance: current applications and future trends in India. Pharmacovigilance J. 2024;5(2):45-52.
38. Stade EC, Stirman SW, Ungar LH, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. NPJ Ment Health Res. 2024;3(1):12.
39. Insilico Medicine. INS018_055 phase 2a trial demonstrates safety and tolerability in idiopathic pulmonary fibrosis [press release]. 2024.
40. Heinrich M, Rabault A, Lopes L, et al. Phenotypic screening using deep learning at

scale: applications and limitations. Drug Discov Today. 2023;28(2):103480.
41. Caudle KE, Klein TE, Hoffman JM, et al. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. Curr Drug Metab. 2014;15(2):209-217.
42. Martin AR, Kanai M, Kamatani Y, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-591.
43. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-1224.
44. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet. 2019;28(R2): R133-R142.
45. Marigorta UM, Rodríguez JA, Gibson G, et al. Replicability and prediction: lessons and challenges from GWAS. Trends Genet. 2018;34(7):504-517.
46. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010;363(2):166-176.
47. Singh A, Kumar S, Mishra S. Pharmacogenomic considerations in the Indian population: a review. Int J Mol Sci. 2023;24(9):8237.
48. Brindley DA. UK 100,000 Genomes Project: implementing the largest single genome sequencing initiative in the world. Front Genet. 2014; 5:308.
49. All of Us Research Program Investigators. The 'All of Us' research program. N Engl J Med. 2019;381(7):668-676.
50. Stenzinger A, Edsjö A, Ploeger C, et al. Trailblazing precision medicine in Europe: a joint view by Genomic Medicine Sweden and the Centers for Personalized Medicine, ZPM, in Germany. Semin Cancer Biol. 2022; 84:242-254.
51. Vamadeva S, Kumar R, Sharma N. Pharmacogenomic testing in India: current status, challenges, and the path forward. Indian J Pharmacol. 2024;56(2):93-101.
52. Sengupta D, Choudhury A, Basu A, et al. Population stratification and underrepresentation of Indian subcontinent genetic diversity in the 1000 Genomes Project dataset. Genome Biol Evol. 2016;8(11):3460-3470.
53. Department of Biotechnology, Government of India. Genome India Programme. 2023-2024.
54. Mukherjee S, Mahapatra K. India's National AI Mission and biomedical applications: an analysis. Health Policy Plan. 2024;39(4):412-419.
55. Tomar SK, Sharma D. Indigenous AI capabilities in Indian healthcare: a roadmap. Indian J Public Health. 2024;68(4):587-594.
56. Saha S, Bhargava K. Indian pharmaceutical industry: emerging opportunities and challenges. Indian J Pharm Educ. 2024;58(1):1-12.
57. Nair AB, Atal S, Sadasivam B, et al. Artificial intelligence implementations in Indian pharmaceutical industry. Indian J Pharmacol. 2023;55(3):162-169.
58. U.S. Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence/machine learning enabled device software functions. Draft guidance for industry. April 2023.
59. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med. 2018;378(11):981-983.
60. Chen IY, Pierson E, Rose S, et al. Ethical machine learning in healthcare. Annu Rev Biomed Data Sci. 2021; 4:123-144.
61. Cresswell K, Williams R, Sheikh A. Developing and applying a formative evaluation framework for health information technology implementations. JMIR Med Inform. 2020;8(6): e15068.
62. Kothari U, Joshi R, Patil A. Generative artificial intelligence in pharmacy education: opportunities and challenges. Indian J Pharm Educ Res. 2024;58(2):298-307.

How to cite this article: Adiba Begum, R. L. Manisha, Muvvala Sudhakar. Artificial intelligence and large language models in drug discovery and pharmacogenomics: a comprehensive pharmacological review of Alphafold, generative AI for de novo design, clinical decision support, and the translational pipeline. *International Journal of Research and Review*. 2026; 13(6): 93-106. DOI: *https://doi.org/10.52403/ijrr.20260610*

******