



Artificial Intelligence in Pharmaceutical Research and Development: A Comprehensive Review

Bodake Ashvin Pandurang¹, Radhika S. Subhedar², Nilesh B. Chougule³

¹Student B. Pharmacy, ²Assistant Professor, ³Principal and HOD (Pharmacology),
Ashokrao Mane Institute of Pharmacy, Ambap, Kolhapur, Maharashtra, India (416 112)
Email ID: ashubodake45@gmail.com

ABSTRACT:

The global pharmaceutical industry faces a productivity crisis marked by rising development costs, often estimated to exceed \$2.6 billion per new molecular entity, and timelines that extend beyond a decade. Artificial intelligence (AI), including machine learning (ML), deep learning (DL) and generative AI, has emerged as a rational, data-driven alternative to the traditional trial-and-error approach. This review discusses the application of AI across the pharmaceutical value chain, covering drug discovery, preformulating, solid dosage form engineering, clinical trials, personalized medicine and manufacturing (Pharma 4.0). It also examines the regulatory landscape of biosimilars, highlighting the divergence in global standards and the role of analytical convergence in harmonizing approval pathways. Case studies such as the polymorphism of Nicergoline, the hydrolysis of Zidovudine and the generative design of microcrystalline cellulose networks illustrate how these tools are applied in practice. The challenges of model interpretability, data quality and regulatory acceptance, along with future directions such as federated learning and quantum computing, are also discussed.

KEYWORDS: Artificial intelligence, machine learning, drug discovery, preformulation, biosimilars, Pharma 4.0.

INTRODUCTION:

Over the last two decades, scientific understanding of disease biology, genomics and proteomics has grown exponentially, yet the efficiency of converting this knowledge into approved therapies has stagnated or even declined. This trend is described as Froom's Law (Moore's Law spelt backwards), which observes that the number of new drugs approved per billion dollars of research spending has fallen by roughly half every nine years.¹ Bringing a single new drug to market typically requires 10 to 15 years and a financial outlay exceeding \$2.6 billion.²

This inefficiency arises largely from the complexity of biological systems and the stochastic nature of conventional discovery, which has relied on serendipity and labor-intensive high-throughput screening (HTS) of large chemical libraries.^{1,2} Such brute-force screening often ends in late-stage failure because of unforeseen toxicity or lack of efficacy.^{1,2} Formulation, clinical testing and manufacturing have also operated in operational silos, using retrospective validation rather than prospective design, so problems such as poor solubility or instability are identified only after a lead candidate has been selected.^{1,2}

AI, defined as the capability of machines to mimic human cognitive functions such as learning, problem solving and pattern recognition, is now regarded as a foundational infrastructure of pharmaceutical science rather than a supplementary tool.³ Artificial neural networks (ANNs), deep neural networks (DNNs) and recurrent neural networks (RNNs) allow intelligent modeling of complex biological and physicochemical phenomena that lie beyond human intuition and traditional statistical methods.³ AI is reshaping the industry by accelerating discovery through virtual screening and *de novo* design, optimizing formulation through predictive preformulation and digital twin modeling, improving manufacturing through real-time process control and automated defect detection, and supporting the regulation of biologics and biosimilars through the totality-of-evidence approach.³ This review discusses each of these areas in turn.

HISTORICAL EVOLUTION OF AI:

The intellectual roots of AI lie in the mid-20th century. Warren McCulloch and Walter Pitts proposed artificial neurons in 1943, establishing the mathematical basis of neural networks,⁴ and Alan Turing proposed the Turing Test in 1950 as a criterion for judging whether a machine's behavior is indistinguishable from that of a human.⁴ The term "Artificial Intelligence" was coined by John McCarthy at the Dartmouth Conference in 1956, which is widely considered the birth of the field.⁴ Progress was interrupted by two "AI winters", periods of skepticism and reduced funding after early systems failed to meet inflated expectations, the first between about 1974 and 1980 and the second between 1987 and 1993.⁴

The resurgence of AI since the 2010s, often called the “AI summer”, was driven by three factors: the growth of big data (genomics, proteomics and electronic health records), more efficient algorithms such as deep learning and backpropagation, and greater computing power, particularly through graphics processing units (GPUs).⁴ Milestones such as IBM’s Deep Blue defeating Garry Kasparov in 1997 and IBM’s Watson winning the quiz show *Jeopardy* in 2011 showed that AI was ready for complex, unstructured tasks relevant to healthcare and life sciences.⁴

CLASSIFICATION OF AI:

Distinguishing between the types of intelligence systems helps to set realistic expectations of what the technology can and cannot do in pharmaceutical applications.⁵ AI is classified by caliber and by functionality.

Based on caliber:

1. **Artificial Narrow Intelligence (ANI):** Also called “weak AI”, this is the current state of the art in the industry. These systems are highly proficient at single tasks such as predicting protein folding, optimizing a synthetic route or identifying tablet defects, but they lack general adaptability and cannot transfer learning from one domain to another without retraining.⁵
2. **Artificial General Intelligence (AGI):** A theoretical state in which machines possess human-level cognitive abilities across a wide range of tasks. AGI does not currently exist.⁵
3. **Artificial Super Intelligence (ASI):** A hypothetical future state in which machine intellect surpasses human capability in all fields, including creativity, general wisdom and social skills.⁵

Based on functionality:

1. **Type 1 (Reactive machines):** The most basic form of AI, analyzing only the current situation without memory of past experience. IBM’s Deep Blue is a classic example.⁵
2. **Type 2 (Limited memory):** Systems that use historical data to make predictions. This is the dominant form of AI in pharma today; for example, an algorithm that predicts drug stability from a training set of historical degradation data. Predictive maintenance systems also belong to this category.⁵
3. **Type 3 (Theory of mind):** Systems that would understand human emotions. Not yet in existence.⁵
4. **Type 4 (Self-awareness):** Systems that would possess consciousness. Not yet in existence.⁵

AI IN DRUG DISCOVERY:

The initial phase of discovery is the most attrition-prone and resource-intensive stage of the pipeline. Traditionally it relies on HTS of physical chemical libraries against a biological target, which yields low hit rates and often fails to identify candidates with optimal druggability.^{1,2} AI alters this paradigm by shifting screening *in silico* and enabling rational design.

Target identification and protein structure prediction: AI algorithms can integrate multi-omics data, including genomics, proteomics and metabolomics, to build robust disease models and identify targets that are causally linked to disease pathology rather than merely correlated with it.⁶ Knowing the three-dimensional structure of a target is essential for designing ligands that bind it effectively. Experimental determination by X-ray crystallography or cryo-electron microscopy (Cryo-EM) can take years for a single protein. DeepMind’s AlphaFold uses deep learning to predict 3D protein structures from amino acid sequences with near-experimental accuracy, which accelerates the understanding of protein–ligand interactions and allows rational drug design against targets once considered “undruggable” because of a lack of structural data.⁷

Virtual screening: ML algorithms can rapidly screen virtual libraries containing billions of compounds and predict the biological activity of untested compounds, along with their physicochemical properties (quantitative structure–property relationship, QSPR) and activity (quantitative structure–activity relationship, QSAR).⁸ This is faster and more cost-effective than physical screening and allows researchers to prioritize only the most promising candidates for synthesis and testing.⁸

De novo design: Generative models such as generative adversarial networks (GANs) can create entirely new molecular structures. A GAN consists of two competing neural networks: a generator that proposes molecular structures and a discriminator that evaluates them against desired criteria such as binding affinity, solubility and low toxicity. Through repeated iterations the system evolves molecules optimized for specific pharmacological and safety profiles and explores chemical space more efficiently than human chemists.⁹

Drug repurposing: AI can identify new therapeutic indications for existing, approved drugs by comparing the molecular signatures of drugs with the biological networks of different diseases. Because the safety profiles of these drugs are already well established, this strategy substantially reduces development time and cost compared with *de novo* discovery.¹⁰

PREFORMULATION STUDIES:

After a lead candidate has been identified, the molecule enters the preformulation phase, which serves as the interface and roadmap between the new drug entity and final formulation development. Its main objective is to generate comprehensive physicochemical data on the drug substance, its potential salts and its compatibility with excipients.¹¹

Solubility and BCS classification: A large percentage of new molecular entities (NMEs) fall into Biopharmaceutics Classification System (BCS) Class II (low solubility, high permeability) or Class IV (low solubility, low permeability).¹¹ Poor solubility severely limits bioavailability and complicates formulation design, which calls for advanced engineering strategies.¹¹

Physicochemical characterization: The active pharmaceutical ingredient (API) is characterized using a battery of analytical techniques, including:¹²

1. **Differential scanning calorimetry (DSC) and thermogravimetry (TG):** thermal methods used to determine melting point, thermal stability and phase transitions.¹²
2. **X-ray powder diffractometry (XRPD):** used to identify crystallinity and distinguish between polymorphic forms.¹²

Polymorphism, the ability of a substance to exist in more than one crystalline form, is a Critical Quality Attribute (CQA) that can profoundly affect solubility, stability and bioavailability. In the case of the poorly water-soluble drug Nicergoline, preformulation studies identified two polymorphic forms, and characterization showed that Form II was structurally more suitable for pharmaceutical use than Form I, which demonstrates the value of identifying the optimal polymorph early in development.¹³

Stability and drug–excipient compatibility: Stability studies guide the development of a stable product by identifying degradation pathways and incompatibilities.

1. **Zidovudine:** Studies on Zidovudine derivatives revealed susceptibility to base-catalyzed hydrolysis, which warns formulators to avoid alkaline environments or excipients that may raise the micro-environmental pH.¹²
2. **Benznidazole:** A chemical reaction was evidenced between Benznidazole and anhydrous lactose, driven by the free hydroxyl groups of the excipient and likely of the Maillard type. Lactose must therefore be excluded from Benznidazole formulations to ensure stability.¹⁴

Such early insights are essential for predicting shelf life and meeting the regulatory requirements of the Chemistry, Manufacturing and Controls (CMC) section of applications such as Investigational New Drug (IND) and New Drug Application (NDA) filings.¹¹

AI-driven predictive modeling: Preformulation has traditionally been experiment-heavy. AI-optimized predictive modeling, built on molecular descriptor datasets, can now predict the behavior of drug molecules and their interactions with excipients without extensive physical testing.¹⁵ In a comparison of ML algorithms, DNNs outperformed traditional models such as linear regression, random forest and gradient boosting, achieving a coefficient of determination (R^2) of 0.90 for solubility and 0.80 for stability (Table 1).¹⁵ This suggests that deep architectures better capture the complex, non-linear relationships between molecular structure and physicochemical behavior. Feature importance analysis identified the topological polar surface area (TPSA) as the most influential descriptor for both solubility and stability. TPSA correlates with hydrogen-bonding capacity and polarity, which govern a molecule's interaction with solvents and excipients, so candidates can be filtered on TPSA *in silico* before any material is synthesized.¹⁵

Table 1: Performance of AI models in preformulation studies

Model type	Application	R^2	Key feature identified
Deep neural network (DNN)	Solubility prediction	0.90	TPSA
Deep neural network (DNN)	Stability prediction	0.80	TPSA
Linear regression, random forest, gradient boosting	Solubility / stability	Lower than DNN	Not applicable

TPSA: topological polar surface area. Conventional models performed below the DNN in the same comparison.¹⁵

SOLID DOSAGE FORM DEVELOPMENT:

AI is changing the development of solid dosage forms such as tablets, capsules and implants, a phase that has long depended on time-consuming, labor-intensive trial and error.¹¹

Generative AI and digital twins: Generative AI for *in silico* structure synthesis acts as an image-informed, AI-powered structure engineer. Using images of sub-optimal formulations, such as those obtained by X-ray microtomography (XRM), the AI creates “digital twins” of the drug product, which allow scientists to simulate how microstructural changes affect performance without manufacturing physical prototypes.¹⁶ The generative model is guided by

interpretable CQAs, such as particle size distribution and drug loading, and can synthesize structural features outside the range of its training data, thereby exploring formulation spaces that human intuition might overlook.¹⁶

Case studies of *in silico* optimization: The approach has been validated in two case studies.

1. **Oral tablets and the percolation threshold:** Connectivity of the pore network is crucial for disintegration and dissolution. The method was applied to a tablet containing microcrystalline cellulose (MCC) and accurately predicted the percolation threshold of the MCC matrix at 4.2% weight. This is the critical point at which MCC particles form a connected network, and it determines the mechanical strength and dissolution properties of the tablet. Predicting it digitally avoids manufacturing dozens of batches with different MCC concentrations.¹⁶
2. **Long-acting HIV inhibitor implants:** In this more complex delivery system, drug release is governed by the spatial distribution of drug particles within the polymer matrix. The AI was used to optimize this distribution to achieve a desired sustained-release profile, which shows that generative AI is applicable to complex controlled-release systems as well as simple tablets.¹⁶

Limitation: the realizability gap: The main limitation of generative structure synthesis is the gap between theoretical optima and practical realization. The AI may generate a microstructure with ideal dissolution kinetics *in silico* that is not always practically realizable with existing manufacturing processes; for example, a highly ordered particle distribution may require precision that standard tablet presses or extruders cannot deliver. Decision-makers must therefore remain aware of the gap between the AI-generated ideal and the real capabilities of the manufacturing line.¹⁶

Advanced solubility enhancement technologies: Preformulation findings also guide the choice of manufacturing technology for solubility enhancement. Hot melt extrusion (HME) can be detrimental for heat-sensitive drugs because of thermal degradation of the API.

1. **KinetiSol dispersing (KSD):** An alternative to HME for thermally labile drugs. KSD uses high shear and friction to generate the heat required for fusion, which markedly reduces residence time and processing temperature. For heat-sensitive compounds such as a Roche research compound and Hydrocortisone, KSD minimized decomposition and produced amorphous solid solutions with enhanced potency, whereas HME caused degradation.^{17,18}
2. **Cyclodextrin complexation:** For drugs such as Nifedipine and Irbesartan, co-evaporation was identified as a highly effective method of preparing inclusion complexes that improve solubility.^{19,20}

BIOSIMILARS: REGULATORY LANDSCAPE AND ANALYTICAL CONVERGENCE:

Biologics represent a large and growing share of the global pharmaceutical market. Biosimilars, which are biological products designed to be highly similar to approved reference products, are therefore important for global health equity and cost containment.²¹

Economic and access imperative: Biosimilars promote price competition and thereby improve the affordability and accessibility of vital treatments, particularly for chronic autoimmune diseases and oncology. It has been estimated that they could save the U.S. healthcare system up to \$100 billion over the next decade, depending on market competition and adoption rates.²¹ Their development remains expensive, however, often costing \$100 to \$300 million and taking six to nine years.²¹

Regulatory divergence: A major barrier to the global expansion of biosimilars is the lack of regulatory harmonization. A systematic analysis of the regulations of eight countries (the US, Japan, South Korea, Canada, Brazil, Argentina, Australia and South Africa) revealed inconsistent guidelines that complicate global development programs.²¹ The key areas of divergence are:

1. **Reference product (RP) selection:** Jurisdictions differ on whether clinical trials must use a locally sourced RP or may use a foreign-sourced one. A local RP requirement necessitates redundant bridging studies to show that the foreign and local products are identical, which significantly increases development cost.²¹
2. **In vivo animal studies:** The necessity of animal toxicity testing is increasingly debated. The US FDA, empowered by the 2022 FDA Modernization Act 2.0, and the UK MHRA (2021) have revised their guidelines to largely remove or reduce the requirement for animal studies, citing their limited predictive value for immunogenicity in humans. Other regions may still require them by default, which forces global developers to run unnecessary animal studies to satisfy the most conservative regulator and raises ethical and financial concerns.^{21,22}

Totality-of-evidence and analytical convergence: Despite clinical divergence, there is strong global consensus on the totality-of-evidence approach, in which bio similarity is established from a composite of data beginning with a comprehensive analytical assessment.^{21,23} Regulatory requirements show minimal divergence for this assessment, which focuses on comparing CQAs (the physicochemical and functional attributes that affect safety and efficacy) using state-of-the-art orthogonal methods based on different scientific principles.^{21,23} The robustness of modern analytical science gives a high degree of confidence early in development and has led to the argument that large comparative efficacy studies in patients may be redundant for confirming bio similarity. The 2022 WHO guidelines reflect this by stating that a comparative efficacy trial may not be necessary when sufficient evidence is established through analytical and pharmacokinetic/pharmacodynamic (PK/PD) studies.^{21,23}

Role of AI in biosimilar development: AI supports the totality-of-evidence approach and accelerates development in several ways:²¹

1. **Bioequivalence and stability:** ML and DL models analyze large volumes of preclinical and clinical data to improve bioequivalence studies and predict molecular stability, reducing the risk of late-stage failure.²¹

2. **Automated characterization:** Convolutional neural networks (CNNs) analyze biological images and classify cellular and molecular structures, allowing precise, automated comparison of a biosimilar with the originator biologic to detect even minute structural differences.²¹
3. **Patent landscape:** Natural language processing (NLP) reviews large volumes of scientific literature and patents, helping developers assess recent advances and navigate the “patent thickets” built around originator products.²¹

Nocebo effect and clinical adoption: Despite proven efficacy, the adoption of biosimilars in oncology practice has been slow. Besides regulatory caution, this is attributed to the nocebo effect, in which negative treatment outcomes are driven by a patient’s negative expectations, often fueled by poor communication and a lack of knowledge that leads patients to equate lower price with lower quality. Comprehensive educational programs for patients and physicians are needed to build confidence in biosimilars.²¹

AI IN CLINICAL TRIALS AND PERSONALIZED MEDICINE:

Clinical trials are the most expensive and time-consuming phase of drug development. AI-driven predictive analytics are used to optimize trial protocols and designs, and this has been reported to reduce trial duration by 25–30%.²⁴ Patient recruitment is a major bottleneck. AI algorithms analyze electronic health records (EHRs), genetic profiles and real-world evidence (RWE) to stratify patients more accurately, which can improve recruitment rates by as much as 35% and ensures that participants match the inclusion and exclusion criteria and are most likely to respond.²⁴

The ultimate goal of AI in pharma is precision medicine, in which treatment is tailored to the individual. AI builds detailed patient profiles by integrating genomic, clinical, lifestyle and environmental data.²⁵ Its main applications are:

1. **Treatment selection:** Selection based on pharmacogenomic factors, predicting which patients will respond to a drug and which may experience adverse effects.²⁵
2. **Dose optimization:** Calculation of the optimal dose for an individual to maximize benefit and minimize toxicity.²⁵
3. **Holistic patient view:** Incorporation of social determinants of health (such as socioeconomic status) and wearable-device data (physical activity, sleep) to predict treatment efficacy more accurately.²⁵

PHARMACEUTICAL MANUFACTURING:

PHARMA 4.0: The application of AI in manufacturing, often termed Pharma 4.0, aims at high efficiency, consistency and rigorous quality control through automation and data analytics.

Process optimization and PAT: AI enables continuous manufacturing through Process Analytical Technology (PAT). DL models analyze sensor data such as temperature, pressure and flow rate in real time and continuously adjust critical process parameters, so that the product consistently meets its CQAs in line with Quality by Design (QbD) principles. Moving from batch processing to continuous monitoring can significantly reduce waste and improve yield.²⁶

Computer vision in quality control: Human visual inspection of parenteral products and tablets is slow, subjective and prone to fatigue. AI-powered computer vision systems use CNNs for automated inspection and can detect subtle defects such as cracks, discoloration or foreign particulates that may escape human detection. Industry data indicate success rates of up to 95% for quality assurance with such systems.²⁷

Supply chain and predictive maintenance: ML models predict demand and manage inventory, as in Novartis’s algorithm-based procurement platform. AI also supports predictive maintenance by analyzing sensor data to anticipate equipment failure before it occurs, preventing costly unplanned downtime.²⁸

OVERVIEW OF TRADITIONAL VERSUS AI-ENABLED DEVELOPMENT:

Table 2 compares the traditional and AI-enabled approaches across the stages of pharmaceutical development discussed above.

Table 2: Comparative analysis of traditional and AI-enabled pharmaceutical development

Development phase	Traditional approach	AI-enabled approach	Key AI technologies	Impact
Discovery	Physical HTS; serendipity-driven	Virtual screening; <i>de novo</i> design; protein structure prediction	GANs, AlphaFold, ML classifiers	Faster hit-to-lead; wider chemical space explored

Development phase	Traditional approach	AI-enabled approach	Key AI technologies	Impact
Preformulation	Trial and error; labor-intensive physical testing	Predictive modeling; <i>in silico</i> solubility and stability profiling	DNNs, QSPR/QSAR	CQAs predicted without material consumption
Formulation	Iterative physical prototyping; univariate analysis	Generative structure synthesis; digital twins	Generative AI, CNNs	Optimized microstructure; percolation threshold predicted
Clinical trials	Manual recruitment; broad inclusion criteria	EHR analysis; patient stratification; precision dosing	NLP, predictive analytics	Shorter trials (25–30%); improved recruitment
Manufacturing	Batch processing; post-production QC testing	Continuous manufacturing; real-time PAT; automated visual inspection	Computer vision, deep learning	Real-time control; up to 95% success in QA; less waste
Biosimilars	<i>In vivo</i> animal studies; redundant bridging studies	Analytical assessment; totality-of-evidence; <i>in silico</i> comparison	ML, NLP (patent analysis)	Reduced animal testing; faster regulatory submission

HTS: high-throughput screening; PAT: process analytical technology; QC: quality control; QA: quality assurance.

ADVANTAGES OF AI IN PHARMACEUTICAL DEVELOPMENT:

1. Rapid screening of billions of virtual compounds, with lower cost than physical screening.⁸
2. Prediction of solubility and stability *in silico*, without consuming material.¹⁵
3. Digital twins allow formulation changes to be simulated without physical prototypes.¹⁶
4. Shorter clinical trials and improved patient recruitment.²⁴
5. Real-time process control, reduced waste and automated defect detection in manufacturing.^{26,27}
6. Support for the totality-of-evidence approach in biosimilar development.²¹

LIMITATIONS OF AI IN PHARMACEUTICAL DEVELOPMENT:

1. Black-box behavior of deep learning models limits interpretability and regulatory acceptance.²⁹
2. Model performance depends on the quality, completeness and representativeness of training data.⁶
3. Data privacy concerns and the risk of algorithmic bias in clinical datasets.⁶
4. AI-generated optimal structures may not be realizable with existing manufacturing processes.¹⁶
5. Divergent global regulatory requirements, particularly on animal testing and reference products.^{21,22}

CHALLENGES AND FUTURE PERSPECTIVES:

The “black box” problem and Explainable AI (XAI): DL models, particularly deep neural networks, often work as “black boxes” in which the input leads to an output but the internal logic remains opaque. In a regulated industry, knowing why a model predicts a molecule to be toxic or safe is as important as the prediction itself. This opacity limits the usefulness of AI in guiding rational, mechanism-based drug design and hampers regulatory acceptance. The emerging field of Explainable AI (XAI) aims to build trust by providing transparent, evidence-based reasoning for complex predictions.²⁹

Data quality and governance: AI models are only as good as the data on which they are trained, so high-quality, unbiased and comprehensive datasets are essential. Missing values must be handled carefully, for example through normalization, before training. Regulatory and ethical considerations such as data privacy (for example, HIPAA compliance) and the prevention of algorithmic bias in clinical datasets call for rigorous data governance and ethical guidelines that ensure fair data representation and decision-making.⁶

Future directions: The future of AI in pharma lies in the convergence of technologies. Quantum computing may solve molecular simulation problems that are currently intractable for classical computers. Federated learning would allow companies to collaborate on training models without sharing proprietary raw data, thereby preserving privacy. Graph neural networks (GNNs) are emerging as a powerful tool for analyzing chemical structures and component interactions as relational data, offering deeper insight into molecular behavior.³⁰⁻³²

CONCLUSION:

The pharmaceutical industry is undergoing a paradigm shift from empirical observation and serendipity to computational prediction and rational design. AI, from generative models in molecule design to computer vision in manufacturing, offers a promising answer to the high costs, high attrition rates and long timelines of traditional drug development. The evidence reviewed here shows that AI is already delivering tangible value: it predicts physicochemical properties such as solubility with high accuracy ($R^2 = 0.90$), synthesizes novel solid dosage structures such as MCC networks, improves clinical trial recruitment by up to 35%, and streamlines biosimilar development through analytical convergence. Realizing the full potential of the technology will require addressing the black-box interpretability problem, harmonizing global regulatory standards, particularly on animal testing and reference products, and ensuring the ethical use of patient data. As the industry moves toward Pharma 4.0, the synergy between human expertise and machine intelligence will shape the next generation of therapeutics.

REFERENCES:

1. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discover.* 2012;11(3):191–200.
2. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ.* 2016; 47:20–33.
3. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discover Today.* 2021;26(1):80–93.
4. Haenlein M, Kaplan A. A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. *Calif Manage Rev.* 2019;61(4):5–14.
5. Hintze A. Understanding the four types of AI, from reactive robots to self-aware beings. *The Conversation* [Internet]. 2016 [cited 2026 Oct 2]. Available from: <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>
6. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* 2019;18(6):463–77.
7. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9.
8. Gawehn E, Hiss JA, Schneider G. Deep learning in drug discovery. *Mol Inform.* 2016;35(1):3–14.
9. Kadurin A, Nikolenko S, Khrabrov K, Aliper A, Zhavoronkov A. drug AN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Mol Pharm.* 2017;14(9):3098–104.
10. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov.* 2019;18(1):41–58.
11. Amidon GL, Lenners' H, Shah VP, Crison JR. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm Res.* 1995;12(3):413–20.
12. Raviolo MA, Briñón MC. Preformulation studies of zidovudine derivatives: acid dissociation constants, differential scanning calorimetry, thermogravimetry, X-ray powder diffractometry and aqueous stability studies. *Sci Pharm.* 2011;79(3):479–92. doi:10.3797/scipharm.1105-04.
13. Malaj L, Censi R, Capsoni D, Pellegrino L, Bini M, Ferrari S, et al. Characterization of nicergoline polymorphs crystallized in several organic solvents. *J Pharm Sci.* 2011;100(7):2610–22.
14. Soares-Sobrinho JL, et al. A preformulation study of a new medicine for Chagas disease treatment: physicochemical characterization, thermal stability, and compatibility of benznidazole. *AAPS PharmSciTech.* 2010;11(3):1391–6.
15. Artificial intelligence in pharmaceutical preformulation: a paradigm shift towards predictive drug development.
16. Hornick T, Mao C, Koynov A, Yawman P, Thool P, Salish K, et al. In silico formulation optimization and particle engineering of pharmaceutical products using a generative artificial intelligence structure synthesis method. *Nat Commun.* 2024;15(1). doi:10.1038/s41467-024-54011-9.

17. Hughey JR, DiNunzio JC, Bennett RC, Brough C, Miller DA, Ma H, et al. Dissolution enhancement of a drug exhibiting thermal and acidic decomposition characteristics by fusion processing: a comparative study of hot melt extrusion and KinetiSol dispersing. *AAPS PharmSciTech*. 2010;11(2):760–74.
18. DiNunzio JC, Brough C, Miller DA, Williams RO 3rd, McGinity JW. Fusion production of solid dispersions containing a heat-sensitive active ingredient by hot melt extrusion and KinetiSol dispersing. *Eur J Pharm Biopharm*. 2010;74(2):340–51.
19. Jagdale SC, Jadhav VN, Chabukswar AR, Kuchekar BS. Solubility enhancement, physicochemical characterization and formulation of fast-dissolving tablet of nifedipine-betacyclodextrin complexes. *Braz J Pharm Sci*. 2012;48(1). doi:10.1590/S1984-82502012000100015.
20. Dudhipala N, Ettireddy S, Youssef AAA, Puchchakayala G. Cyclodextrin complexed lipid nanoparticles of irbesartan for oral applications: design, development, and in vitro characterization. *Molecules*. 2021;26(24):7538. doi:10.3390/molecules26247538.
21. Bas TG, Duarte V. Biosimilars in the era of artificial intelligence – international regulations and the use in oncological treatments. *Pharmaceuticals (Basel)*. 2024;17(7):925. doi:10.3390/ph17070925.
22. FDA Modernization Act 2.0. Pub. L. No. 117-328, Div. FF, Title III, Subtitle C (2022).
23. World Health Organization. Guidelines on evaluation of biosimilars. Replacement of Annex 2 of WHO Technical Report Series, No. 977. Geneva: World Health Organization; 2022.
24. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci*. 2019;40(8):577–91.
25. Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision medicine, AI, and the future of personalized health care. *Clin Transl Sci*. 2021;14(1):86–93.
26. Arden NS, Fisher AC, Tyner K, Yu LX, Lee SL, Kopcha M. Industry 4.0 for pharmaceutical manufacturing: preparing for the smart factories of the future. *Int J Pharm*. 2021; 602:120554.
27. Rački D, Tomazečič D, Skočaj D. Detection of surface defects on pharmaceutical solid oral dosage forms with convolutional neural networks. *Neural Comput Appl*. 2021. doi:10.1007/s00521-021-06397-6.
28. Amazon Web Services. How machine learning is changing the pharmaceutical industry and reshaping the way drugs are produced in the 21st century. AWS for Industries Blog [Internet]. [cited 2026 Oct 2]. Available from: <https://aws.amazon.com/blogs/industries/in-the-news-the-intelligent-era-design-detect-discover/>
29. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell*. 2020;2(10):573–84.
30. Cao Y, Romero J, Olson JP, Degroote M, Johnson PD, Kieferová M, et al. Quantum chemistry in the age of quantum computing. *Chem Rev*. 2019;119(19):10856–915.
31. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020; 3:119.
32. Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst*. 2021;32(1):4–24.