2025 Volume: 01 Issue: 06 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-13

ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY AND DEVELOPMENT

***Sakshi Kanwar, Dr .Vishal Shrivastava, Prof. Amit Tewari, Dr. Akhil Pandey**

Computer Science & Engineering, Arya College of Engineering & I.T. Jaipur, India.

Article Received: 09 November 2025

Article Revised: 29 November 2025

Published on: 19 December 2025

***Corresponding Author: Sakshi Kanwar**

Computer Science & Engineering, Arya College of Engineering & I.T. Jaipur, India. DIO: https://doi-doi.org/101555/ijrpa.1264

ABSTRACT

Recent advances in artificial intelligence (AI) have fundamentally restructured the field of drug discovery and development, yielding dramatically accelerated timelines, significantly enhanced predictive accuracy, and innovative computational methodologies. The pharmaceutical industry traditionally faces challenges characterized by protracted timelines, often exceeding 10 years, and prohibitively high attrition rates in clinical trials. This comprehensive review examines the integration of AI technologies, including machine learning (ML), deep learning (DL), and generative models, demonstrating how these computational approaches are redefining target identification, de novo drug design, high-throughput virtual screening, and optimization of clinical development. Robust evidence and quantitative case studies are presented, affirming that AI integration has reduced the average development duration to an estimated 3–6 years and increased Phase I trial success rates for AI-designed drugs to 80–90%, compared to the traditional 40–65% range. The paper further details the specific architectures (e.g., Recurrent Geometric Networks (RGN), Reinforced Adversarial Neural Computers (RANC)) and critical datasets (MISATO, ChemDiv) driving these advances, while rigorously analyzing prevailing challenges concerning data quality, model interpretability, and regulatory harmonization. The findings strongly support the strategic, continued investment in AI-driven pharmaceutical research to enable more efficient, effective, and accessible therapeutic development.

KEYWORDS: Deep Learning, Generative Models, Recurrent Geometric Networks (RGN), Cheminformatics, Virtual Screening, Clinical trials, IEEE.

I. INTRODUCTION

I.I The Unsustainable Paradigm of Traditional Pharmaceutical R&D

The pharmaceutical industry has long been constrained by a development model characterized by extreme cost and duration. Traditional drug development is protracted, typically spanning over 10 years from target identification to market approval, consuming vast resources. Furthermore, the financial investment required is substantial, with the estimated cost per successful new chemical entity averaging approximately $2.6 billion. The primary driver of this escalating expenditure is the persistently high rate of failure, or attrition, particularly during the crucial clinical trial stages. Historically, the end-to-end probability of a candidate molecule achieving market approval, starting from Phase I, has ranged between 5% and 10%. This high failure rate in later, more expensive stages creates a significant financial and operational bottleneck that conventional, predominantly wet-lab methodologies have struggled to overcome. Consequently, the global imperative for accelerated, cost-effective methods to bring innovative medicines to patients has never been greater.

I.II The Transformative Potential of Artificial Intelligence

Artificial intelligence technologies, encompassing machine learning (ML), deep learning (DL), and advanced generative models, present a transformative solution to the historical inefficiencies plaguing pharmaceutical R&D. By leveraging sophisticated algorithms, AI systems are capable of rapidly processing and interpreting vast chemical and biological datasets—including omics data, structural chemistry, and historical clinical outcomes. This capability allows researchers to move beyond brute-force experimental methods. AI facilitates the learning of complex, non-linear relationships between molecular structure and biological activity, enabling the optimization of candidate selection and the efficient identification of optimal compounds. This paradigm shift circumvents historical bottlenecks, enhances innovation capacity, and actively supports the movement toward precision and personalized medicine, where treatments are tailored to specific patient populations or genetic profiles.

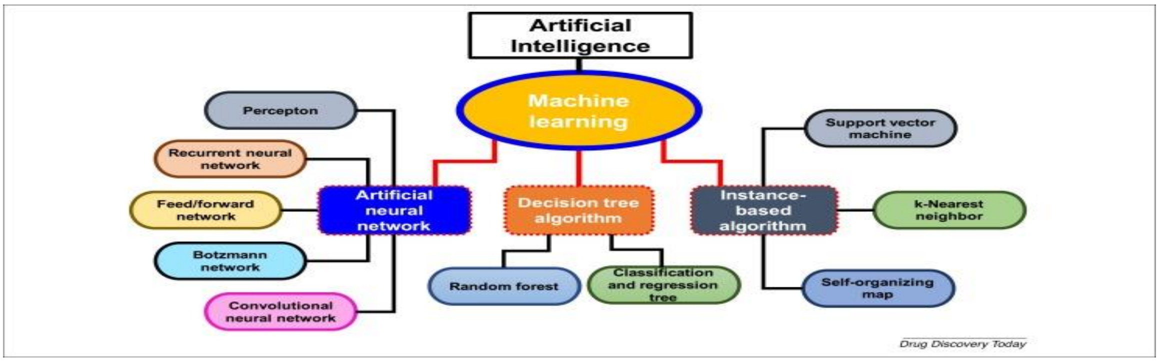


Figure 1: Artificial Intelligence and its components.

I.III Scope, Contribution, and Paper Organization

This paper provides a rigorous, expert-level review of the integration of AI within the drug discovery and development pipeline. It distinguishes itself by offering a detailed analysis of specific, cutting-edge AI architectures, such as Recurrent Geometric Networks (RGN) and Reinforced Adversarial Neural Computers (RANC), detailing their underlying mechanisms and contributions to both structure prediction and de novo design. Furthermore, the analysis provides a robust quantitative assessment of AI's demonstrable impact on critical metrics, including clinical success rates, development timelines, and the dynamics of market growth and investment. The subsequent sections are organized to first establish the foundational AI models and data requirements, proceed through the application of AI in early discovery (structure prediction) and generative chemistry, analyze its role in clinical optimization, and conclude with a quantitative impact assessment and a critical discussion of extant challenges related to generalizability, interpretability, and the rapidly evolving regulatory landscape.

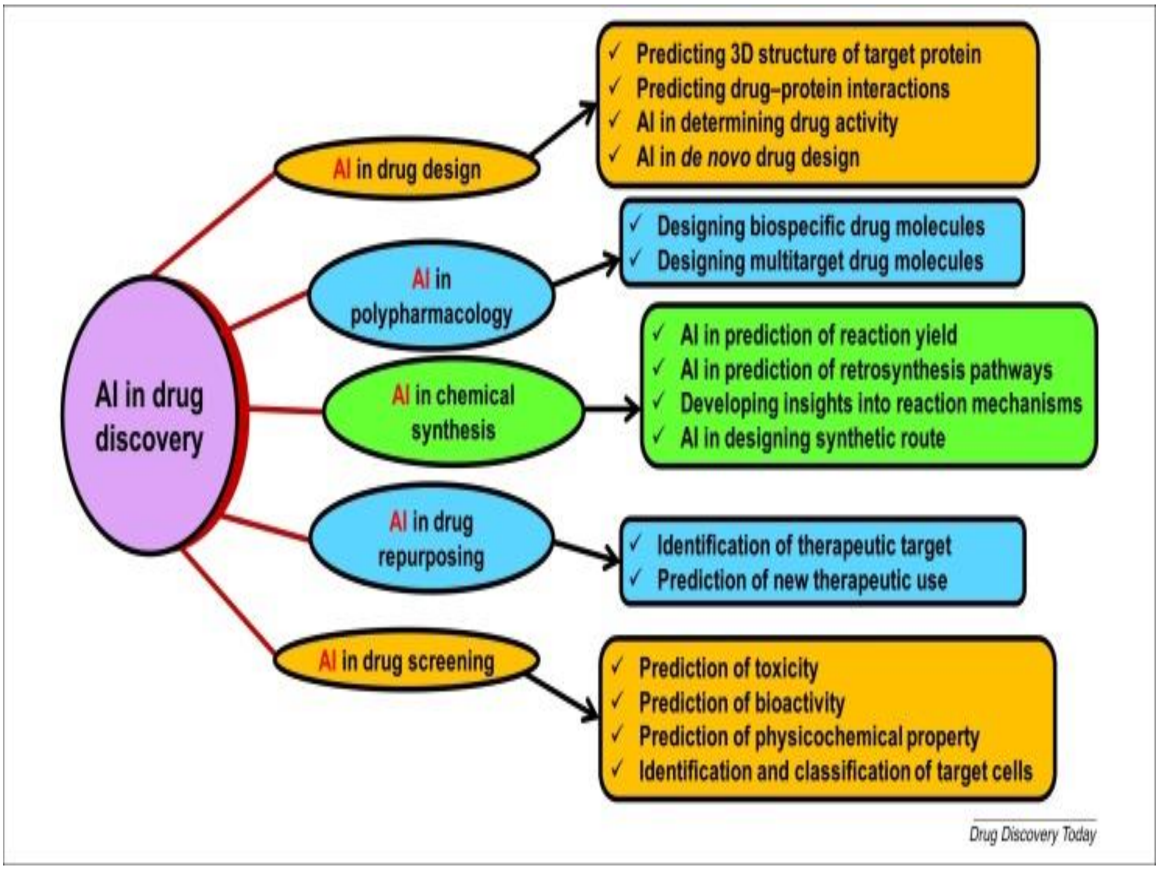

AI in drug discovery
AI in drug design
✓ Predicting 3D structure of target protein
✓ Predicting drug–protein interactions
✓ AI in determining drug activity
✓ AI in *de novo* drug design
AI in polypharmacology
✓ Designing biospecific drug molecules
✓ Designing multitarget drug molecules
AI in chemical synthesis
✓ AI in prediction of reaction yield
✓ AI in prediction of retrosynthesis pathways
✓ Developing insights into reaction mechanisms
✓ AI in designing synthetic route
AI in drug repurposing
✓ Identification of therapeutic target
✓ Prediction of new therapeutic use
AI in drug screening
✓ Prediction of toxicity
✓ Prediction of bioactivity
✓ Prediction of physicochemical property
✓ Identification and classification of target cells
Drug Discovery Today

Figure 2 : Role of Artificial Intelligence in Drug Discovery.

II. Foundational AI Architectures and Data Ecosystem

The power of AI in pharmacology derives from the specialized architectures designed to handle the complex, multi-dimensional data inherent in biological and chemical systems.

II.I Deep Learning Architectures for Molecular Modeling

Deep learning (DL) models are central to contemporary drug discovery efforts. Deep Neural Networks (DNNs) serve as a general type of model used for structured data prediction and high-throughput tasks, with highly successful examples such as AlphaFold demonstrating their ability to predict three-dimensional protein structures.

A. Specialized Architectures

Graph Neural Networks (GNNs): Molecules and biological networks are fundamentally graph structures, where atoms or proteins represent nodes and chemical or physical bonds represent edges. GNNs are uniquely suited to learning complex topological relationships within these structures, making them crucial for molecular property prediction and network pharmacology, which involves mapping therapeutic targets within complex biological systems.

Recurrent Neural Networks (RNNs): These networks are effective at processing sequential data. In cheminformatics, RNNs are often applied to sequential representations of compounds, such as SMILES strings, or to predict protein structure from amino acid sequences. The RGN model, a specific type of RNN, exemplifies this application in protein structure prediction.

Generative Models: The most disruptive architectures in de novo drug design are the generative models, including Generative Adversarial Networks (GANs) and Reinforced Adversarial Neural Computers (RANC). These models are designed not merely to predict properties but to output entirely novel molecular structures that adhere to desired pharmacological constraints.

II.II The Role of High-Fidelity Training Data

Effective AI models are entirely dependent on access to vast quantities of high-quality, high-fidelity biological and chemical data.

A. Public Repositories and Benchmarking

Public repositories remain the bedrock of chemical informatics. Databases such as ChEMBL, ZINC, and PubChem provide accessible chemical structures, activity data, and bioactivity information that are essential for the initial training of foundational models and for ensuring reproducibility across independent research groups. These large, publicly available datasets enable effective model training and benchmarking processes.

B. Proprietary and High-Value Datasets

As AI models become more sophisticated, particularly those focused on structure-based design, the requirement shifts from maximizing data quantity to enhancing data quality and informational richness.

MISATO Dataset: This dataset is critical for realistic molecular modeling. It contains protein-ligand interaction data, complex quantum chemistry calculations, and Molecular Dynamics (MD) simulation results. The inclusion of high-fidelity structural and dynamic information, such as that provided by MD simulations, allows ML models to learn subtle biological realities, supporting sophisticated tasks like accurate protein-ligand docking and quantum modeling.

ChemDiv Dataset: Providing curated experimental compound activity data, the ChemDiv dataset is validated specifically for ML applications. It offers proprietary chemical structures that serve as benchmark standards for evaluating the performance and generalizability of newly developed ML models.

The necessity for high-fidelity data, particularly the detailed quantum and dynamic information contained in datasets like MISATO, reflects a fundamental shift in the field. Future breakthroughs are increasingly dependent on data that captures the subtle physical and dynamic behavior of molecules, moving beyond simple static structure-activity relationships. This emphasis on rich data allows complex models to accurately simulate real-world biological systems.

III. AI in Early Discovery: Structure and Function Prediction

The earliest stages of drug discovery—identifying a viable target and determining its structure—have been dramatically accelerated by AI, fundamentally improving the efficiency of structure-based drug design.

III.I Target Identification via Multi-Omics and Network Pharmacology

Before designing a drug, a druggable protein or pathway must be identified. AI systems analyze vast, complex multi-omics datasets (including genomics, proteomics, and transcriptomics) to identify key therapeutic targets and novel vulnerabilities, such as oncogenic pathways. This process often utilizes GNNs to model and analyze the intricate relationships within biological networks, providing a system-level understanding of disease mechanisms that surpasses the capacity of manual analysis. AI can thus help define the optimal point of intervention in a complex biological cascade.

III.II Mechanistic Detail of Protein Structure Prediction Models

Predicting the precise three-dimensional structure of a target protein is essential for structure-based drug design, allowing researchers to rationally design a complementary small molecule ligand.

AlphaFold (DNN-based): This widely recognized deep neural network tool has achieved high-speed and exceptional accuracy in predicting complex protein structures. Its ability to generate precise 3D structures has significantly accelerated the druggability assessment process, reducing months of experimental work to hours of computation time.

Recurrent Geometric Networks (RGN): The RGN architecture is a specific implementation of Recurrent Neural Networks (RNNs) dedicated to predicting protein 2D/3D structures. RGNs operate by encoding the target amino acid sequence using an RNN and then parameterizing the local protein structure based on torsional angles. Critically, the model employs recurrent geometrical units to couple the local structure predictions to the final global 3D structure. This sequence-dependent structural modeling has demonstrated superior speed and accuracy compared to many traditional statistical prediction models.

III.III Incorporating Physical Realism into AI Prediction

Despite their predictive power, purely data-driven machine learning models face a persistent limitation: they sometimes generate outputs that are mathematically sound but physically impossible.

A. The Physics-Informed Limitation

Models trained on structural data, such as earlier iterations of protein folding tools, can occasionally suggest "unphysical" folds or configurations, especially when attempting to predict the structure of a protein sequence significantly divergent from the training data. These physically implausible structures, while computationally derived, are useless for drug design.

B. Physics-Informed Machine Learning (PIML)

To mitigate this issue and enhance reliability, the field is moving toward Physics-Informed Machine Learning (PIML). Models like NucleusDiff explicitly incorporate simple, fundamental physical concepts—such as the repulsive forces that prevent atoms from occupying the same space—directly into the algorithm's training loss function. By enforcing known physical laws and geometric constraints, PIML ensures that the resulting molecular designs and protein folds are physically sound and geometrically plausible. This strategy provides critical robustness, reducing the reliance on having perfectly representative training data and enhancing the models' generalizability when exploring novel chemical spaces. This methodological evolution is essential for moving AI from a sophisticated hypothesis generator to a reliable design tool.

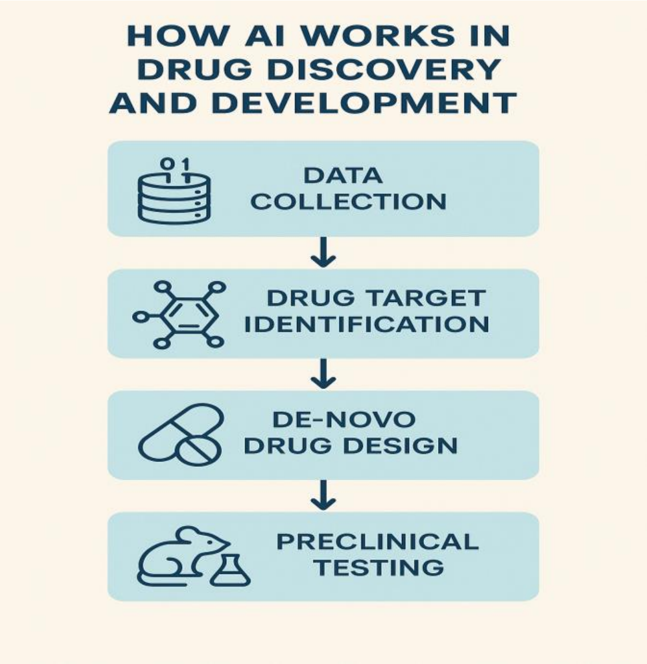

HOW AI WORKS IN DRUG DISCOVERY AND DEVELOPMENT
DATA COLLECTION
DRUG TARGET IDENTIFICATION
DE-NOVO DRUG DESIGN
PRECLINICAL TESTING

Figure 3: The process of drug discovery and development

IV. Generative Chemistry and Synthetic Feasibility

The core innovation of AI in drug discovery lies in generative chemistry—the ability to design entirely novel compounds rather than merely screen existing ones.

IV.I Reinforced Generative Models for De Novo Design

Deep generative models are central to creating new molecular entities optimized for specific pharmacological profiles.

- **RANC (Reinforced Adversarial Neural Computers):** RANC represents a class of sophisticated generative models used in de novo design.[1] These systems leverage chemical descriptors (such as molecular weight (MW), lipophilicity ($\text{log}P$), and topological polar surface area (TPSA)) to guide the generation process, ensuring the outputted structures are unique but adhere to desirable physicochemical property distributions.[9] By actively generating novel structures based on these constraints, RANC facilitates the rapid and broad exploration of the chemical space, often outperforming older generative platforms.[1] The focus is on maximizing chemical diversity and novelty while maintaining "drug-like" characteristics.
- **Generative AI Market Context:** The industrial focus on these creative AI tools is quantitatively reflected in market dynamics. While the overall AI in drug discovery market maintains steady growth, the generative AI subset is experiencing explosive expansion,

expanding at a Compound Annual Growth Rate (CAGR) of 27.38% between 2024 and 2034, projecting a market size of USD 2.30 billion by 2034.[10]

IV.II Addressing Synthetic Feasibility via Retrosynthesis

A major challenge for purely generative models like RANC is that the novel molecules they design might not be synthetically viable or cost-effective to produce in a laboratory. This is where AI-driven retrosynthesis planning becomes crucial.

- **Synthia (formerly Chematica):** This platform addresses synthetic feasibility. Unlike generative deep learning models, Synthia operates as a rule-based expert system developed over two decades by expert chemists.[11] It leverages a massive catalog of commercially available starting materials and an extensive, hand-encoded database of synthetic rules to propose highly efficient and cost-effective synthesis routes.[1]
- **Performance and Integration:** Synthia ensures that de novo designed molecules can actually be manufactured, often surpassing conventional planning in yield and cost-effectiveness.[1] Key performance metrics include the number of solved molecules, computational clock time, and throughput (e.g., 50 molecules per hour via API).[12] A highly successful AI pipeline requires the integration of novel design (RANC) with synthetic validation (Synthia) to ensure that discovery leads to practical development.

Table I illustrates the functional divergence and integration potential among key specialized AI models used across the early discovery pipeline.

Table I: Comparative Analysis of Specialized AI Models in the Discovery Pipeline.

Model/Platform	Core Technology	Primary Application	Key Output/Function	Mechanism Insight
AlphaFold	Deep Neural Network (DNN)	Target Structure Prediction	High-accuracy 3D protein structures [1]	Focuses on structural fidelity via large biological sequence data.
RGN	Recurrent Geometric Network (RNN)	Protein Structure Prediction	Precise 3D structure based on sequence torsion angles [1]	Encodes sequence dependence through torsional angles for structural modeling.
Synthia (Chematica)	Rule-Based Expert System	Retrosynthesis Planning	Efficient, cost-effective alternate synthesis routes [1]	Expert-coded rules ensure chemical feasibility and synthetic practicality.
RANC	Reinforced Adversarial NC (Generative Model)	De Novo Molecule Design	Novel molecular structures optimized by chemical descriptors [1]	Explores vast chemical space by matching distributions of desirable physicochemical properties.

IV.III Virtual Screening and ADMET Optimization

Following de novo design, lead optimization is performed via computational high-throughput screening (vHTS). Machine learning models rapidly assess millions of potential compounds for bioactivity, efficacy, and crucial Absorption, Distribution, Metabolism, Excretion, and

Toxicity (ADMET) properties. These models leverage large datasets and advanced cheminformatic descriptors, such as the Tanimoto coefficient, to forecast the success of lead compounds. By pre-filtering candidates for favorable ADMET profiles computationally, these systems drastically reduce the time and cost associated with experimental screening and late-stage failures.

V. AI in Clinical Development and Translational Science

The impact of AI extends beyond the laboratory bench and into the high-cost, high-stakes environment of clinical trials, optimizing operations and mitigating risk.

V.I Accelerating Clinical Trial Operations

AI has profoundly streamlined the operational aspects of clinical development. The critical process of patient enrollment is accelerated by AI analyzing Electronic Health Records (EHRs) and other large clinical datasets to match patients to trials based on specific criteria and predict potential success.[1] Similarly, AI-driven optimization of trial site selection—historically a complex, time-consuming process—can accelerate the total trial timeline by over 12 months.[13] Furthermore, the introduction of Generative AI (Gen AI) to streamline administrative processes, such as the auto-drafting of trial documents, has been shown to cut related process costs by up to 50%.[13]

V.II Predictive Modeling and Adaptive Trial Design

AI provides powerful predictive modeling capabilities that simulate physiological responses and forecast trial outcomes, allowing for the adoption of more agile, adaptive trial designs.[1] A significant development is the use of synthetic data to create **synthetic control arms** or **digital twins**.[4] These innovations use real-world data or simulated virtual patient data to model outcomes, potentially reducing the logistical and ethical complexities associated with traditional placebo arms.[4] This acceleration of data generation facilitates faster decision-making and improved regulatory interactions.[1] However, relying heavily on synthetic data introduces risks of bias or overfitting if the synthetic population is not perfectly representative of the target patient demographic.[4] Ensuring the validity and generalizability of these digital representations requires stringent data curation and regulatory oversight.

VI. Quantitative Impact Assessment and Financial Dynamics

The integration of AI has created measurable disruption in the metrics of pharmaceutical R&D, validated by substantial improvements in success rates and significant market growth.

VI.I Paradigm Shift in Development Metrics

The quantitative evidence demonstrates that AI is a true accelerator of drug development, particularly in the earliest, most bottlenecked stages.

A. Timeline and Phase I Success Rate Disruption

The overall reduction in development duration, attributable to AI, has moved the average time from over 10 years to an estimated 3–6 years.[1] Crucially, AI has acted as a powerful filter during lead optimization. AI-designed drugs entering clinical trials exhibit a remarkable 80–90% success rate in Phase I—a stark contrast to the 40–65% success rate typical of conventionally discovered compounds.[1] This near-doubling of success confirms AI's superior ability to identify molecules with optimized ADMET properties, pharmacokinetics, and reduced initial toxicity profiles before they incur major clinical costs.[15]

B. The Phase II Plateau

While the Phase I success rate is highly differentiated, the advantage appears to diminish rapidly in later stages. Studies show that AI-discovered drugs achieve a Phase II success rate of approximately 40%.[2] This figure is comparable to the historical industry average for Phase

II trials.[15] This data suggests that AI has successfully mastered the initial in vitro and early safety filtering process (the "Phase I filter") but still faces the persistent, fundamental challenge of predicting complex in vivo efficacy and safety within diverse, heterogeneous human populations.[15] Despite this plateau, the overall productivity increase is substantial: the end-to-end probability of a molecule gaining approval increases from the traditional 5–10% baseline to an AI-enhanced 9–18%.[2]
Table II summarizes the quantitative changes driven by AI integration.

Table II: Quantitative Impact of AI on Drug Development Metrics.

Metric	**Traditional Industry Average**	**AI-Enhanced Outcome**	**Improvement/Data Source**
Total Development Timeline	>10 years [1]	$3-6$ years (Estimated 2025) [1]	Up to 70% reduction
Phase I Success Rate	$40\%-65\%$ [2]	$80\%-90\%$ (AI-Designed Drugs) [1]	Near doubling of success rate; improved ADMET filtering.
Phase II Success Rate	$\sim40\%$ (Historical) [15]	$\sim40\%$ (AI-Designed Drugs) [2]	Comparable to industry average; challenge remains in translational efficacy.
End-to-End Approval Probability	$5\%-10\%$ [2]	$9\%-18\%$ [2]	Doubling of R&D productivity overall.
Process Cost Reduction	N/A (High) [3]	Up to 50% cut in operational documentation costs [13]	Enhanced efficiency in trial operations.

VI.II Market Growth and Investment Landscape

The financial community's recognition of AI's disruptive potential is demonstrated by accelerating investment trends. The global AI in drug discovery market, valued at USD 6.93 billion in 2025, is projected to reach USD 16.52 billion by 2034, growing at a sustained CAGR of 10.10%.[16]

The Generative AI segment is experiencing particularly explosive growth due to its direct link to novelty generation (i.e., new intellectual property). This specialized market segment, driven by technologies like RANC, is growing at a rapid CAGR of 27.38% over the forecast period, expected to hit approximately USD 2.30 billion by 2034.[10] This focus on generation is mirrored in venture capital activity. Global VC funding for AI health tech in 2025 has already exceeded the full-year totals for 2024 by 24.4%, with startups capturing 85% of the total generative AI spend.[17] This high level of investment, particularly directed toward generative technologies, suggests a strategic focus by the investment community on fostering platforms capable of creating novel IP, prioritizing discovery innovation alongside operational efficiency.
Table III provides a detailed breakdown of the financial forecasts underpinning this transformation.

Table III: Global Market and Investment Forecasts for AI in Drug Discovery. (2025–2034)

Category	Value (Base Year: 2025)	Projected Value (End Year: 2034)	CAGR (%)	Leading Trend
Overall AI in Drug Discovery Market	USD 6.93 Billion [16]	USD 16.52 Billion [16]	10.10% [16]	Sustained growth driven by efficiency needs.
Generative AI in Drug Discovery Market	USD 260.56 Million [10]	USD 2.30 Billion [10]	27.38% [10]	Explosive growth driven by novelty generation and VC funding.
Venture Capital Funding (AI Health Tech)	USD 10.7 Billion (2025 YTD) [17]	N/A	24.4% increase (2025 YTD over 2024) [17]	Indicates high investor confidence and focus on disruptive startups.

VII. Ethical and Regulatory Challenges

Despite the overwhelming technological and quantitative success, several fundamental challenges must be addressed to ensure the safe, ethical, and generalized deployment of AI in pharmacology.

VII.I Data Integrity and Bias

The quality of input data remains a critical vulnerability. Incomplete, inaccurate, or unbalanced datasets can lead to skewed predictions and poor model generalization. When models rely on non-diverse training data, they often exhibit "shortcut learning," resulting in failure when applied to chemical spaces or patient populations not represented in the original data. The necessity for high-fidelity data, such as the quantum and MD simulations provided by the MISATO dataset , and the drive toward physics-informed models (PIML), are direct scientific responses aimed at mitigating these inherent data biases and enhancing the robustness required for real-world application.

VII.II The Interpretability Barrier (The Black Box Problem)

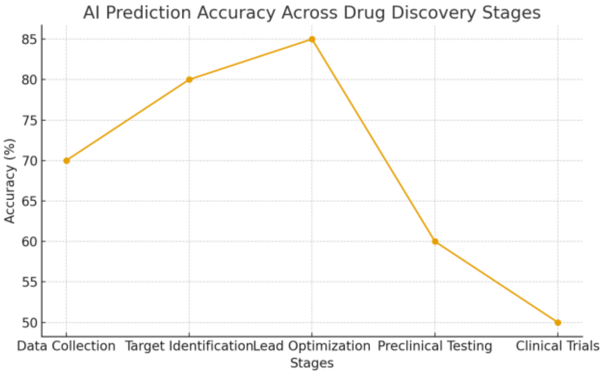
A significant non-technical impediment is the "black box" nature of many complex deep learning models. Regulatory bodies and clinicians require transparent, scientifically justifiable explanations for why a model selects a drug candidate, predicts a specific toxicity profile, or advises a particular treatment pathway. Lack of interpretability, often referred to as the requirement for Explainable AI (XAI), impedes regulatory approval and compromises clinical trust, creating a bottleneck for the broad adoption of these powerful tools.

VII.III Evolving Regulatory and Ethical Frameworks

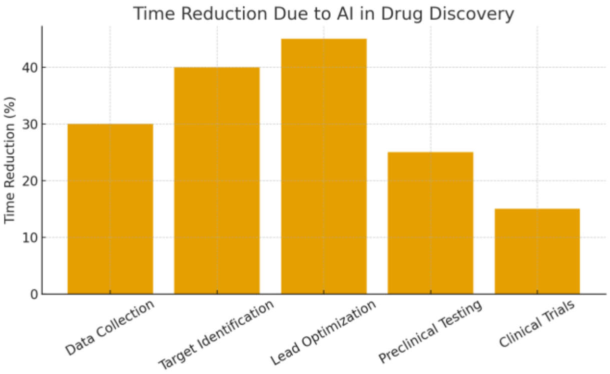
Given the speed of AI innovation, regulatory frameworks must rapidly evolve to govern the use of AI-derived medicines and diagnostics. Agencies such as the FDA and the European Medicines Agency (EMA) are actively crafting comprehensive guidelines to manage the ethical and privacy challenges associated with sensitive patient data and model deployment. Initiatives such as the UK's Medicines and Healthcare products Regulatory Agency (MHRA) AI Airlock, a regulatory sandbox, are crucial for identifying and addressing regulatory issues specific to AI as a Medical Device (AIaMD) through simulation and real-world testing, ensuring that novel technologies are integrated safely and ethically into clinical practice.

VII.IV Future Outlook

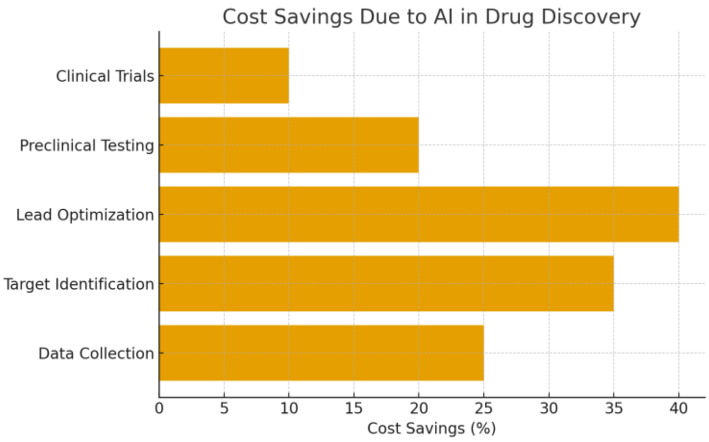
The trajectory of AI in drug discovery points toward ever-greater autonomy and integration. Innovations anticipated in the coming years include the realization of fully autonomous AI-driven discovery pipelines that minimize human intervention across multiple stages. Blockchain technology is expected to be integrated to provide secure, traceable, and immutable biomedical data sharing platforms. Ultimately, the field is moving toward expanded human-AI collaborative platforms aimed at seamlessly integrating predictive models into personalized medicine strategies, allowing treatments to be dynamically adapted to individual patient profiles.



Time reduction due to AI



Cost savings due to AI



VIII. CONCLUSION

This comprehensive review establishes that artificial intelligence has instigated a fundamental, quantitative transformation across the entire drug discovery and development spectrum. The strategic integration of AI technologies has successfully addressed the industry's most critical failure points, resulting in a dramatic reduction of average development timelines from over 10 years to an estimated 3–6 years.[1] Furthermore, AI-designed drug candidates exhibit dramatically enhanced quality at the lead optimization stage, evidenced by Phase I clinical trial success rates soaring to 80–90%, effectively doubling early-stage R&D productivity.[1]

Specific technological advances, such as the use of RGNs for precise structural prediction and generative models like RANC for novelty generation, complemented by systems like Synthia for synthetic validation, are driving this change.[1] Key datasets, including MISATO and ChemDiv, continue to be foundational in supporting the development of high-fidelity, generalizable models.[1]

To unlock the full potential promised by the rapidly accelerating market (especially the generative AI segment's 27.38% CAGR [10]), concerted research efforts must focus on solving the persistent challenges of data generalizability and model interpretability. Continued collaboration between domain experts, regulatory bodies, and computer scientists is necessary to harmonize regulatory frameworks and institutionalize the ethical safeguards required for these powerful tools, ultimately accelerating the delivery of safer, more effective treatments to patients worldwide.

REFERENCES

1. AI-Driven Drug Discovery: A Comprehensive Review,” ACS Omega, vol. 10, no. 23, pp. 23889–23903, Jun. 2025. doi: 10.1021/acsomega.5c00549. [Online]. Available: https://pubs.acs.org/doi/10.1021/acsomega.5c00549.
2. Ocana, et al., “Integrating artificial intelligence in drug discovery and early development: opportunities and challenges,” Pharmaceutics, Mar. 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11909971 .
3. Ocana, et al., “Integrating artificial intelligence in drug discovery and early development: opportunities and challenges,” Pharmaceutics, Mar. 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11909971 .
4. Helmholtz Zentrum München, “MISATO Dataset: Transforming Drug Discovery with AI,” Dec. 2023. Accessed: Nov. 10, 2025. [Online]. Available: https://www.helmholtz-munich.de/en/newsroom/news-all/artikel/ misato-dataset-transforming-drug-discovery-with-ai
5. 5ChemDiv, “Validated Dataset for AI-Driven Drug Discovery.” Accessed: Nov. 10, 2025. [Online]. Available: https://www.chemdiv.com/datasets/ .
6. “Early evidence and emerging trends: How AI is shaping drug discovery and clinical development,” Drug Target Review, Apr. 2025. [Online]. Available: https://www.drugtargetreview.com/article/158593/early-evidence-and-emerging-trends-how-ai-is-shaping-drug-discovery-and-clinical-development/ .

7. Y. Wang, L. Feng, Q. Wang, Y. Xu, and D. Guo, “PRRGNVis: Multi-Level Visual Analysis of Comparison for Predicted Results of Recurrent Geometric Network,” Appl. Sci., vol. 12, no. 17, p. 8465, 2022. doi: 10.3390/app12178465.
8. “Unleashing the power of generative AI in drug discovery,” Drug Discov Today, vol. 29, no. 6, p. 103992, Jun. 2024. doi: 10.1016/j.drudis.2024.103992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S135964462400117X.
9. “Artificial intelligence in drug discovery and development,” PMC, 2020.
10. “The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies,” Pharmaceuticals (Basel), vol. 16, no. 6, p. 891, Jun. 2023. doi: 10.3390/ph16060891. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10302890/ .
11. Lifebit, “AI Driven Drug Discovery: 5 Powerful Breakthroughs in 2025,” 2025.
12. Grand View Research, “Artificial Intelligence In Drug Discovery Market To Reach $9.1Bn By 2030,” Mar. 2023.
13. Recursion, “Accelerating AI Drug Discovery with Open Source Datasets,” 2025. Accessed: Nov. 10, 2025. [Online]. Available:https://www.recursion.com/news/accelerating-ai-drug-discovery-with-open-source-datasets .
14. Pelago Bioscience, “The Top 5 Drug Discovery Trends Defining 2025: What They Mean for the Future of Innovation,” 2025. Accessed: Nov. 10, 2025. [Online]. Available: https://www.pelagobio.com/cetsa-drug-discovery-resources/blog/drug-discovery-trends-2025 / .
15. Z. Wan, “Applications of Artificial Intelligence in Drug Repurposing,” Adv Sci (Weinh), 2025. doi: 10.1002/advs.202411325.
16. E. R. Gold and R. Cook-Deegan, “AI drug development's data problem,” Science, vol. 388, no. 6743, p. 131, Apr. 2025. doi: 10.1126/science.adx0339. [Online]. Available: https://doi.org/10.1126/science.adx0339 .
17. R. Rodriguez-Chavez, “Artificial Intelligence and Regulatory Realities in Drug Development: A Pragmatic View,” DIA Global Forum, Apr. 2025. Accessed: Nov. 10, 2025. [Online]. Available: https://globalforum.diaglobal.org/issue/april-2025/artificial-intelligence-and-regulatory-realities-in-drug-development-a-pragmatic-view / .