Proceedings of the 5th International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/2027.38309

Generative AI Drives a Paradigm Shift in Drug Discovery: An Analysis of Technical Approaches, Application Cases, and Challenges

Hongyu Mu

School of Pharmaceutical Sciences, South-Central Minzu University, Wuhan, China
cooperalbert99@163.com

Abstract. Traditional drug research and development has long been characterised by long development cycles, high investment costs and a high attrition rate among candidate compounds, conventional experimental methods for screening candidate compounds are of limited efficiency when faced with the vast chemical space and complex biological mechanisms. Generative artificial intelligence can provide a new technological approach to new drug research and development by generating conditions based on the distribution patterns of molecular structures and their properties, and by enabling de novo molecular design. This paper systematically reviews the main generative models—including RNNs, Transformers, VAEs, GANs, normalising flows and diffusion models—and their technical characteristics. It analyses their applications in drug target discovery, protein structure prediction, candidate compound generation and lead optimisation, with a particular focus on how generative AI, when combined with active learning, computational physics and feedback from wet-lab experiments, reshapes the drug discovery and development process. Research indicates that generative AI, when combined with active learning, physical computing and feedback from wet-lab experiments, is driving a shift in drug discovery from the traditional linear open-loop model towards a data-driven iterative model of 'generation–evaluation–feedback–regeneration–experimental validation'. By examining the technical characteristics of various generative models and their suitability for drug discovery tasks, this paper provides guidance on selecting generative AI technologies, identifying suitable application scenarios and integrating them into drug discovery workflows.

Keywords: generative artificial intelligence, drug research and development, de novo molecular design, lead optimization, artificial intelligence-aided drug discovery

1. Introduction

Traditional drug research and development (R&D) is characterised by long cycles and high investment costs, whilst the high attrition rate of candidate compounds under stringent regulatory standards keeps R&D risks persistently high, one of the key reasons for this is that manual experimental screening of candidate compounds struggles to efficiently and accurately identify active drug molecules within the vast chemical space and complex biological mechanisms [1-3].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2027.38309

Traditional machine learning models in artificial intelligence (AI) are capable of predicting molecular properties based on existing molecular data, thereby improving the efficiency of chemical molecule discovery [2]. The emergence of generative artificial intelligence (generative AI) has enabled the further study of patterns in chemical structures to generate new candidate compounds, this advancement has transformed the way the chemical space is explored, paving the way for a paradigm shift in drug R&D [1, 3, 4].

The different underlying generative mechanisms of various generative AI models mean that their applications in drug R&D also differ in focus. Among these, Generative Adversarial Networks (GANs), Variational Auto-Encoders (VAEs) and Normalising Flows (NFs) are capable of learning the distribution of existing molecules and are suitable for candidate compounds generation and structural exploration, Recurrent Neural Networks (RNNs) and Transformers, drawing on their sequence modelling capabilities, can be used for molecular representation learning and the analysis of complex bioinformatics relationships, diffusion models enhance the quality of structural generation through an iterative process and have demonstrated significant potential in the design of novel drug molecules [3, 5].

In recent years, generative AI has begun to move from theoretical research into practical drug R&D processes. For example, Insilico Medicine utilises AI technology in areas such as disease target discovery, design of candidate compounds and drug optimisation, demonstrating the potential for generative AI to be applied throughout the early stages of drug R&D and driving AI-assisted drug R&D into the clinical exploration phase [2]. Existing research is often conducted from a single perspective and lacks an analysis of the links between technology, applications and actual drug R&D processes. This paper focuses on how generative AI is transforming the drug R&D process, summarises the current challenges and outlines future directions.

2. Technical approaches and application scenarios for generative AI-driven drug R&D

2.1. Mainstream technical approaches and underlying principles of generative AI

2.1.1. The fundamental principles of generative AI-driven drug R&D

Traditional high-throughput screening (HTS) and some virtual screening (VS) methods primarily search for candidate compounds within existing compound libraries, factors such as chemical scaffolds and patent barriers limit the discovery of new drugs [1-3, 6]. Generative AI learns from vast amounts of molecular data to extract features such as chemical grammar, topological rules, and the probability distributions of physicochemical properties and bioactivity. Once training is complete, the model can generate new drug-like molecules through probabilistic sampling, whilst guiding the generation of target molecules based on constraints such as target activity, solubility, synthetic difficulty and toxicity [6-8]. Unlike traditional VS, which primarily involves screening candidate compounds from an existing compound library, generative AI is gradually transforming molecular exploration from 'screening within a library' to 'on-demand generation'.

2.1.2. Representation of drug molecular data and model adaptation

To train generative AI to generate target molecules, the molecular structures must first be converted into computer-readable data. At present, the three main representations are SMILES, molecular graphs and three-dimensional structures. SMILES encodes molecules as character sequences and is suitable for training models such as RNNs, Transformers and sequence-based VAEs [6], molecular

DOI: 10.54254/2753-8818/2027.38309

graphs use atoms as nodes and chemical bonds as edges to preserve molecular topological relationships, and are suitable for graph neural networks as well as graph VAEs, graph GANs and graph diffusion models, three-dimensional structures further incorporate atomic spatial coordinates and geometric information, making them suitable for models requiring spatial conformation modelling, such as 3D diffusion models and equivariant graph neural networks [5, 9]. Different representations preserve varying amounts of structural information, which in turn influences the choice of generative model and the tasks for which it is suitable.

2.1.3. Mainstream generative models and their technical characteristics

Once a digital representation of the molecular structure has been established, different generative models utilise distinct learning mechanisms to model molecular features and are applied to various drug R&D tasks (Table 1). RNNs and Transformers primarily generate candidate compounds by learning sequence representations such as SMILES. RNNs use an autoregressive strategy to generate molecular sequences token by token, while Transformers leverage the self-attention mechanism to strengthen long-range dependency modelling and benefit from pre-training and transfer learning [3, 5, 6, 10, 11], VAEs map molecules onto a continuous latent space and can be used for molecule generation, scaffold hopping and property optimisation [6, 11, 12], GANs learn the distribution of real molecules through adversarial training between generators and discriminators [3, 5, 6]. For example, VeGA utilises a lightweight Transformer for model pre-training and fine-tuning with target activity data, demonstrating the value of the Transformer's transfer learning capabilities in the generation of molecules for low-data targets [10].

The NF model performs explicit probability modelling of molecular distributions via inverse transformation, whilst the diffusion model learns molecular distributions through iterative noise addition and removal, and is capable of generating three-dimensional conditions by incorporating contextual information such as atomic three-dimensional coordinates and protein pockets [3, 5, 13]. For example, PharmacoForge, proposed by Flynn et al., utilises an isotropic diffusion model to generate three-dimensional pharmacophores based on protein binding pockets and uses these to screen real-world compound libraries, demonstrating the potential of diffusion models in the generation of protein pocket conditions and three-dimensional drug design [13]. Overall, different generative models have their own specific focuses in terms of data representation, generative mechanisms and suitable tasks. As drug design tasks gradually expand from two-dimensional molecule generation to multi-objective optimisation, target-conditioned generation and three-dimensional conformation design, generative models have also begun to evolve towards greater controllability and three-dimensional structural modelling.

Table 1. Summary of the five categories of deep generative models

Model category	Representative models	Core mechanism	Main applications	Key advantages	Key limitations
Sequence-based generation	RNNs, Transformers	RNNs: autoregressive sequence generation, Transformers: sequence modelling based on self-attention mechanisms	De novo molecular design, condition-driven molecular synthesis, and target-oriented molecular design	RNN architectures are relatively simple, Transformers are particularly effective at long-range dependency modelling and utilising pre-training transfer.	RNNs have limited capabilities for long-range dependency modelling, standard sequence models struggle to directly represent 3D conformational information, Transformers incur significant computational overhead.

DOI: 10.54254/2753-8818/2027.38309

Table 1. (continued)

Latent space generation	VAEs, AAEs	Encoding molecules into a continuous latent space, and generating molecules through sampling and decoding	Molecular optimisation, scaffold hopping, conditional generation	Continuous latent space facilitates molecular interpolation, structure exploration and property optimisation	There is a trade-off between generation quality and potential spatial constraints, under certain representations, invalid structures may be generated.
Adversarial generation	GANs	The generator and discriminator learn the distribution of molecular data through adversarial training	Molecule synthesis, multi-objective property optimisation	It is capable of learning the distribution of complex molecules and can carry out goal-oriented generation in conjunction with conditions or incentive mechanisms.	Training stability is poor, and model collapse is likely to occur, the use of discrete molecules makes gradient propagation difficult.
Normalizing flows	NFs	Mapping data distributions to simple distributions through a series of reversible transformations	Molecular generation, probability density and likelihood estimation	The likelihood can be explicitly calculated and is invertible	Reversible structures are subject to strong constraints, the model design is complex and involves significant computational overhead.
Probabilistic diffusion generation	Diffusion models	Learning the distribution of molecular data through gradual addition and removal of noise	3D molecular generation, protein pocket condition generation, conformation generation	Strong capabilities in 3D modelling and condition generation	Training and sampling are computationally expensive, the generated structures still require further evaluation in terms of their synthetic feasibility and physical plausibility.

2.2. Applications of AI-assisted drug target discovery and structure prediction

The design of a drug first requires research into the mechanisms of the disease, the identification of abnormal protein targets, and then the design of the drug based on those targets. Therefore, the identification of biological targets that can be modulated by drugs and that influence the course of a disease is central to targeted therapy and drug R&D. Traditional target discovery methods primarily include affinity pull-down and genome-wide knockout screening, these are time-consuming, labour-intensive and carry a high risk of failure. Furthermore, once a target has been identified, it must be validated through a combination of genetic, pharmacological and functional experimental evidence. The value of AI in target identification lies in its ability to integrate large-scale, multi-source biological data, screen candidate targets in bulk, and then prioritise them—thereby reducing the risk of identifying invalid targets at a later stage [1, 2, 4, 7].

At present, AI-assisted target discovery primarily involves the integration of multi-omics, literature and clinical data to perform feature learning, association analysis and the prediction of candidate targets. Common strategies include statistical analysis combining AI with genome-wide association studies (GWAS), network analysis such as knowledge graphs and gene co-expression networks, as well as machine learning classification or regression models. In addition, integrated AI platforms have been developed, for example, PandaOmics can integrate multi-omics data, literature and clinical information to screen and prioritise candidate targets [1, 2, 4 ,8].

Building on traditional prediction and ranking methods, generative AI has further expanded knowledge extraction and simulation in target discovery. Large language models (LLMs) can extract information such as disease mechanisms, targets and disease relationships from large-scale

DOI: 10.54254/2753-8818/2027.38309

biomedical literature, thereby assisting in the formulation of research hypotheses, some generative models can also simulate molecular and phenotypic changes following gene knockout or drug intervention, providing guidance for target screening and subsequent experimental validation [8].

Once a target has been identified,its three-dimensional (3D) protein structure must be determined to guide subsequent drug design. Structures are typically determined experimentally using methods such as X-ray crystallography, nuclear magnetic resonance and cryo-electron microscopy, however, these methods are costly and time-consuming, and the absence of experimentally determined structures can also hinder the implementation of structure-based drug design (SBDD) [1, 7]. Currently, the use of AI in protein structure and protein-ligand binding modelling primarily involves two approaches: one involves predicting the protein 3D structure based on its amino acid sequence, which is then used for molecular docking or VS, models such as AlphaFold have enabled computational studies on targets for which experimental structures are lacking [1, 2, 8], another approach utilises both protein and ligand information for joint modelling, directly predicting the structure of the protein–ligand complex [2, 8]. However, this type of joint modelling currently suffers from issues such as insufficient physical constraints, limited flexibility in receptor pockets, and a lack of precision in the description of interactions [2].

In summary, traditional AI methods are primarily involved in the prediction and prioritisation of candidate targets and in relationship analysis, whilst generative AI extends further into knowledge extraction, VS and the modelling of certain protein-ligand structures, thereby laying the groundwork for subsequent molecular design and optimisation.

2.3. Generative AI for drug optimisation and the redesign of R&D processes

2.3.1. Problems with traditional lead optimisation

Once a target has been identified, the next step is to screen for candidate compounds and optimise them. In traditional small-molecule drug discovery and lead optimisation, initial active molecules are typically identified through HTS, followed by the hit-to-lead and lead optimisation stages. This approach presents four main issues [2]: Firstly, traditional lead optimisation primarily involves local structural modifications to existing active molecules, as the number of molecules actually synthesised and tested in each round is limited, the scope of exploration within the chemical space is consequently restricted, Secondly, drug candidates' activity, selectivity, solubility, toxicity and pharmacokinetic properties may constrain one another, requiring researchers to continually refine the molecular structure in order to strike a balance [9], Thirdly, the iterative 'design–synthesis–testing–analysis' experimental cycle requires the repeated synthesis and testing of a large number of molecules, thereby increasing both time and experimental costs, fourthly, there is a scarcity of data on known active molecules and structure-activity relationships for new targets, and it is often necessary to establish the relationship between structure and activity gradually through multiple rounds of experimentation. These issues have also driven the gradual adoption of generative AI in drug design, with de novo molecular design and computational evaluation used to explore and optimise molecules [1].

2.3.2. Drive the optimisation of drugs and the restructuring of R&D processes

Traditional molecular optimisation relies primarily on researchers using experimental results to guide the next round of molecular design, feedback loops are often lacking between molecular generation, computational evaluation and experimental validation. Consequently, in recent years,

DOI: 10.54254/2753-8818/2027.38309

researchers have begun to explore combining generative AI with active learning, computational evaluation and experimental validation, thereby creating a continuous, iterative optimisation process that integrates experimentation and computation. This shift is illustrated below by a VAE combined with a two-layer nested active learning approach.

The VAE-AL closed-loop process (Figure 1) consists primarily of four major modules: molecule generation, inner-layer cheminformatics evaluation, outer-layer physics-based evaluation, and feedback-based updating. The VAE generates candidate compounds using target-related molecular data, the inner-layer chemical predictor then screens these candidate compounds based on metrics such as the Quantitative Estimate of Drug-likeness, synthetic accessibility and structural similarity, whilst the outer-layer physical predictor evaluates the predicted binding affinity between the compounds and the target using methods such as molecular docking. The molecules identified through the two-stage screening process are added to the target-specific dataset and used to update the generation module, thereby guiding the generation of the next round of candidate compounds [14].

In the CDK2 case study, the model screened tens of thousands of synthesised molecules over multiple rounds and identified 324 candidate compounds, 10 of these were ultimately selected for synthesis and validation. Of these, nine molecules underwent in vitro activity assays, with eight exhibiting an IC50 below 50 μM, the most potent molecule had an IC50 of 70 nM [14]. These results indicate that multi-tiered computational evaluation can bring the preliminary screening of a large number of candidate compounds forward to the computational stage, thereby narrowing down the range of candidates that proceed to synthesis and wet-lab experiments. However, the wet-lab experiments in this study were primarily used to validate final candidate compounds, it cannot therefore be concluded that the entire wet-lab process has been fully automated.

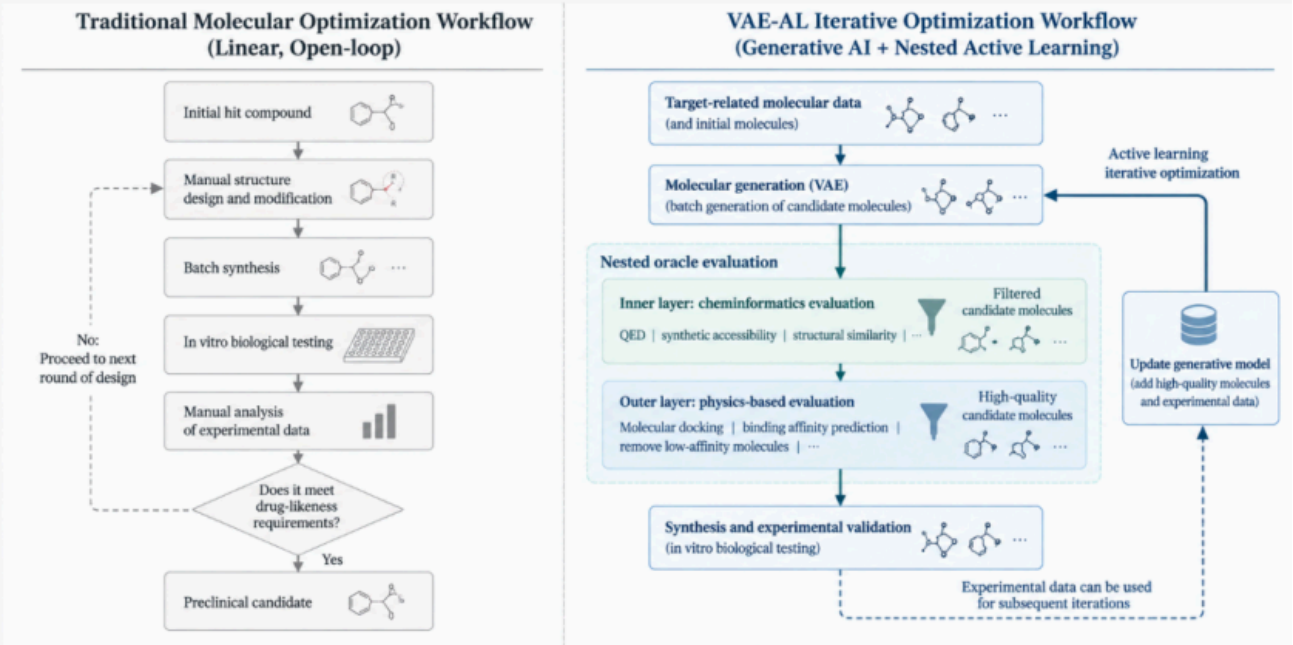


Figure 1. Comparison of traditional molecular optimisation with the molecular optimisation workflow incorporating generative AI

In a study on idiopathic pulmonary fibrosis, an AI platform proposed the TNIK target hypothesis by integrating multi-omics and other data, subsequently, a generative chemistry platform carried out de novo molecular design and optimisation for TNIK, ultimately leading to the development of the

DOI: 10.54254/2753-8818/2027.38309

TNIK inhibitor Rentosertib (INS018-055). According to relevant literature, this project took approximately 18 months from target identification to the nomination of preclinical candidate compounds, demonstrating the potential of AI platforms to shorten the early R&D cycle by bridging target discovery and candidate molecule design [8, 15].

The above examples demonstrate that the impact of generative AI on drug R&D is not limited to the generation of candidate compounds, but also extends to the integration of computational evaluation, experimental validation and data feedback. Many candidate compounds can be initially screened computationally, with those showing promising predicted performance then selected for synthesis and experimental validation. The experimental results can also serve as data for subsequent model optimisation, enabling a gradual shift from traditional workflows reliant on manual curation towards a data-driven, iterative approach. At the same time, researchers' focus has begun to shift from repetitive screening towards the analysis of candidate results, the establishment of evaluation criteria and the design of experimental protocols. However, because these results are based on only a small number of cases, it cannot be concluded that generative AI can universally reduce the time and cost of drug R&D.

3. Practical challenges and solutions for the application of generative AI in drug R&D

Although generative AI has made some progress in the field of drug R&D, it still faces limitations in areas such as models, data, interpretability, practical application and industry regulation. At the level of model algorithms, generative AI still faces issues such as pattern collapse, unstable generation of three-dimensional structures, and the difficulty of expressing three-dimensional conformations using SMILES [3, 5], some models are also prone to favouring existing skeletons and lack the ability to generate structures under conditions involving chirality and multiple objectives [10]. To address the issues outlined above, appropriate molecular representations and generation methods can be selected according to the specific drug design task. To prevent models from repeatedly generating similar molecules or favouring existing skeletons, constraints such as molecular similarity and skeletal diversity can be incorporated into the generation and screening processes to reduce the repeated generation of similar molecules. For problems including insufficient physical constraints and limited accuracy of binding prediction, conformational optimisation and secondary screening can also be performed by integrating physical computational methods after candidate molecules are generated [14]. Furthermore, model evaluation should not rely solely on virtual physicochemical properties, it should also incorporate additional evaluation criteria such as skeletal diversity, synthetic feasibility and experimental activity, and the generated results must ultimately be validated experimentally [5].

At the data level, a lack of high-quality activity data and negative experimental data can limit model training and generalisation capabilities [2], at the same time, scaffold leakage in datasets may lead to artificially inflated evaluation metrics [10]. Peptide drug discovery also faces additional challenges, such as a shortage of annotated and immunogenicity data [9]. For scenarios with limited data, one approach is to first carry out pre-training using large-scale molecular data, followed by fine-tuning using a small amount of target-specific activity data, alternatively, virtual positive molecules can be obtained through HTS to augment the training data [14]. In terms of dataset partitioning, the Bemis-Murcko scaffold-based method can be adopted to reduce evaluation bias caused by structurally similar molecules appearing in both the training set and test set simultaneously, in the field of peptides, protein language models can also be utilised to alleviate the problem of insufficient annotated data.

In terms of model interpretability, deep generative models are often regarded as 'black boxes', making it difficult to explain the basis for their predictions of molecular activity, toxicity and other

DOI: 10.54254/2753-8818/2027.38309

properties [5]. When LLMs are involved in literature mining and the integration of target knowledge, they may also exhibit factual hallucinations and attribution errors. To address the above issues, explainable artificial intelligence could be further developed to identify the key molecular structures and features that influence prediction outcomes, as for factual errors in LLM-generated content, retrieval-enhanced generation could be utilised to combine the model's output with existing biomedical literature and database information [15]. However, regardless of how much the model's capabilities improve, their outputs still require assessment by experts in fields such as medicinal chemistry and pharmacology [1].

In terms of computational power and practical implementation, large-scale generative models and three-dimensional molecular simulations still require significant computational resources [10, 15], moreover, a gap remains between computational results and wet-lab experiments, and relatively few AI-driven de novo design approaches have progressed to clinical trials [5]. Currently, lightweight models such as VeGA, by reducing the number of parameters and lowering hardware requirements, enable laboratories with limited computational resources to carry out model fine-tuning and candidate compound generation using relatively small datasets of target activity data [10]. Furthermore, a closed-loop Design-Make-Test-Analyse automation system [3] can be established to link molecular synthesis, experimental testing and data feedback, thereby improving coordination between computational design and wet-lab experiments.

Across the entire R&D value chain and in the context of industry regulation, issues such as data silos, privacy protection, insufficient regulatory frameworks and a shortage of interdisciplinary talent still persist [2, 4]. To address the difficulty of directly sharing raw data between different organisations, federated learning can be employed to facilitate cross-organisational collaborative training, enabling different organisations to participate jointly in model training whilst retaining their local data [3]. At the same time, regulatory frameworks governing data sharing, intellectual property and generative AI-driven drug R&D need further refinement, and interdisciplinary training in pharmacy, biology and computer science should be strengthened.

4. Conclusions and outlook

Generative AI is driving a gradual shift in drug R&D from the traditional trial-and-error model towards a data-driven model of prediction, generation and validation. Through an analysis of existing technical approaches and case studies, this paper finds that the changes brought about by generative AI are no longer limited to the generation of new candidate compounds, but are also beginning to manifest in molecular evaluation, experimental feedback and the integration of different stages of the R&D process. Some of the screening and optimisation work that would otherwise require multiple rounds of experimentation can be carried out in advance in the computational space, this can then be combined with the results of subsequent experiments to proceed to the next round of molecular optimisation, enabling drug R&D to gradually evolve into an iterative process of 'generation–evaluation–feedback–regeneration–experimental validation'.

However, the analysis in this paper is still limited by the number of existing studies and case studies. At present, relatively few cases have involved generative AI in actual drug R&D and clinical validation. Furthermore, as the data, evaluation metrics and R&D stages used in different studies are not entirely consistent, it is difficult to make direct comparisons between existing results, nor is it possible to conclude, on the basis of a small number of cases, that generative AI can universally reduce the time and cost of drug R&D. Future work should further expand the application and long-term validation of generative AI in real-world drug R&D projects and establish more standardised evaluation methods to clarify the practical role that generative AI can play at different stages of

DOI: 10.54254/2753-8818/2027.38309

R&D. At present, generative AI is still primarily used as a supporting tool in drug R&D, the results generated by the models and computational predictions ultimately require experimental validation, whilst key stages such as target selection, candidate evaluation and R&D decision-making still require the involvement of researchers.

References

[1] Zhang, K., Yang, X., Wang, Y. et al. Artificial intelligence in drug development. *Nature Medicine, 31*, 45–59. https://doi.org/10.1038/s41591-024-03434-4

[2] Jarallah, S. J., Almughem, F. A., Alhumaid, N. K., Fayez, N. A., Alradwan, I., Alsulami, K. A., Tawfik, E. A., & Alshehri, A. A. (2025). Artificial intelligence revolution in drug discovery: A paradigm shift in pharmaceutical innovation. *International Journal of Pharmaceutics, 680*, 125789. https://doi.org/10.1016/j.ijpharm.2025.125789

[3] Gangwal, A., & Lavecchia, A. (2024). Unleashing the power of generative AI in drug discovery. *Drug Discovery Today, 29*(6), 103992. https://doi.org/10.1016/j.drudis.2024.103992

[4] Huanbutta, K., Burapapadh, K., Kraisit, P., Sriamornsak, P., Ganokratanaa, T., Suwanpitak, K., & Sangnim, T. (2024). Artificial intelligence-driven pharmaceutical industry: A paradigm shift in drug discovery, formulation development, manufacturing, quality control, and post-market surveillance. *European Journal of Pharmaceutical Sciences, 203*, 106938. https://doi.org/10.1016/j.ejps.2024.106938

[5] Lavecchia, A. (2024). Navigating the frontier of drug-like chemical space with cutting-edge generative AI models. *Drug Discovery Today, 29*(9), 104133. https://doi.org/10.1016/j.drudis.2024.104133

[6] Liu, D., Song, T., Na, K., & Wang, S. (2024). PED: A novel predictor-encoder-decoder model for Alzheimer drug molecular generation. *Frontiers in Artificial Intelligence, 7*, 1374148. https://doi.org/10.3389/frai.2024.1374148

[7] Albani, F. G., Alghamdi, S. S., Almutairi, M. M., & Alqahtani, T. (2025). Artificial intelligence-driven innovations in oncology drug discovery: Transforming traditional pipelines and enhancing drug design. *Drug Design, Development and Therapy, 19*, 5685–5707. https://doi.org/10.2147/DDDT.S509769

[8] Seyhan, A. A., & Carini, C. (2026). Artificial intelligence in drug discovery and development: Current landscape, challenges, and future perspectives. *Critical Reviews in Oncology/Hematology, 226*, 105461. https://doi.org/10.1016/j.critrevonc.2026.105461

[9] Lin, H., Vogel, H., & Zhang, H. (2026). A comparative review of artificial intelligence applications in small molecule versus peptide drug discovery. *International Journal of Molecular Sciences, 27*(7), 3142. https://doi.org/10.3390/ijms27073142

[10] Pietro Delre, Antonio Lavecchia, VeGA: A versatile generative architecture for bioactive molecules across multiple therapeutic targets. *J. Chem. Inf. Model., 65*(20), 10918–10931. https://doi.org/10.1021/acs.jcim.5c01606

[11] Blaschke, T., Olivecrona, M., Engkvist, O., Bajorath, J., & Chen, H. (2018). Application of generative autoencoder in de novo molecular design. *Molecular Informatics, 37*(1–2), 1700123. https://doi.org/10.1002/minf.201700123

[12] Guo, J., Fan, W., Amayri, M., & Bouguila, N. (2025). Deep clustering analysis via variational autoencoder with Gamma mixture latent embeddings. *Neural Networks, 183*, 106979. https://doi.org/10.1016/j.neunet.2024.106979

[13] Flynn, E. L., Shah, R., Dunn, I., Aggarwal, R., & Koes, D. R. (2025). PharmacoForge: Pharmacophore generation with diffusion models. *Frontiers in Bioinformatics, 5*, 1628800. https://doi.org/10.3389/fbinf.2025.1628800

[14] Filella-Merce, I., Molina, A., Díaz, L., Orzechowski, M., Berchiche, Y. A., Zhu, Y. M., Vilalta-Mor, J., Malo, L., Yekkirala, A. S., Ray, S., & Guallar, V. (2025). Optimizing drug design by merging generative AI with a physics-based active learning framework. *Communications Chemistry, 8*(1), 238. https://doi.org/10.1038/s42004-025-01635-7

[15] Mansour, G. K., & Sukkarieh, H. H. (2026). Generative artificial intelligence transitions pharmaceutical development from empirical screening to predictive molecular design and clinical trial optimization. *Pharmaceuticals, 19*(4), 614. https://doi.org/10.3390/ph19040614