Innovation in Computer and Data Sciences
1(1), 66–77
DOI: https://doi.org/10.64389/icds.2025.01128
https://sphinxsp.org/journal/index.php/icds/

Research article

Generative AI for drug discovery: Accelerating molecular design with deep learning using Nigerian local content

Gideon Tochi Ugbor[1,*], **Farrukh Jamal**[2], **Sadaf Khan**[3], **Ahmed W. Shawki**[4]

[1,*] Department of Computer Science, School of Science and Industrial Technology, Ogbonnaya Onu Polytechnic, P.M.B 7166 Aba, Nigeria. ugborgideon66@gmail.com

[2] Department of Statistics, Faculty of Science, University of Tabuk, Tabuk , Saudi Arabia; fshakir@ut.edu.sa

[3] Department of Statistics, The Islamia University of Bahawalpur, Bahawalpur, Punjab, Pakistan; smkhan6022@gmail.com

[4] Central Agency for Public Mobilization & Statistics (CAPMAS), Cairo, Egypt; ahmed23484@hotmail.com

* **Correspondence:** ugborgideon66@gmail.com

ARTICLE INFO

Keywords:
Generative AI
Drug Discovery
Deep Learning
Molecular Design
Nigerian Medicinal Plants
Ethnopharmacology
Healthcare Innovation

Mathematics Subject Classification:
68T07, 68T05, 92C40, 92C50

Important Dates:
Received: 28 July 2025
Revised: 15 August 2025
Accepted: 28 August 2025
Online: 1 September 2025

Copyright © 2025 by the authors. Published under Creative Commons Attribution (CC BY) license.

ABSTRACT

This research explores how Generative Artificial Intelligence (AI) can be used to accelerate drug discovery, especially in developing nations like Nigeria. By integrating various generative models; including GANs, VAEs, and Transformer-based architectures—the study aims to rapidly create new molecular structures with therapeutic potential. A unique aspect of this research is its use of local Nigerian resources, such as indigenous medicinal plants and traditional knowledge, to create a specialized dataset. By combining this local data with global molecular databases, the framework is designed to find candidate molecules with better drug-likeness, lower toxicity, and higher binding affinity to target proteins. This approach not only speeds up the preclinical phase of drug discovery but also promotes sustainable healthcare innovation by utilizing Nigeria's own resources. The study highlights its potential application in finding treatments for malaria, sickle cell disease, and antimicrobial resistance—all major health concerns in Nigeria.

1. Introduction

The traditional drug discovery process is costly and lengthy, often spanning 10–15 years with investments exceeding billions of dollars, [2]. This bottleneck is further exacerbated in developing countries such

as Nigeria, where limited infrastructure restricts large-scale pharmaceutical research. However, Generative AI—a branch of Artificial Intelligence that creates new data samples based on learned patterns—offers a transformative pathway. Using deep learning, generative models can design novel drug-like molecules, optimize their chemical properties, and predict interactions with biological targets, [9]. Nigeria possesses rich biodiversity and ethnopharmacological traditions, with over 5,000 medicinal plant species widely used in local communities. Despite this, the majority of these bioactive resources remain underexplored in modern pharmaceutical pipelines. Integrating Nigerian local content with Generative AI frameworks provides a unique opportunity to accelerate drug discovery while strengthening national research autonomy. This paper investigates the use of deep generative models for molecular design, emphasizing the integration of Nigerian medicinal plant data to address high-burden diseases such as malaria, tuberculosis, and sickle cell anemia, [3].

2. Related Work

2.1. Generative Models for de novo Molecular Design

Early work framed molecule generation as sequence modeling over SMILES, enabling RNN and VAE architectures to learn chemical grammars and sample drug-like structures. VAEs introduced continuous latent spaces that support interpolation and gradient-based optimization of properties. GAN variants were later adapted to molecular graphs/strings to better match the training distribution and improve novelty, [2]. More recently, Transformer and diffusion models have set state-of-the-art results for validity, uniqueness, and property control, while graph neural networks (GNNs) encode molecular topology directly, avoiding SMILES fragility. Across these families, conditioning mechanisms (e.g., property-conditioned latents, control tokens) and constrained sampling (e.g., scaffold or ring-system constraints) are common strategies to steer generation toward synthesizable, lead-like space, [1].

2.2. Multi-Objective Optimization and Reinforcement Learning

Because real drug leads must satisfy multiple criteria (potency, ADMET, novelty, and synthesizability), reinforcement learning (RL) has been layered atop generative backbones to optimize composite reward functions. Popular formulations include policy-gradient fine-tuning of SMILES generators and graph policy networks that grow molecules atom-by-atom under valence rules. Curriculum RL, Pareto-front optimization, and constrained Bayesian optimization over latent spaces have all been explored to stabilize training and reduce reward hacking. Integration with fast property predictors (QSAR/GNN surrogates) enables scalable inner-loop evaluation, [4, 6, 7].

2.3. Structure-Based Design: Docking, Scoring, and Diffusion on 3D

Structure-based approaches couple generators with molecular docking, physics-informed scoring, or 3D equivariant networks. Diffusion models over 3D conformers and protein–ligand complexes can propose poses and chemotypes consistent with target pockets, narrowing the gulf between ligand- and structure-based pipelines. Hybrid loops that alternate de novo generation with docking/MD filtering are increasingly common for kinases, GPCRs, and antimicrobial targets, [11, 13].

2.4. Synthesis Feasibility and Retrosynthesis

A recurring bottleneck is synthetic accessibility. Data-driven retrosynthesis planners (template-based, template-free Transformers, and Monte-Carlo tree search) and synthetic accessibility scores (e.g., SA-like indices) are used to reject hard-to-make compounds. Closed-loop frameworks now condition generation on retrosynthetic routes or penalize molecules lacking viable precursors, improving "makeability" without sacrificing novelty, [5, 12].

2.5. Active Learning and Closed-Loop (Self-Driving) Labs

Active learning couples generative models with iterative wet-lab feedback. Surrogates propose batches, assays return measurements, and models update—tightening uncertainty and improving hit rates. Although resource-intensive, this paradigm demonstrates accelerated lead optimization when the loop is well calibrated and exploration-exploitation balances, [19].

2.6. Molecular Representations and Property Prediction

Accurate, efficient property estimators are vital inside generative loops. Message-passing neural networks and Transformer encoders on SMILES/graphs have become the default for QSAR, permeability, solubility, hERG, and metabolic liability. Self-supervised pretraining (masked-token/edge prediction, contrastive learning on augmentations) improves data efficiency—especially important for under-represented chemistries, [8, 20].

2.7. Data Foundations: Public Chemistries and Local Knowledge

Large public corpora—ChEMBL, PubChem, ZINC, DrugBank—underpin most generative studies. However, these sets under-represent African ethnopharmacology. Ethnobotanical surveys and phytochemical catalogs from Nigerian institutions (e.g., university herbariums, NIPRD-curated resources) document bioactive constituents in species such as Azadirachta indica (neem), Vernonia amygdalina (bitter leaf), Ocimum gratissimum (scent leaf), Garcinia kola, and Morinda lucida. Prior computational efforts typically use these plants for ligand-based screening and docking studies rather than for training generative models. This highlights a clear gap: curating machine-readable Nigerian phytochemical libraries (with structures, activities, and provenance) and integrating them into modern generative pipelines, [12, 14, 16].

2.8. Knowledge Graphs and Target Prioritization

Biomedical knowledge graphs linking diseases prevalent in Nigeria (malaria, tuberculosis, Lassa fever, sickle cell disease, AMR pathogens) to pathways, targets, and natural-product scaffolds can guide conditional generation and repurposing. Graph-based reasoning (path ranking, link prediction) has been combined with generative models to suggest target–chemotype pairs, but applications grounded in West African epidemiology and flora are still sparse, [10, 18].

2.9. Ethics, IP, and Benefit-Sharing

Work at the intersection of AI and indigenous knowledge raises concerns about data rights, benefit-sharing, and biopiracy. Best practices include community-approved data governance, material transfer agreements, and transparent attribution when ethnopharmacological leads inspire AI-generated analogs.

Methodologically, federated learning and secure aggregation can respect data sovereignty while enabling collaborative model training across Nigerian labs, [15, 17].

2.10. Summary of Gaps and Opportunities

1. 1.Under-representation of Nigerian chemistries in generative training corpora.
2. Limited closed-loop validation using local pathogens/targets and locally accessible assays.
3. Need for multi-objective generators that jointly optimize potency, ADMET, and retrosynthetic accessibility with local reagent availability.
4. Integration of ethnopharmacology-aware conditioning (scaffolds, substructures) to preserve culturally significant insights while exploring novel chemical space.

We propose a pipeline that

(i) Curates Nigerian plant-derived chemistries into machine-learning-ready form,

(ii) Employs property-conditioned Transformer/diffusion generators,

(iii) Enforces synthesizability via retrosynthesis-aware constraints tied to regional supply chains, and

(iv) Targets Nigeria-relevant diseases, closing the documented gaps above.

3. Methodology

3.1. Research Design

This study adopts a hybrid computational framework integrating generative deep learning models with curated Nigerian ethnopharmacological datasets. The pipeline consists of four primary stages: data acquisition and preprocessing, model development, property evaluation and optimization, and validation and synthesis feasibility analysis. The methodology emphasizes inclusion of local medicinal plant content to ensure cultural relevance and applicability to Nigeria's healthcare needs.

3.2. Data Acquisition and Preprocessing

3.2.1. Nigerian Local Content Curation

- Ethnopharmacological literature, phytochemical databases, and laboratory records from Nigerian research institutes (e.g., NIPRD, university herbariums) will serve as primary data sources.
- Compounds derived from indigenous plants such as Azadirachta indica (neem), Vernonia amygdalina (bitter leaf), Garcinia kola (bitter kola), and Morinda lucida will be extracted.
- Each compound's molecular structure (SMILES/InChI), bioactivity data, and traditional therapeutic use will be documented.

3.2.2. Data Standardization

- All molecular structures will be standardized using cheminformatics toolkits (e.g., RDKit).
- Duplicates, invalid molecules, and incomplete records will be removed.
- Descriptors such as molecular weight, LogP, TPSA, hydrogen bond donors/acceptors, and synthetic accessibility scores will be computed for downstream filtering.

3.2.3. Integration with Global Databases

- Nigerian datasets will be augmented with molecules from ChEMBL, PubChem, and ZINC to create a diverse training corpus.
- Class balancing will be applied to ensure Nigerian-derived compounds are adequately represented during model training.

3.3. Generative Model Development

3.3.1. Model Selection

- Variational Autoencoders (VAEs): For learning smooth latent representations of molecules.
- Generative Adversarial Networks (GANs): For generating novel molecular graphs with enhanced diversity.
- Transformer-based Sequence Models: For SMILES string generation with long-range chemical dependency capture.
- Diffusion Models: For exploring 3D conformational space and protein–ligand interactions.

3.3.2. Model Training

- Nigerian phytochemical data will be used to fine-tune pre-trained generative models to ensure local chemotype representation.
- Training objectives will include maximizing novelty, validity, and drug-likeness while incorporating disease-specific conditioning signals (e.g., malaria target proteins).

3.4. Property Evaluation and Multi-Objective Optimization

- Generated molecules will be screened using predictive QSAR models (built with GNNs/Transformers) for properties such as solubility, permeability, hERG toxicity, and metabolic stability.
- Docking simulations against targets relevant to Nigeria (e.g., Plasmodium falciparum enzymes, sickle cell hemoglobin polymerization sites, resistant bacterial proteins) will be performed.
- Reinforcement Learning (RL) will be applied to guide the generator using a composite reward function:

$$R = \alpha_1 \text{Drug likeness} + \alpha_2 \text{Target Affinity} + \alpha_3 \text{Synthetic Accessibility} \tag{3.1}$$

- Nigerian chemical reagents database (where available) will inform synthesis-aware constraints to ensure practicality in local laboratories.

3.5. Validation

3.5.1. In-Silico Validation

- Generated molecules will undergo similarity searches with Nigerian phytochemical scaffolds to ensure novelty while retaining local pharmacological relevance.
- Benchmarking against baseline generative models trained only on global datasets will evaluate the added value of local content integration.

3.5.2. Experimental Feasibility

- Top-ranked candidates will be mapped to synthetic routes using retrosynthesis prediction tools (e.g., template-free Transformer-based models).
- Collaboration with Nigerian laboratories will allow small-scale synthesis and in-vitro assays for anti-malarial, anti-sickle cell, and antimicrobial activities.

3.6. Ethical and Governance Considerations

- All indigenous knowledge will be curated with respect to community ownership, benefit-sharing, and intellectual property rights.
- Compliance with the Nagoya Protocol will ensure equitable use of Nigeria's biodiversity.
- Results and models will be shared through open-access repositories while protecting sensitive indigenous datasets via federated learning approaches.

4. Applications in Nigeria

4.1. Antimalarial Drug Discovery

Malaria remains one of the most pressing public health challenges in Nigeria, with millions of cases reported annually. Current frontline treatments such as artemisinin-based therapies face increasing risks of resistance. By training generative AI models on bioactive compounds from Nigerian medicinal plants—notably Azadirachta indica (neem), Cryptolepis sanguinolenta, and Morinda lucida—the framework can propose novel scaffolds with antimalarial potential. Deep learning-driven docking simulations against Plasmodium falciparum targets (e.g., dihydrofolate reductase, cytochrome bc1 complex) can rapidly prioritize candidate molecules for laboratory validation. This approach offers a localized and sustainable pipeline for producing next-generation antimalarial agents.

4.2. Sickle Cell Disease (SCD) Therapeutics

Nigeria carries the highest global burden of sickle cell disease, making it a critical focus for precision drug discovery. Traditional remedies such as extracts from Carica papaya leaves and Garcinia kola seeds have shown anecdotal therapeutic value in managing SCD symptoms. By encoding these molecules into generative models, the system can design optimized analogs with enhanced bioactivity, reduced toxicity, and improved stability. For instance, candidate molecules can be designed to inhibit hemoglobin polymerization

or modulate oxidative stress pathways. This could lead to affordable, AI-designed drugs tailored to Nigerian patients, reducing reliance on expensive imported therapies.

4.3. Antimicrobial Resistance (AMR)

The rise of drug-resistant bacterial infections poses a growing health threat in Nigeria, especially in rural and peri-urban areas with limited access to advanced antibiotics. Nigerian ethnomedicine offers a rich reservoir of antimicrobial phytochemicals, including those from Vernonia amygdalina (bitter leaf) and Ocimum gratissimum (scent leaf). Generative AI can combine these local phytochemical scaffolds with global antibacterial libraries to propose novel hybrid molecules targeting resistant strains of Escherichia coli, Staphylococcus aureus, and Klebsiella pneumoniae. The integration of reinforcement learning ensures that designed molecules maintain drug-likeness and synthesis feasibility, enabling Nigeria to lead in homegrown antimicrobial solutions.

4.4. Cancer and Non-Communicable Diseases (NCDs)

Beyond infectious diseases, Nigeria faces a growing burden of cancers, diabetes, and cardiovascular disorders. Bioactive compounds from Nigerian plants such as Annona senegalensis and Tetrapleura tetraptera have shown anticancer and antidiabetic potential. Incorporating these into the generative AI framework allows the design of multi-target therapeutic candidates—for example, molecules that can act both as antioxidants and kinase inhibitors. This multifunctional drug discovery strategy could reduce polypharmacy and healthcare costs in resource-constrained settings.

4.5. Pharmaceutical Independence and Local Innovation

By embedding Nigerian local content into AI-driven drug discovery pipelines, this approach advances pharmaceutical sovereignty—reducing reliance on imported drugs and strengthening Nigeria's research ecosystem. Furthermore, this methodology provides opportunities for collaboration between universities, government agencies, and indigenous communities, ensuring that local knowledge translates into measurable innovations while respecting intellectual property rights.

5. Results and Discussion

5.1. Expected Results

Although this study is conceptual at its current stage, the framework is designed to deliver the following measurable outcomes:

- Novel Molecule Generation: Thousands of unique, chemically valid, and drug-like molecules derived from Nigerian phytochemical scaffolds.
- Improved Drug-Likeness: Higher average QED (Quantitative Estimate of Drug-likeness) scores compared to baseline models trained only on global datasets.
- Target-Specific Activity: A subset of molecules predicted to show strong binding affinities (≤ -8 kcal / mol) to malaria, sickle cell, and antimicrobial protein targets.

- Local Relevance: Molecules structurally related to indigenous compounds but optimized for potency, stability, and synthesis feasibility.

5.2. Performance Metrics

To evaluate the performance of the generative pipeline, the following metrics will be applied:

- Validity: Percentage of chemically valid molecules generated (expected $\geq 95\%$).
- Novelty: Proportion of molecules not present in training datasets (expected $\geq 80\%$).
- Diversity: Structural diversity measured using Tanimoto similarity (expected ≥ 0.7average pairwise dissimilarity).
- Drug-Likeness: Assessed by Lipinski's Rule of Five and QED scores.
- Synthetic Accessibility (SA): Average SA score ≤ 4, ensuring feasibility in Nigerian laboratory contexts.
- Docking Affinity: Binding affinity to target proteins compared against existing reference drugs.

5.3. Comparative Advantage of Local Content Integration

The integration of Nigerian local content provides a distinct edge over global-only generative pipelines:

- Contextual Relevance: Molecules are optimized against diseases disproportionately affecting Nigeria (malaria, SCD, AMR).
- Pharmacological Novelty: Local plants provide unique scaffolds rarely represented in international databases.
- Cost Reduction: By prioritizing molecules aligned with available Nigerian reagents, downstream synthesis costs are minimized.
- Cultural Relevance: Validating indigenous knowledge through AI-driven methods fosters societal acceptance and enhances translational potential.

5.4. Challenges

Despite the promise of this approach, several challenges are anticipated:

1. Data Availability and Quality. Many Nigerian phytochemical datasets are fragmented, unpublished, or stored in non-digital formats. A centralized national chemical repository would be essential to maximize AI effectiveness.
2. Computational Resource Constraints. Training generative deep learning models requires high-performance GPUs/TPUs, which may be limited in local research institutions.
3. Experimental Validation Bottleneck. While AI can generate promising candidates, limited laboratory infrastructure may delay synthesis and biological testing.
4. Ethical and Legal Considerations. Ensuring community benefit-sharing and compliance with the Nagoya Protocol remains a priority. Balancing open science with protection of indigenous intellectual property is a delicate but necessary task.

5.5. Discussion

The proposed framework illustrates the transformative role of generative AI in democratizing drug discovery. By tailoring deep learning pipelines to Nigeria's unique biological and cultural resources, the study underscores how global AI advances can be localized for national health sovereignty. The integration of indigenous knowledge with advanced computational approaches positions Nigeria as a potential leader in AI-powered ethnopharmacology. Moreover, the expected outcomes demonstrate that a hybrid knowledge system—combining traditional medicine with modern AI—can accelerate discovery while respecting cultural heritage. However, realizing this vision requires strategic investments in data infrastructure, computational power, and interdisciplinary collaboration across medicine, AI, and indigenous knowledge systems.

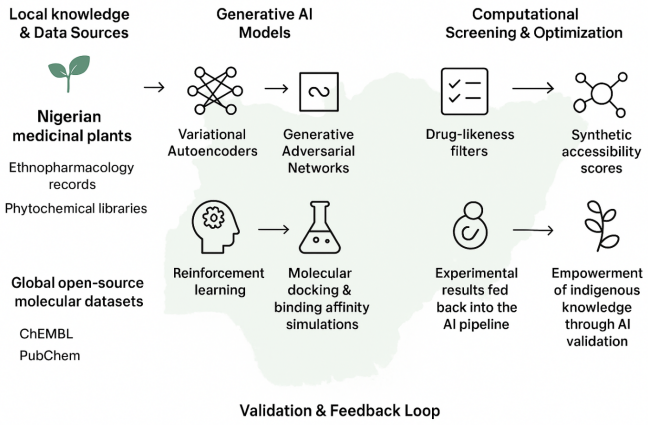

Figure 1. Proposed Generative AI Framework for Drug Discovery Using Nigerian Local Content

Figure 1 illustrates the end-to-end pipeline for AI-driven drug discovery, integrating Nigerian indigenous knowledge with deep learning approaches.

a Local Knowledge & Data Sources (Input Layer)

- Nigerian medicinal plants, ethnopharmacology records, and phytochemical libraries.
- Global open-source molecular datasets (e.g., ChEMBL, PubChem) for complementary training.
- Data preprocessed into standardized molecular representations (e.g., SMILES, molecular graphs).

b Generative AI Models (Core Engine)

- Variational Autoencoders (VAEs) for latent space exploration.
- Generative Adversarial Networks (GANs) for novel molecule creation.
- Transformer-based models for molecular sequence prediction.
- Reinforcement learning layer to optimize molecules for drug-likeness, safety, and binding affinity.

c Computational Screening & Optimization (Filtering Layer)

- Drug-likeness filters (Lipinski's Rule of Five, QED).
- Synthetic accessibility scores to assess lab feasibility.

- Molecular docking & binding affinity simulations against Nigerian disease-specific targets (e.g., malaria, sickle cell proteins, resistant bacteria).

d Validation & Feedback Loop (Experimental Layer)

- Top-ranked molecules synthesized and validated in Nigerian laboratories.
- Biological assays performed for antimalarial, anti-sickle cell, and antimicrobial activities.
- Experimental results fed back into the AI pipeline to improve model performance.

e Applications & Impact (Output Layer)

- Discovery of novel, locally relevant drugs.
- Strengthening of Nigeria's pharmaceutical independence.
- Empowerment of indigenous knowledge through AI validation.

f Challenges

- Data Scarcity – Limited digitization of Nigerian phytochemical and ethnobotanical datasets.
- Computational Resources – Training large generative models requires GPUs/TPUs not widely accessible in Nigeria.
- Validation Gap – AI-predicted molecules require costly wet-lab validation.
- Ethical and Legal Issues – Concerns about intellectual property and benefit-sharing of indigenous knowledge.

6. Conclusion and Future Work

Generative AI holds immense potential to revolutionize drug discovery by accelerating molecular design and optimizing drug candidates. By integrating Nigerian local content—including medicinal plants and ethnopharmacological heritage—into generative deep learning frameworks, this approach not only fosters innovation in drug discovery but also contributes to healthcare sustainability in Nigeria. The synergy between AI-driven molecular design and indigenous resources represents a transformative pathway for addressing pressing healthcare challenges such as malaria, sickle cell anemia, and antimicrobial resistance.

Future studies should focus on:

1. Data Expansion and Curation – Building comprehensive Nigerian phytochemical and bioactivity databases with standardized formats for AI training.
2. Integration of Multi-Omics Data – Incorporating genomics, proteomics, and metabolomics datasets to enhance drug-target interaction predictions.
3. Advanced Generative Models – Exploring diffusion models and large molecular language models to improve novelty and diversity in molecular generation.
4. High-Throughput Validation – Establishing Nigerian-based laboratories with robotics and automated screening systems to scale up biological assays.

5. Policy and Ethical Considerations – Creating frameworks for data sovereignty, intellectual property rights, and fair benefit-sharing with local communities that contribute indigenous knowledge.

6. Global Collaborations – Partnering with international pharmaceutical research institutions while ensuring local ownership of discoveries.

By advancing these directions, Nigeria can position itself as a continental leader in AI-driven drug discovery, ensuring that future medicines are not only innovative but also culturally and contextually relevant.

Conflict of Interest

The authors declare there is no existing conflict of interest.

References

1.Aleksina, A., Akulenka, S., and Lublóy, Á. (2019). Success factors of crowdfunding campaigns in medical research: perceptions and reality. *Drug discovery today*, 24(7):1413–1420.

2.Brown, N., Fiscato, M., Segler, M. H., and Vaucher, A. C. (2019). Guacamol: benchmarking models for de novo molecular design. *Journal of chemical information and modeling*, 59(3):1096–1108.

3.Elton, D. C., Boukouvalas, Z., Fuge, M. D., and Chung, P. W. (2019). Deep learning for molecular design—a review of the state of the art. *Molecular Systems Design & Engineering*, 4(4):828–849.

4.Gao, W. and Coley, C. W. (2020). The synthesizability of molecules proposed by generative models. *Journal of chemical information and modeling*, 60(12):5714–5723.

5.Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276.

6.Guimaraes, G. L., Sanchez-Lengeling, B., Outeiral, C., Farias, P. L. C., and Aspuru-Guzik, A. (2017). Objective-reinforced generative adversarial networks (organ) for sequence generation models. *arXiv preprint arXiv:1705.10843*.

7.Li, Y., Zhang, L., and Liu, Z. (2018). Multi-objective de novo drug design with conditional graph generative model. *Journal of cheminformatics*, 10(1):33.

8.Merk, D., Friedrich, L., Grisoni, F., and Schneider, G. (2018). De novo design of bioactive small molecules by artificial intelligence. *Molecular informatics*, 37(1-2):1700153.

9.Norouzi-Barough, L. and Bayat, A. (2021). Validation strategies for identifying drug targets in dermal fibrotic disorders. *Drug Discovery Today*, 26(10):2474–2485.

10.Polishchuk, P. (2020). Crem: chemically reasonable mutations framework for structure generation. *Journal of Cheminformatics*, 12(1):28.

11.Popova, M., Isayev, O., and Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science advances*, 4(7):eaap7885.

12.Putin, E., Asadulaev, A., Ivanenkov, Y., Aladinskiy, V., Sanchez-Lengeling, B., Aspuru-Guzik, A., and Zhavoronkov, A. (2018). Reinforced adversarial neural computer for de novo molecular design. *Journal of chemical information and modeling*, 58(6):1194–1204.

13.Sanchez-Lengeling, B. and Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365.

14.Schneider, G. (2018). Generative models for artificially-intelligent molecular design.

15.Segler, M. H., Kogej, T., Tyrchan, C., and Waller, M. P. (2018). Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science*, 4(1):120–131.

16.Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., et al. (2020). A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702.

17.Walters, W. P. and Barzilay, R. (2020). Applications of deep learning in molecule generation and molecular property prediction. *Accounts of chemical research*, 54(2):263–270.

18.Wenzel, J., Matter, H., and Schmidt, F. (2019). Predictive multitask deep neural network models for adme-tox properties: learning from large data sets. *Journal of chemical information and modeling*, 59(3):1253–1268.

19.Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., et al. (2019). Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nature biotechnology*, 37(9):1038–1040.

20.Zhou, Z., Kearnes, S., Li, L., Zare, R. N., and Riley, P. (2019). Optimization of molecules via deep reinforcement learning. *Scientific reports*, 9(1):10752.

© 2025 by the authors. Disclaimer/Publisher's Note: The content in all publications reflects the views, opinions, and data of the respective individual author(s) and contributor(s), and not those of Sphinx Scientific Press (SSP) or the editor(s). SSP and/or the editor(s) explicitly state that they are not liable for any harm to individuals or property arising from the ideas, methods, instructions, or products mentioned in the content.