Generative AI Meets Natural Product Chemistry: A Critical Review of Molecular Design, Synthetic Feasibility, Bioactivity Prediction, and Safety Validation

Luca Ferraro[1]*, Matteo Ricci[1], Giulia Moretti[2], Paolo Greco[3]

[1]Department of Artificial Intelligence for Medicinal Mushroom and Fungal Metabolites, Faculty of Pharmacy, University of Bologna, Bologna, Italy.
[2]Department of Machine Learning for Drug Synergy and Toxicity, Faculty of Pharmacy, University of Turin, Turin, Italy.
[3]Department of Phytochemical Informatics, Faculty of Pharmacy, Sapienza University of Rome, Rome, Italy.

ABSTRACT

Generative AI is increasingly used to support molecular design, chemical space exploration, and compound prioritisation, creating new opportunities for natural product chemistry while also raising important questions about chemical realism and validation. This critical review evaluates how generative models may contribute to natural product-inspired scaffold generation, analogue design, molecular novelty, diversity, bioactivity prediction, synthetic feasibility assessment, ADMET screening, toxicity prediction, and safety validation. The review takes a cautious interpretive approach, treating generated structures as design hypotheses rather than validated chemical, biological, or translational entities. Particular attention is given to the structural complexity of natural products, including stereochemistry, biosynthetic diversity, functional-group density, and the difficulty of distinguishing natural product likeness from biological relevance. The synthesis argues that generative AI can expand ideation, prioritise candidate chemical regions, and integrate multi-objective design constraints, but its outputs remain vulnerable to invalid structures, unrealistic stereochemistry, synthetic accessibility overestimation, false-positive bioactivity prediction, target-context mismatch, ADMET uncertainty, toxicity underprediction, benchmark bias, data leakage, and weak uncertainty reporting. The main conclusion is that generative AI has genuine value for natural product-inspired molecular exploration only when coupled to chemically informed constraints, retrosynthetic review, experimentally grounded bioactivity testing, safety assessment, uncertainty calibration, and transparent reporting. Future research should prioritise domain-specific datasets, stereochemistry-aware generation, synthesis-linked evaluation, biologically meaningful objectives, and validation pipelines that preserve the expertise of natural product chemists.

Key Words: *Generative AI, Natural product chemistry, Molecular design, Synthetic feasibility, Bioactivity prediction, Safety validation*

eIJPPR 2026; 16(1):60-70

HOW TO CITE THIS ARTICLE: Ferraro L, Ricci M, Moretti G, Greco P. Generative AI Meets Natural Product Chemistry: A Critical Review of Molecular Design, Synthetic Feasibility, Bioactivity Prediction, and Safety Validation. Int J Pharm Phytopharmacol Res. 2026;16(1):xx-xx. https://doi.org/10.51847/qSMlQOLj14

INTRODUCTION

Natural products remain a central source of chemical inspiration for drug discovery because they provide structurally diverse, biologically shaped molecules with functional-group arrangements, stereochemical richness, and scaffold architectures that are often under-represented in conventional synthetic libraries [1]. Their historical contribution to drug discovery also demonstrates that natural product-derived and natural product-inspired

Corresponding author: Luca Ferraro
Address: Department of Artificial Intelligence for Medicinal Mushroom and Fungal Metabolites, Faculty of Pharmacy, University of Bologna, Bologna, Italy.
E-mail: ✉ luca.ferraro@unibo.it
Received: 17 November 2025; **Revised:** 02 February 2026; **Accepted:** 08 February 2026

structures can inform therapeutic innovation, but this contribution should not be confused with automatic developability, safety, or translational readiness [2].

Generative AI has emerged as a major development in molecular design because it can invert the usual screening logic: instead of only evaluating existing molecules, models can propose new structures under specified design objectives [3]. In de novo molecular design, this capability includes generation, optimisation, and prioritisation of candidate structures, yet the quality of generated molecules depends on the design task, training data, objective functions, chemical representation, and downstream evaluation criteria [4].

Natural product chemistry creates special challenges for generative AI because natural products are not merely "complex molecules" in a generic sense. Their biological meaning may depend on stereochemical configuration, biosynthetic origin, conformational behaviour, target context, metabolic stability, and matrix-specific safety considerations. A generated molecule may resemble a natural product by scaffold or descriptor profile while lacking the stereochemical precision, biological context, synthetic plausibility, or safety evidence needed to support a stronger claim.

This article critically reviews generative AI at the intersection of molecular design and natural product chemistry. Its aim is not to present generative chemistry as a replacement for natural product expertise, experimental isolation, synthesis, pharmacology, or toxicology. Instead, it evaluates where generative AI can support ideation and prioritisation, where evidence remains weak, and why chemical realism, synthetic feasibility, bioactivity validation, ADMET evidence, toxicity assessment, uncertainty reporting, and experimentally grounded safety validation are necessary before natural product-inspired generated molecules can be interpreted beyond hypothesis status.

Generative AI in natural product chemistry

Deep learning has broadened computational drug discovery by enabling molecular representation learning, property prediction, and design-oriented workflows, but these capabilities depend strongly on data quality, domain relevance, and model evaluation [5]. In natural product chemistry, this dependence is especially important because the training data must represent stereochemical richness, biosynthetic diversity, functional-group patterns, and biologically relevant chemical contexts rather than only generic small-molecule distributions.

Generative AI approaches relevant to natural product chemistry include variational autoencoders, generative adversarial networks, reinforcement learning systems, graph generative models, transformer-based molecular generators, diffusion models, molecular language models, and foundation-model-assisted design. Evidence from natural-product-inspired design indicates that machine intelligence can be tuned toward biologically motivated chemical hypotheses, as shown in work on natural-product-inspired retinoid X receptor modulator design [6]. However, such outputs should be interpreted as candidates for review and testing, not as evidence of confirmed activity, safety, or synthesizability.

Natural-inspired molecular optimisation can also support exploration of chemical regions adjacent to natural product-like space by combining generation with optimisation objectives [7]. This is useful when the aim is to search beyond known compounds, propose scaffold variations, or identify structures that satisfy multiple computational constraints. The limitation is that optimisation can privilege what the objective function rewards, which may not coincide with biological relevance, synthetic feasibility, or safety.

Molecular language models are particularly relevant because they treat chemical strings as learnable sequences and can be adapted toward natural product-like compound generation [8]. Their appeal lies in scalable generation and flexible conditioning, but natural product chemistry exposes important weaknesses: string validity does not guarantee stereochemical correctness, structural novelty does not guarantee value, and natural product likeness does not establish target engagement, pharmacological activity, or safety. Foundation-model-assisted design may expand these capabilities, but its outputs still require domain-specific interpretation and validation.

Molecular design and chemical space exploration

Natural product-inspired molecular design can use generative models to construct virtual libraries that resemble natural product chemical space, helping researchers explore scaffold families, analogue concepts, and under-sampled chemical regions [9]. This function is valuable as an ideation tool because it can generate hypotheses that would be difficult to enumerate manually. Its central limitation is that virtual library membership does not establish chemical usefulness, biological relevance, or experimental feasibility.

Variational autoencoder-based molecular design introduced the idea that molecules can be embedded into continuous latent spaces where optimisation and interpolation become possible [10]. For natural product-inspired chemistry, this suggests a route to traverse structural neighbourhoods around complex scaffolds, but latent continuity may smooth over chemically discontinuous realities such as stereochemical constraints, conformational dependence, and synthetic route discontinuities. Reinforcement learning adds another layer

by optimising structures toward specified objectives, although this creates the risk that generated molecules satisfy the reward function rather than the broader requirements of medicinal or natural product chemistry [11].

Recurrent molecular generators can produce focused libraries by learning chemical syntax and distributional patterns from training sets [12]. Activity-directed reinforcement learning can further bias molecular generation toward predicted target activity, but such predictions remain dependent on the underlying predictive model and its applicability domain [13]. In natural product-inspired analogue design, this is a major interpretive boundary: a model may generate structures that look plausible and score well computationally while still failing because of assay mismatch, target-context mismatch, stereochemical error, or poor experimental behaviour.

Scaffold hopping from natural products to synthetic mimetics illustrates the appeal of preserving biologically meaningful structural logic while improving chemical tractability or exploring alternative scaffolds [14]. Benchmarking frameworks for de novo molecular design can evaluate properties such as validity, novelty, and diversity, yet these metrics do not establish synthetic feasibility, biological activity, ADMET suitability, or safety [15]. Therefore, molecular design and chemical space exploration should be framed as prioritisation and hypothesis generation rather than evidence of translational readiness.

Table 1 summarises generative AI functions used for molecular design and natural product-inspired chemical space exploration.

Table 1. Generative AI Functions in Natural Product-Inspired Molecular Design: Scaffold Generation, Analogue Design, Chemical Space Exploration, Novelty, Diversity, Natural Product Likeness, Chemical Validity, Stereochemical Representation, and Interpretation Boundaries

Generative design function	Core design question	Potential value	Required evidence	Main limitation	Validation requirement	Failure risk	Critical interpretation
Natural product-inspired generation	Does the output meaningfully resemble natural product chemical space?	Expands ideation around biologically shaped scaffolds	Structural comparison, descriptor context, scaffold analysis	Natural product likeness may be superficial	Expert chemical review and biological testing	Natural product-like but biologically irrelevant molecules	Supports hypothesis generation only
Scaffold generation	Is the scaffold chemically plausible and interpretable?	Enables new scaffold concepts and scaffold hopping	Scaffold validity, novelty, precedent, stability review	Novel scaffold does not imply useful scaffold	Chemical plausibility and synthesis review	Novelty without value	Useful only when linked to function and feasibility
Analogue design	Does the generated analogue preserve meaningful structure–activity logic?	Supports local exploration around known chemotypes	SAR context, matched-pair reasoning, stereochemical checks	Small changes may disrupt activity	Experimental analogue testing	Activity cliffs and misleading similarity	Analogue status must not imply confirmed activity
Chemical space exploration	Which regions are being explored and why?	Broadens prioritisation beyond known libraries	Diversity, coverage, similarity, constraint analysis	Exploration may enter irrelevant or infeasible regions	Domain-aware filtering and expert triage	Large but low-value chemical spaces	Exploration is not equivalent to discovery
Molecular novelty	Is the molecule structurally new in a	Reduces simple reproduction of	Similarity and substructure comparison	Novelty may be trivial or undesirable	Novelty audit plus functional evaluation	Overclaiming innovation	Novelty must be separated from value

ISSN (Online) 2249-6084 (Print) 2250-1029

	meaningful way?	known molecules					
Molecular diversity	Does the generated set avoid narrow clustering?	Improves breadth of hypothesis generation	Internal diversity and scaffold diversity checks	Diversity can be chemically irrelevant	Diversity assessment linked to design objectives	Mode collapse or unfocused diversity	Diversity must be interpreted with task relevance
Natural product likeness	Does likeness correspond to useful biological hypotheses ?	Helps prioritise NP-adjacent structures	NP-likeness metrics and biological rationale	Likeness does not establish activity	Bioactivity and target validation	Misleading biological inference	Natural product-like means structurally similar, not validated
Chemical validity	Is the output a chemically valid structure?	Removes impossible structures early	Valence, charge, aromaticity, tautomer and structure checks	Validity is a minimal threshold	Cheminformatics and manual inspection	Invalid or artefactual molecules	Validity does not prove feasibility or safety
Stereochemical representation	Are stereocentres defined and chemically plausible?	Preserves a key feature of natural products	Chiral annotation, 3D review, stereochemical consistency	String formats may omit or distort stereochemistry	Stereo-aware generation and expert review	Unrealistic or ambiguous stereochemistry	Stereochemical uncertainty weakens biological interpretation
Physicochemical plausibility	Are generated properties compatible with intended use?	Supports early filtering	Descriptor review and structural alerts	Plausibility depends on context	Experimental and ADMET follow-up	Plausible-looking but unsuitable molecules	Plausibility is supportive, not confirmatory

Synthetic feasibility and bioactivity prediction

Synthetic feasibility is a decisive boundary for generated natural product-inspired molecules because chemical validity and synthetic accessibility scores do not necessarily show that a molecule can be prepared in the laboratory. Analyses of molecules proposed by generative models show why synthesizability needs dedicated assessment rather than reliance on generative output quality alone [16]. This issue is amplified for natural product-inspired structures because stereochemical density, ring complexity, unusual functional groups, and biosynthetic-like motifs may complicate route planning.

Retrosynthetic planning can help triage generated molecules by proposing possible disconnections and route hypotheses, but these plans remain dependent on reaction knowledge, training data, and expert interpretation [17]. Reaction prediction models add further support by estimating likely chemical transformations, especially when uncertainty is explicitly represented [18]. For generated natural product-inspired analogues, uncertainty is not a minor technical detail; it is central to deciding whether a proposed structure should be prioritised, redesigned, or retained only as a speculative design idea.

Retrosynthesis tools can support route plausibility review by providing fast planning workflows, but a proposed computational route should not be treated as experimental evidence of synthesis [19]. More broadly, machine learning in computer-aided synthesis planning depends on available reaction data and may underperform when generated structures fall outside well-represented chemistry [20]. Natural product-inspired molecules may contain functional arrangements, stereochemical patterns, or ring systems that are poorly covered by standard reaction datasets, so expert synthetic review remains essential.

Bioactivity prediction, target prediction, docking hypotheses, multi-objective optimisation, and ADMET-related prioritisation can help rank generated structures, but they must remain within clear evidentiary limits. Molecular machine-learning benchmarks provide useful comparison tasks for property prediction, yet benchmark performance may not translate directly to natural product-inspired domains or specific biological assays [21].

Learned molecular representations also influence prediction reliability, which means predicted activity, target relevance, or ADMET suitability should be interpreted as prioritisation evidence rather than proof of pharmacological activity or safety [22].

Table 2 maps synthetic feasibility and bioactivity prediction requirements for evaluating generated natural product-inspired molecules.

Table 2. Synthetic Feasibility and Bioactivity Prediction in Generative Natural Product Chemistry: Synthetic Accessibility, Retrosynthetic Plausibility, Reaction Feasibility, Target Relevance, Bioactivity Prediction, ADMET Screening, Uncertainty, and Validation Needs

Evaluation domain	Core question	Input evidence needed	Potential output	Interpretation limitation	Validation requirement	Risk if ignored	Critical synthesis role
Synthetic accessibility	Is the molecule likely to be synthetically approachable?	Structural complexity, functional groups, stereochemistry, accessibility estimate	Early feasibility triage	Accessibility scores may overestimate laboratory feasibility	Chemist review and route assessment	Unsynthesizable generated structures	Separates chemical validity from feasibility
Retrosynthetic plausibility	Can credible routes be proposed?	Route suggestions, disconnections, reaction precedent	Candidate route hypotheses	Planner output may be incomplete or impractical	Expert retrosynthetic review	Retrosynthetic implausibility	Links generation to practical chemistry
Reaction feasibility	Are proposed transformations chemically reasonable?	Reaction precedent, substrate scope, uncertainty	Feasibility ranking	Reaction models depend on training coverage	Reaction-condition and precedent review	Failed synthesis translation	Adds realism to generated designs
Stereochemical feasibility	Can the required stereochemistry be controlled or assigned?	Chiral centres, stereochemical notation, route stereocontrol	Stereo-aware feasibility judgement	Generated structures may omit or misuse stereochemistry	Stereochemical synthesis and analytical confirmation	Incorrect analogue identity	Critical for natural product-inspired interpretation
Biosynthetic plausibility	Does biosynthetic logic matter for the design claim?	Biosynthetic context where relevant	Mechanistic inspiration or prioritisation	Biosynthetic plausibility is not required for every analogue	Expert natural product chemistry review	Misleading natural product analogy	Clarifies meaning of "natural product-inspired"
Bioactivity prediction	Does the molecule score as active in a relevant model?	Assay data, model domain, target context	Prioritisation hypothesis	Prediction is not experimental activity	Biochemical or cellular assay confirmation	False-positive prioritisation	Keeps bioactivity claims cautious
Target relevance	Is the proposed target biologically meaningful?	Disease context, pathway evidence, target expression	Mechanistic hypothesis	Target prediction can be context-mismatched	Target engagement and pathway validation	Irrelevant target hypothesis	Links chemistry to pharmacology
Docking hypothesis	Does docking suggest plausible	Structure preparation, binding-site	Binding hypothesis	Docking does not prove	Orthogonal binding and	Overinterpreted docking score	Restricts docking to

ISSN (Online) 2249-6084 (Print) 2250-1029

	binding where relevant?	rationale, pose review		binding or activity	activity assays		hypothesis status
Multi-objective optimisation	Are design objectives balanced realistically?	Objective weights, constraints, trade-off analysis	Pareto-like prioritisation	Optimisation may game weak objectives	Sensitivity analysis and experimental follow-up	Artificially optimised molecules	Prevents objective-function overclaiming
ADMET screening	Are early liability signals considered?	ADMET models, applicability domain, uncertainty	Safety triage	ADMET prediction does not validate safety	Experimental ADMET testing where appropriate	False reassurance	Connects design to safety evidence
Uncertainty reporting	How confident is each prediction?	Calibration, applicability domain, confidence estimates	Risk-aware prioritisation	Uncertainty may be absent or poorly calibrated	Calibration and external validation	Overconfident prioritisation	Makes computational evidence interpretable
Experimental validation	Has the prioritised molecule been made and tested?	Synthesis, identity confirmation, biological and safety assays	Evidence beyond computation	Often missing at design stage	Reproducible chemical and biological validation	Translational overclaiming	Defines final evidence boundary

Safety validation and failure modes

Safety validation is the point at which generative AI in natural product-inspired chemistry most clearly moves beyond molecular design aesthetics. Failure modes in molecule generation and optimisation include invalid outputs, poor diversity, optimisation artefacts, and objective-driven structures that appear successful under computational metrics while remaining chemically or biologically questionable [23]. In natural product-inspired design, these risks are intensified because generated molecules may appear complex, scaffold-rich, or natural product-like without preserving the stereochemical, physicochemical, or mechanistic features that make a natural product biologically meaningful.

Benchmarking platforms for molecular generation can improve comparability across models, but they cannot fully determine whether a generated molecule is synthetically feasible, biologically relevant, or safe [24]. Benchmark bias becomes especially important when evaluation tasks are dominated by generic small-molecule datasets rather than natural product-like structures, stereochemically rich molecules, or assay contexts relevant to natural product pharmacology. Data leakage can also inflate apparent predictive quality, making generated outputs look more reliable than they are.

The safety risks of generative molecular design are not limited to benign optimisation failure. Work on the dual-use potential of AI-powered drug discovery shows that generative workflows can be redirected toward harmful chemical objectives, demonstrating why safety-aware design, toxicity filters, and responsible interpretation are necessary parts of generative chemistry [25]. For natural product-inspired chemistry, this means that "biologically active-looking" generated molecules require careful hazard evaluation rather than celebratory framing.

ADMET prediction can support early triage of generated structures, but computational ADMET outputs do not validate safety [26]. Off-target liability assessment remains necessary because small molecules can interact with unintended biological targets even when the intended target hypothesis appears plausible [27]. Graph-based modelling of adverse interaction contexts can help represent complex safety relationships, but such modelling remains context-dependent and should not substitute for experimental safety evaluation [28]. Deep learning approaches to pharmacokinetic and toxicity prediction may strengthen early prioritisation, but toxicity underprediction remains a critical risk when generated molecules are advanced without orthogonal evidence [29].

Figure 1 illustrates major failure modes that can arise when generative AI is applied to natural product-inspired molecular design without adequate chemical, synthetic, biological, and safety validation.

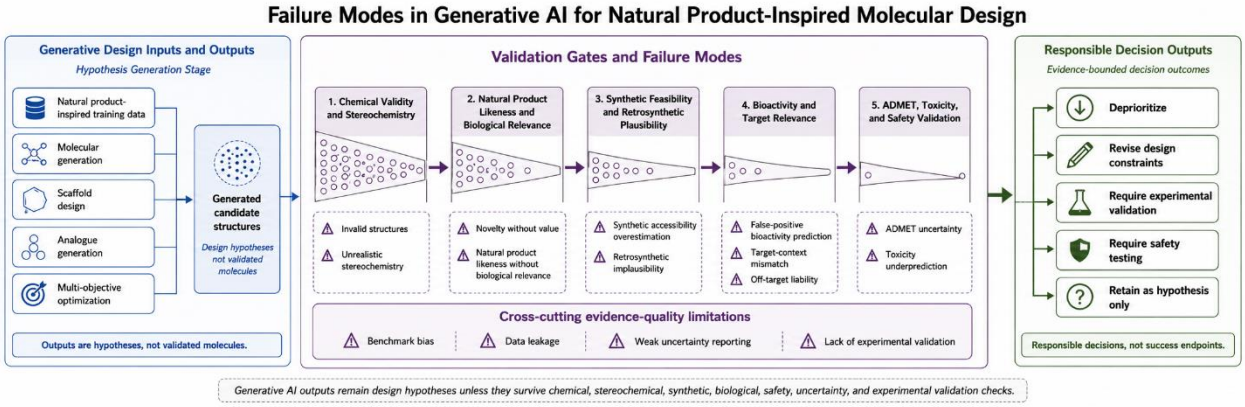

Figure 1. Failure Modes in Generative AI for Natural Product-Inspired Molecular Design

Alt-text description

A conceptual failure-mode map showing generative AI outputs passing through chemical validity, stereochemistry, synthetic feasibility, bioactivity prediction, ADMET and toxicity screening, benchmark bias, uncertainty reporting, and experimental validation checks.

Critical synthesis

The critical value of generative AI in natural product chemistry lies in hypothesis generation, design-space expansion, and candidate prioritisation rather than proof. Machine intelligence can help generate natural product-inspired molecular ideas, but even bioactivity-oriented outputs should be interpreted as proposed chemical hypotheses until synthesis, analytical confirmation, biological testing, and safety assessment are completed [30]. This distinction is essential because generated molecules may look plausible while remaining experimentally unmade, biologically untested, or safety-uncertain.

Interpretability is a central requirement for credible generative molecular design because researchers need to understand why structures were proposed, which features drove prediction, and where model confidence is limited [31]. In natural product-inspired chemistry, explainability is especially important because small structural, stereochemical, or conformational differences can change biological interpretation. Without interpretable reasoning, generated analogues may be difficult to distinguish from artefacts of model optimisation.

Molecular representation is another critical determinant of reliability. Geometric and graph-based molecular learning approaches can better represent structural relationships than purely descriptor-based approaches, but representation choices still influence what models learn, what they omit, and how generated molecules are evaluated [32]. For natural product-inspired structures, representation must handle stereochemistry, ring systems, conformational features, and functional-group context with enough fidelity to support meaningful chemical review.

Bioactivity prediction remains one of the most attractive uses of AI-guided design, but it is also one of the most vulnerable to overinterpretation. Activity cliffs show that small structural changes can produce large biological differences, exposing the fragility of predictions when generated analogues sit near steep structure–activity boundaries [33]. This means that similarity to an active natural product, predicted target activity, or favourable computational ranking should be treated as a prioritisation signal rather than evidence of pharmacological activity.

Overall, the reviewed evidence supports a disciplined interpretation: generative AI can add value by producing chemically diverse ideas, proposing scaffold transitions, and integrating multiple design objectives, but the evidence remains insufficient when outputs are evaluated only by validity, novelty, diversity, predicted activity, or ADMET scores. A credible workflow must combine domain-specific generation with stereochemical checks, route-aware synthesis review, target-context justification, calibrated uncertainty, off-target and toxicity assessment, and experimental validation.

Table 3 presents the critical synthesis of value, limitations, failure modes, and validation priorities for generative AI in natural product chemistry.

Table 3. Critical Synthesis of Generative AI in Natural Product Chemistry: Design Value, Chemical Realism, Synthetic Feasibility, Bioactivity Prediction, ADMET and Toxicity Risk, Failure Modes, Validation Priorities, and Future Research Needs

Critical synthesis theme	Potential contribution	Main evidence gap	Failure mode	Validation priority	Safety relevance	Translational boundary	Research agenda implication
Design-space expansion	Generates natural product-inspired ideas beyond known compounds	Limited proof that expanded space is biologically useful	Irrelevant or infeasible chemical regions	Domain-aware chemical review	Prevents prioritisation of unsuitable structures	Exploration is not discovery	Build NP-specific design benchmarks
Scaffold generation	Supports scaffold hopping and new chemotype ideation	Weak link between generated scaffold and biological function	Novelty without value	Scaffold plausibility and SAR review	Avoids misleading scaffold claims	Novel scaffold is not therapeutic relevance	Combine scaffold design with mechanism
Analogue generation	Proposes structural variants around known chemotypes	Activity often not experimentally confirmed	Activity cliffs and stereochemical mismatch	Experimental analogue testing	Reduces false activity inference	Analogue status is not bioactivity proof	Use assay-linked analogue validation
Natural product likeness	Prioritises NP-adjacent chemical regions	Likeness does not show target relevance	NP-like but inactive molecules	Bioactivity and target validation	Avoids biological overinterpretation	NP-like is descriptive only	Define context-aware NP-likeness metrics
Chemical realism	Filters invalid or implausible structures	Validity checks may be too shallow	Invalid structures or reactive motifs	Structural curation and stability review	Removes unsafe or impossible candidates	Validity is not feasibility	Add stereo-aware and reactivity-aware filters
Synthetic feasibility	Links design to possible laboratory preparation	Accessibility scores may overestimate feasibility	Unsynthesizable outputs	Retrosynthetic and chemist review	Prevents wasted experimental effort	Score does not prove synthesis	Integrate route-aware generation
Reaction feasibility	Evaluates whether proposed transformations are credible	Reaction models may lack domain coverage	Implausible route planning	Reaction precedent and uncertainty review	Reduces failed synthesis translation	Planned route is hypothetical	Expand reaction data for NP-like chemistry
Bioactivity prediction	Prioritises compounds for testing	Model domain may not match NP-like structures	False-positive activity	Biochemical and cellular assays	Prevents unsupported pharmacology claims	Prediction is not activity	Use applicability-domain reporting
Target relevance	Connects generated molecules to biological context	Target predictions may be context-mismatched	Target-context mismatch	Target engagement evidence	Reduces irrelevant safety or efficacy claims	Target hypothesis is not mechanism	Couple models to pathway biology

ISSN (Online) 2249-6084 (Print) 2250-1029

ADMET prediction	Screens early liabilities	Computational outputs remain uncertain	False safety reassurance	Experimental ADMET testing	Improves early risk triage	ADMET prediction is not safety validation	Calibrate ADMET uncertainty
Toxicity prediction	Flags possible harmful structures	Underprediction remains possible	Toxicity underprediction	Orthogonal toxicity assays	Central to safety validation	Negative prediction is not absence of toxicity	Conservative toxicity-aware optimisation
Benchmark transparency	Improves model comparison	Benchmarks may not reflect NP chemistry	Benchmark bias and data leakage	External and domain-specific testing	Prevents inflated confidence	Benchmark success is not translation	Standardise leakage and domain audits
Uncertainty reporting	Supports risk-aware prioritisation	Often weak or missing	Overconfident predictions	Calibration and confidence reporting	Helps avoid unsafe prioritisation	Confidence is not validation	Require uncertainty in reporting standards
Experimental validation	Converts hypotheses into evidence	Often absent in generative studies	Overclaiming readiness	Synthesis, identity confirmation, bioassays, safety testing	Final safety and biological gate	No experiment means no proof	Build validation-linked design pipelines

Figure 2 presents a critical synthesis framework for evaluating generative AI outputs in natural product chemistry across design value, synthetic feasibility, bioactivity prediction, safety validation, and translational boundaries.

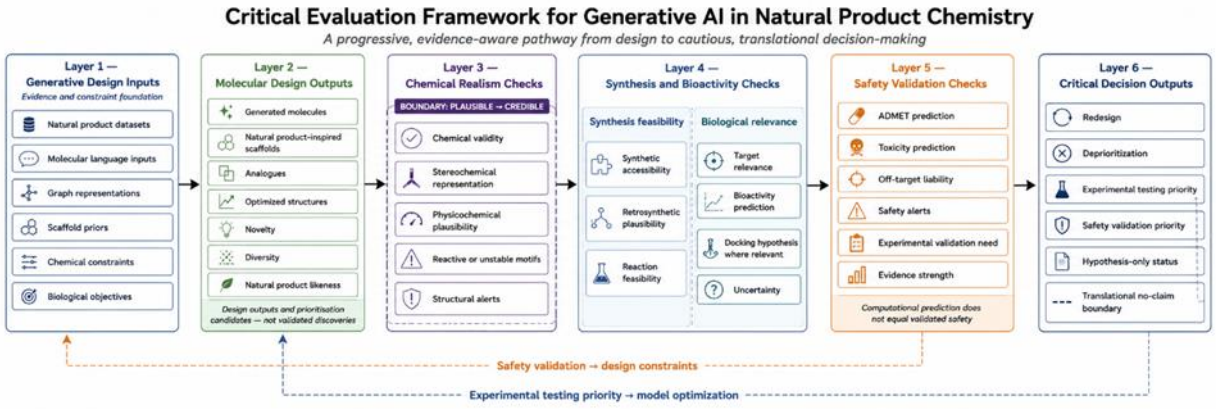

Figure 2. Critical Evaluation Framework for Generative AI in Natural Product Chemistry

Alt-text description

A conceptual evaluation framework showing generated natural product-inspired molecules assessed through chemical realism, novelty, diversity, natural product likeness, synthetic feasibility, bioactivity prediction, safety validation, uncertainty, experimental confirmation, and translational boundaries.

Future research agenda

Future work should prioritise higher-quality natural product datasets that preserve stereochemistry, structural provenance, assay context, and chemical annotation quality. Chemical language models for natural product discovery point toward domain-aware generation, but their future value will depend on whether they can represent natural product-specific complexity while reporting uncertainty and avoiding unsupported claims of biological relevance [34]. Dataset curation should therefore include stereochemical completeness, duplicate control, scaffold annotation, assay reliability, and clear separation between training, validation, and external testing sets.

Generative design should also become more tightly linked to synthetic feasibility, retrosynthetic planning, biological objectives, and safety-aware optimisation. Deep generative molecular design is reshaping ideation workflows, but the next step is not simply larger models or more generated

molecules; it is integration with route-aware generation, applicability-domain reporting, target-context evidence, ADMET uncertainty, toxicity triage, and experimental validation pipelines [35]. Natural product chemists, synthetic chemists, pharmacologists, and toxicologists should be involved early so that design objectives reflect feasible chemistry and meaningful biology.

A further priority is transparent evaluation of structural novelty and value. AI-designed molecules require careful novelty assessment because apparent novelty can arise from superficial structural difference, representation artefacts, or insufficient comparison against known chemical space [36]. Future reporting should distinguish chemical novelty from functional novelty, natural product likeness from biological relevance, predicted activity from measured activity, and safety prediction from safety validation. This distinction will help generative AI mature from a structure-producing technology into a more reliable component of evidence-based natural product-inspired design.

CONCLUSION

Generative AI can expand natural product-inspired molecular design and chemical space exploration by proposing scaffolds, analogues, and prioritisation hypotheses that may support early discovery thinking. Its value, however, depends on chemical realism, stereochemical correctness, synthetic feasibility, bioactivity validation, target relevance, ADMET and toxicity assessment, uncertainty reporting, and experimental confirmation. Generated molecules should not be treated as synthesizable, biologically active, safe, clinically useful, or translationally ready merely because they are valid, novel, natural product-like, or computationally optimised. A rigorous future for generative AI in natural product chemistry will require close integration between model development, cheminformatics, synthetic chemistry, pharmacology, toxicology, and transparent evidence-bound reporting.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, Supuran CT. Natural products in drug discovery: Advances and opportunities. Nat Rev Drug Discov. 2021;20(3):200-16. doi:10.1038/s41573-020-00114-z

[2] Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod. 2020;83(3):770-803. doi:10.1021/acs.jnatprod.9b01285

[3] Sánchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. Science. 2018;361(6400):360-5. doi:10.1126/science.aat2663

[4] Meyers J, Fabian B, Brown N. De novo molecular design and generative models. Drug Discov Today. 2021;26(11):2707-15. doi:10.1016/j.drudis.2021.05.019

[5] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-50. doi:10.1016/j.drudis.2018.01.039

[6] Merk D, Grisoni F, Friedrich L, Schneider G. Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators. Commun Chem. 2018;1:68. doi:10.1038/s42004-018-0068-1

[7] Shen X, Zeng T, Chen N, Li J, Wu R. NIMO: A natural product-inspired molecular generative model based on conditional Transformer. Molecules. 2024;29(8):1867. doi:10.3390/molecules29081867

[8] Sakano K, Furui K, Ohue M. NPGPT: Natural product-like compound generation with GPT-based chemical language models. J Supercomput. 2025;81:352. doi:10.1007/s11227-024-06860-w

[9] Li Y, Zhou X, Liu Z, Zhang L. Designing natural product-like virtual libraries using deep molecule generative models. J Chin Pharm Sci. 2018;27(7):451-9. doi:10.5246/jcps.2018.07.046

[10] Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268-76. doi:10.1021/acscentsci.7b00572

[11] Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. J Cheminform. 2017;9(1):48. doi:10.1186/s13321-017-0235-x

[12] Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018;4(1):120-31. doi:10.1021/acscentsci.7b00512

[13] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv. 2018;4(7):eaap7885. doi:10.1126/sciadv.aap7885

[14] Grisoni F, Merk D, Consonni V, Hiss JA, Giani Tagliabue S, Todeschini R, et al. Scaffold hopping from natural products to synthetic mimetics by holistic molecular similarity. Commun Chem. 2018;1:44. doi:10.1038/s42004-018-0043-x

[15] Brown N, Fiscato M, Segler MHS, Vaucher AC. GuacaMol: Benchmarking models for de novo molecular design. J Chem Inf Model. 2019;59(3):1096-108. doi:10.1021/acs.jcim.8b00839

[16] Gao W, Coley CW. The synthesizability of molecules proposed by generative models. J Chem Inf Model. 2020;60(12):5714-23. doi:10.1021/acs.jcim.0c00174

[17] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018;555(7698):604-10. doi:10.1038/nature25978

[18] Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. ACS Cent Sci. 2019;5(9):1572-83. doi:10.1021/acscentsci.9b00576

[19] Genheden S, Thakkar A, Chadimová V, Reymond JL, Engkvist O, Bjerrum E. AiZynthFinder: A fast, robust and flexible open-source software for retrosynthetic planning. J Cheminform. 2020;12(1):70. doi:10.1186/s13321-020-00472-1

[20] Coley CW, Green WH, Jensen KF. Machine learning in computer-aided synthesis planning. Acc Chem Res. 2018;51(5):1281-9. doi:10.1021/acs.accounts.8b00087

[21] Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: A benchmark for molecular machine learning. Chem Sci. 2018;9(2):513-30. doi:10.1039/C7SC02664A

[22] Yang K, Swanson K, Jin W, Coley CW, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370-88. doi:10.1021/acs.jcim.9b00237

[23] Renz P, Van Rompaey D, Wegner JK, Hochreiter S, Klambauer G. On failure modes in molecule generation and optimization. Drug Discov Today Technol. 2019;32-33:55-63. doi:10.1016/j.ddtec.2020.09.003

[24] Polykovskiy D, Zhebrak A, Sanchez-Lengeling B, Golovanov S, Tatanov O, Belyaev S, et al. Molecular Sets (MOSES): A benchmarking platform for molecular generation models. Front Pharmacol. 2020;11:565644. doi:10.3389/fphar.2020.565644

[25] Urbina F, Lentzos F, Invernizzi C, Ekins S. Dual use of artificial-intelligence-powered drug discovery. Nat Mach Intell. 2022;4(3):189-91. doi:10.1038/s42256-022-00465-9

[26] Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021;49(W1):W5-W14. doi:10.1093/nar/gkab255

[27] Van Vleet TR, Liguori MJ, Lynch JJ 3rd, Rao M, Warder S. Screening strategies and methods for better off-target liability prediction and identification of small-molecule pharmaceuticals. SLAS Discov. 2019;24(1):1-24. doi:10.1177/2472555218799713

[28] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457-i466. doi:10.1093/bioinformatics/bty294

[29] Myung Y, de Sá AGC, Ascher DB. Deep-PK: Deep learning for small molecule pharmacokinetic and toxicity prediction. Nucleic Acids Res. 2024;52(W1):W469-W475. doi:10.1093/nar/gkae254

[30] Schneider P, Altmann KH, Schneider G. Generating bioactive natural product-inspired molecules with machine intelligence. Chimia (Aarau). 2022;76(5):396-401. doi:10.2533/chimia.2022.396

[31] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4

[32] Atz K, Grisoni F, Schneider G. Geometric deep learning on molecular representations. Nat Mach Intell. 2021;3(12):1023-32. doi:10.1038/s42256-021-00418-8

[33] van Tilborg D, Alenicheva A, Grisoni F. Exposing the limitations of molecular machine learning with activity cliffs. J Chem Inf Model. 2022;62(23):5938-51. doi:10.1021/acs.jcim.2c01073

[34] Sakano K, Furui K, Kengkanna A, Kikuchi Y, Ohue M. Chemical language models for natural product discovery. Nat Prod Rep. 2026. Advance article. doi:10.1039/D6NP00002A

[35] Zeng X, Wang F, Luo Y, Kang SG, Tang J, Lightstone FC, et al. Deep generative molecular design reshapes drug discovery. Cell Rep Med. 2022;3(12):100794. doi:10.1016/j.xcrm.2022.100794

[36] Xie S, Zhu H, Huang N. AI-designed molecules in drug discovery, structural novelty evaluation, and implications. J Chem Inf Model. 2025;65(17):8924-33. doi:10.1021/acs.jcim.5c00921