Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

Chemical Novelty without Therapeutic Advantage: A Theoretical Account of How Generative Artificial Intelligence Should Reconcile Molecular Originality, Synthetic Accessibility, Selectivity, and Developability in Medicinal Chemistry

James Whitfield[1], Aisha Rahman[2]*, Kenji Sato[3], Sofia Lindqvist[4]

1. ***Department of Medicinal Chemistry and Generative AI, Warwick Business School, Coventry, United Kingdom.***
2. ***Department of Pharmaceutical Chemistry and Drug Design, Lahore University of Management Sciences, Lahore, Pakistan.***
3. ***Department of Synthetic Chemistry and Molecular Design, Keio University, Tokyo, Japan.***
4. ***Department of Drug Discovery and Development, Stockholm University, Stockholm, Sweden.***

ARTICLE INFO

Received:
20 April 2026
Received in revised form:
23 August 2026
Accepted:
28 August 2026
Available online:
28 October 2026

Keywords: Generative artificial intelligence, De novo molecular design, Synthetic accessibility, Selectivity, Developability, Multiobjective optimization

ABSTRACT

Generative artificial intelligence has expanded the range and speed with which candidate chemical structures can be proposed, but molecular originality is not equivalent to therapeutic advantage. A molecule may be unfamiliar in chemical space while remaining difficult to synthesize, poorly aligned with the intended target profile, insufficiently selective for its pharmacological purpose, or unsuitable for further development. This theoretical article examines how medicinal chemistry should interpret these competing dimensions without collapsing them into a single measure of molecular quality. The analysis distinguishes structural novelty from target relevance, synthetic accessibility, intended selectivity, physicochemical and ADME-related developability, and the reliability of the computational predictions on which these judgments depend. It argues that generative molecular design should be treated as a multiobjective decision problem in which some objectives may trade off, some operate as contextual constraints, and some should remain non-compensatory when their failure invalidates otherwise attractive candidates. The resulting account favors explicit objective-specific evaluation, route-aware synthetic assessment, pharmacology-dependent interpretation of selectivity, prospective developability validation, and reliability-qualified comparison of candidate trade-offs. Therapeutic usefulness therefore depends less on whether generative AI can produce chemically unusual molecules than on whether those molecules remain defensible after medicinal-chemistry constraints are applied.

*This is an **open-access** article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, and build upon the work non commercially.*

To Cite This Article: Whitfield J, Rahman A, Sato K, Lindqvist S. Chemical Novelty without Therapeutic Advantage: A Theoretical Account of How Generative Artificial Intelligence Should Reconcile Molecular Originality, Synthetic Accessibility, Selectivity, and Developability in Medicinal Chemistry. Pharmacophore. 2026;17(5):1-8. https://doi.org/10.51847/z45gBPpBCS

Introduction

Generative artificial intelligence has altered the scale and organization of molecular design by enabling models to propose structures directly from learned chemical representations, target information, or optimized property objectives. Yet the ability to generate candidates quickly does not establish that the generated compounds are medicinally useful. Contemporary drug-design research increasingly treats AI as a component of a wider discovery process rather than as an autonomous source of successful medicines [1]. This distinction becomes especially important when novelty itself is used as evidence of progress. Structural analyses of AI-designed compounds show that apparent originality depends on how novelty is measured and on whether comparison is performed through fingerprints, scaffolds, target-specific precedent, or other structural relationships [2]. Recent reviews of generative molecular design likewise emphasize that constrained generation, multiobjective optimization, uncertainty, and experimental alignment remain central unresolved problems despite substantial advances in model architecture [3].

Corresponding Author: Aisha Rahman; Department of Pharmaceutical Chemistry and Drug Design, Lahore University of Management Sciences, Lahore, Pakistan. E-mail: aisha.rahman@lums.edu.pk.

The resulting difficulty is not simply that some generated molecules are better than others. It is that several conceptually different properties can all be described loosely as indicators of "quality." Generative models can broaden exploration of drug-like chemical space, but wider exploration does not determine whether the sampled molecules have useful biological, synthetic, or developmental characteristics [4]. Even basic evaluation terms such as novelty can conceal substantial variation: scaffold-aware benchmarking captures information that conventional validity and novelty statistics may miss [5]. Structure-based generators introduce further requirements because chemically plausible two-dimensional structures may still fail to reproduce credible three-dimensional ligand geometries or relevant protein–ligand interactions [6]. Molecular originality, scaffold diversity, geometric plausibility, target relevance, and practical usefulness therefore cannot be treated as interchangeable manifestations of one latent property.

This article addresses that problem theoretically rather than proposing another generative architecture. Its central question is how medicinal chemistry should reconcile originality with synthetic accessibility, selectivity, developability, and target-related value when these dimensions do not necessarily move together. The need for such an account follows from the difference between expanding a search space and deciding which outputs deserve experimental commitment [1]. Structural novelty can be characterized systematically [2], but the presence of novelty alone does not determine whether the resulting chemistry represents a productive departure from precedent [5]. The argument developed below therefore treats generative design as a multiobjective medicinal-chemistry problem in which candidate value depends on the configuration of objectives, the reliability of their measurement, and the consequences of allowing improvement on one dimension to compensate for weakness on another.

The Novelty–Therapeutic Value Problem

A first source of confusion arises when optimization performance is interpreted as evidence that a generator has discovered a superior region of chemical space. Goal-directed molecular generators can obtain favorable objective scores while concentrating proposals within relatively narrow regions, and comparative work shows that diversity-sensitive evaluation can alter how such generators are judged [7]. Benchmarking frameworks such as GuacaMol were developed precisely because distribution learning, novelty, validity, and goal-directed optimization describe different capabilities [8]. A model may therefore appear highly successful under one benchmark while offering little information about whether its high-scoring structures are sufficiently diverse, biologically meaningful, synthetically tractable, or developmentally credible.

The problem becomes sharper when three-dimensional and target-conditioned design are considered. Three-dimensional generative methods attempt to incorporate geometric and structural information that is absent from purely sequence- or graph-based novelty measures, but they introduce their own requirements concerning conformational realism, binding-site representation, and structure-conditioned sampling [9]. Conversely, methods that deliberately combine scaffold novelty with predicted drug-like or target-related properties demonstrate that novelty can be optimized alongside other objectives rather than standing in for them [10]. Such work is important because it exposes the inferential error in treating a novel scaffold, a favorable docking result, and a plausible medicinal-chemistry profile as three interchangeable signs of the same underlying advantage.

Chemical novelty should instead be viewed as an analytically separable property whose value depends on what it enables and what it displaces. Scaffold-sensitive assessment shows that different structural metrics can produce materially different judgments about a generator's output [5]. Three-dimensional benchmarking similarly indicates that apparently valid molecules can still fail tests of geometry or protein–ligand interaction quality [6]. Diversity-focused comparisons show that aggressive objective optimization may narrow the range of candidate solutions even when headline scores rise [7]. These findings do not establish that novelty, diversity, and structural plausibility are statistically independent in every dataset. They establish the more limited but medicinally important point that none can be assumed to substitute for the others.

The key theoretical issue is therefore objective displacement. A generated molecule becomes less useful when improvement in one attractive characteristic is purchased by deterioration in another characteristic that matters for experimental progression. A highly original scaffold may have limited value if it lacks a credible synthesis route; an optimized target score may be misleading when the compound exhibits an unsuitable selectivity profile; and attractive in silico properties may have little practical significance if developability predictions are unreliable outside their training domain. Structural novelty studies show why originality requires explicit verification rather than intuitive interpretation [2]. Scaffold-aware benchmarking demonstrates that conventional novelty metrics can overlook relevant structural relationships [5]. Diversity-sensitive evaluation shows that apparent optimization can hide solution-space collapse [7]. Methods that jointly manipulate novelty and drug-like objectives further illustrate that these quantities can be deliberately traded rather than assumed to coincide [10].

The resulting distinction is summarized in **Table 1.** The table does not rank the objectives or assign universal weights. Its purpose is to show why optimizing any one of them in isolation leaves a characteristic medicinal-chemistry failure unresolved.

Table 1. Objective conflicts that prevent molecular originality from functioning as a stand-alone measure of generative-design success

Medicinal-chemistry objective	What apparent optimization success can look like	What remains unresolved when optimized alone	Characteristic failure if treated as sufficient	Qualifying question before candidate progression

Molecular originality	Low similarity to known compounds; unfamiliar scaffold or substitution pattern	Whether the structure is makeable, target-relevant, selective, stable, or developable	Chemically unusual output with no corresponding therapeutic advantage	What medicinally useful property is gained by departing from precedent?
Structural diversity	Broad dispersion across generated structures or scaffolds	Whether diverse outputs satisfy target and development constraints	Diversity for its own sake, including chemically varied but uniformly poor candidates	Does diversity increase the number of credible solution paths or merely the number of distinct structures?
Target-related potency or affinity	High predicted activity or favorable structure-based score	Off-target activity, mechanism relevance, assay transfer, synthesis, ADME behavior	Optimization toward a strong proxy that does not survive downstream testing	Is the predicted target advantage reliable and pharmacologically meaningful in the intended setting?
Synthetic accessibility	Favorable heuristic score or computable retrosynthetic route	Route robustness, precursor availability, operational complexity, scale, or experimental success	"Makeable" molecules whose proposed routes are impractical or fragile	What level of synthetic evidence is sufficient for the decision being made?
Selectivity	Low predicted activity at nominated off-targets or favorable target ratio	Whether selectivity is actually desirable for the intended mechanism	Rejection of useful polypharmacology or acceptance of unmeasured off-target liabilities	What pattern of target engagement is therapeutically intended?
Developability	Favorable predicted physicochemical or ADME profile	Endpoint-specific uncertainty, model drift, formulation and later-stage liabilities	Attractive aggregate drug-likeness despite a decisive weakness in a specific endpoint	Which development property is decision-critical at the present stage?

Reconciling Chemical, Synthetic and Developability Objectives

Synthetic accessibility is the most immediate example of an objective that cannot safely be inferred from molecular novelty or model validity. Comparative evaluation of generated molecules using computer-aided synthesis planning has shown that strong performance on a primary generative objective can coexist with substantial synthetic difficulty and that introducing synthesis-oriented bias may reduce performance on the original objective [11]. This is a genuine multiobjective conflict rather than a cosmetic filtering problem. Route-informed measures such as RAscore attempt to approximate whether a retrosynthesis planner is likely to identify a route, making synthetic feasibility available earlier in virtual screening [12]. Full retrosynthetic systems such as AiZynthFinder add further information by decomposing targets recursively toward purchasable precursors [13]. Neither result should be interpreted as proof that a proposed compound will be synthesized successfully, but both supply evidence that molecular feasibility is richer than a validity check on the generated structure.

Synthetic accessibility is itself multidimensional. A binary distinction between "easy" and "hard" compounds can conceal differences in step burden, route confidence, precursor requirements, and the quality of the transformations available to the planner. Multiclass approaches have therefore been developed to estimate synthetic accessibility at a more granular level than binary prediction [14]. Route-oriented methods such as RetroScore further combine characteristics including route length, prediction confidence, and graph-edit information when evaluating accessibility [15]. The theoretical implication is that medicinal chemistry should resist replacing one scalar simplification with another. A single synthetic-accessibility number can be useful operationally, but it should not erase the distinction between finding any route, finding a plausible route, and finding a route whose chemistry is practical for the project.

The strongest evidence for synthesis-aware generative design arises when candidate generation is connected to molecules that can actually be made and tested. In a CXCR4-focused program, a reinforcement-learning approach constrained generation using reaction templates and available building blocks and linked the resulting compounds to synthesis and experimental evaluation [16]. This type of work does not show that one generative strategy will generalize across targets or chemical series. It demonstrates that makeability can be incorporated before experimental commitment rather than treated as an afterthought. Recent synthesis-centered accounts similarly argue that generative design becomes more actionable when proxy estimates of accessibility are connected to retrosynthetic planning, available chemistry, and eventual laboratory execution [17]. For theoretical purposes, this suggests a useful progression from structurally plausible, to routeable, to practically executable, and finally to experimentally testable. The progression is not a universal maturity scale, but it prevents a favorable synthetic-accessibility proxy from being mistaken for demonstrated chemical actionability.

Selectivity introduces a different kind of constraint because its desirable direction depends on the therapeutic mechanism. Target-aware generation can incorporate biological context directly into molecular proposal and refinement; TamGen, for example, couples a chemical language model to target information and was evaluated through experimental activity in a specific discovery program [18]. At the same time, generative design has been used deliberately to identify multi-target ligands, showing that activity across several targets may be an intended pharmacological objective rather than a failure of specificity [19]. Structure-based diffusion methods that condition generation on pharmacophoric information likewise demonstrate that expert knowledge about the intended binding environment can restrict the search space in ways that generic molecular novelty

cannot [20]. Consequently, "greater selectivity" cannot function as a universal objective. What matters is correspondence between the generated molecule's target-engagement pattern and the pharmacology required by the therapeutic hypothesis.

Developability broadens the problem further because no single descriptor captures the many properties that influence whether a compound can progress. Prospective industrial evaluation of machine-learning models across six in-vitro ADME endpoints found that performance depended on algorithm choice, endpoint, and model-updating strategy across repeated datasets collected over 20 months [21]. The importance of this evidence is not that one algorithmic family should dominate generative design. It is that developability prediction is endpoint-specific and temporally contingent. A candidate that appears satisfactory according to a broad drug-likeness criterion may remain vulnerable to a decisive weakness in permeability, metabolism, clearance, protein binding, or another property that becomes important in its project context.

The reconciliation of these objectives therefore requires progressively stronger forms of evidence as a molecule moves from virtual generation toward experimental commitment. **Table 2** specifies the distinct evaluative jobs performed by novelty, synthesis, selectivity, target relevance, and developability. The entries deliberately separate a minimally informative computational signal from stronger evidence that would justify greater confidence.

Table 2. Evaluation requirements for judging generated molecules across novelty, synthesis, pharmacology, and developability

Evaluation dimension	Minimum informative evidence	Stronger evidence for candidate progression	What the evidence cannot establish by itself	Consequence for medicinal-chemistry judgment
Chemical novelty	Explicit structural comparison using declared representation or scaffold definition	Agreement across complementary similarity, scaffold, and expert structural analyses	Therapeutic superiority, biological relevance, or synthetic feasibility	Treat novelty as a descriptive design property, not a surrogate for value
Synthetic accessibility	Heuristic or learned accessibility estimate	Retrosynthetic route with credible transformations and accessible precursors	Laboratory yield, robustness, scale-up feasibility, cost, or operational simplicity	Increase confidence only as evidence moves from score to route to execution
Target relevance	Target-conditioned generation, docking, pharmacophore fit, or activity prediction	Orthogonal computational support followed by biochemical or cellular confirmation appropriate to the hypothesis	Clinical efficacy or acceptable off-target behavior	Interpret favorable target evidence within the intended mechanism and assay context
Selectivity or intended polypharmacology	Predicted activity profile over nominated targets	Experimentally measured target-engagement pattern across therapeutically relevant comparators	Safety against unmeasured targets or universal desirability of narrow specificity	Define the desired engagement profile before scoring selectivity
Physicochemical and ADME developability	Property predictions from validated models	Prospective endpoint-specific testing, with model performance monitored over time	Complete downstream development success	Treat developability as a vector of decision-relevant endpoints rather than a generic score
Prediction reliability	Model validation within a defined chemical or structural domain	Prospective or external validation on compounds resembling the intended decision setting	Absence of uncertainty or immunity to distribution shift	Qualify every attractive predicted objective by the reliability of the model producing it
Therapeutic relevance	Favorable configuration of the preceding dimensions	Convergent experimental evidence appropriate to the stage of discovery	Eventual clinical benefit	Reserve the strongest claims until downstream evidence warrants them

Taken together, the evidence supports a non-concentric design landscape rather than a hierarchy in which novelty sits at the top. Molecular originality can expand the available search space, but candidate value emerges only where originality intersects with viable synthesis, pharmacologically appropriate target engagement, acceptable development properties, and predictions that remain credible for the region of chemical space being explored. **Figure 1** represents these relationships as overlapping but non-equivalent objective regions so that improvement in one region cannot visually stand in for adequacy in the others.

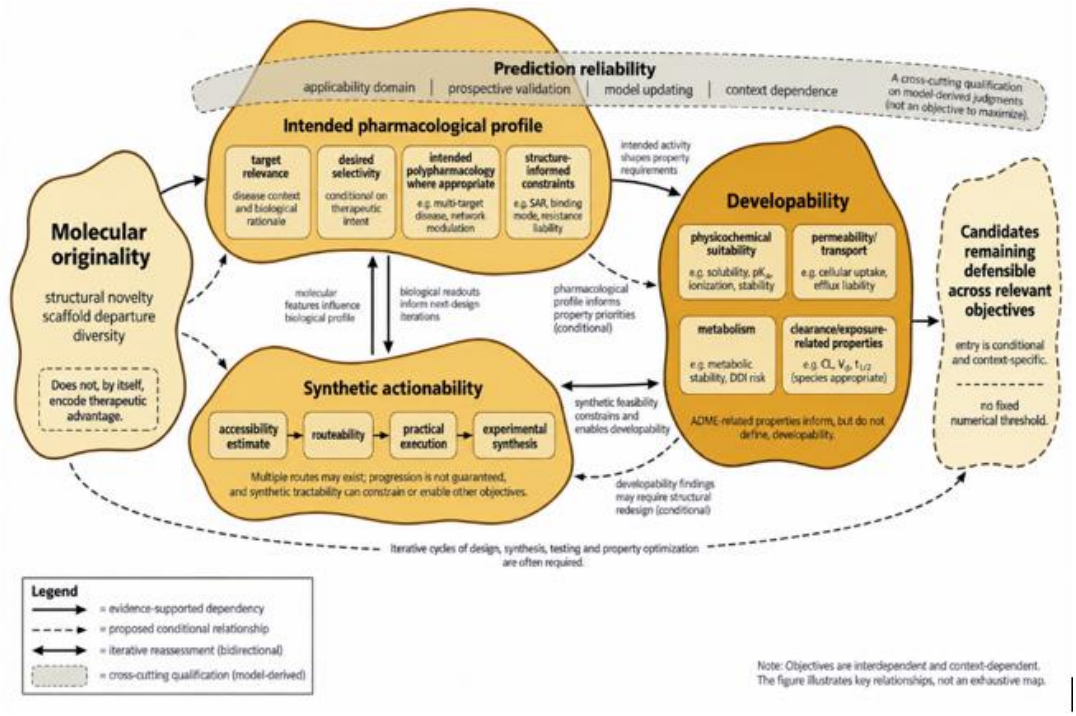


Figure 1. Where molecular originality ceases to be sufficient for medicinal-chemistry value.

A Multiobjective Framework for Generative Molecular Design

The preceding analysis implies that generative molecular design cannot be reduced to maximizing a single notion of molecular quality. Medicinal chemistry routinely confronts properties that improve at different rates, respond differently to structural modification, and sometimes move in opposite directions. Deep reinforcement-learning approaches to multiparameter optimization illustrate that molecular properties with partly orthogonal trends can be addressed simultaneously within a generative process [22]. Monte Carlo tree-search approaches likewise demonstrate that several pharmacological and physicochemical objectives can be incorporated into the search for candidate structures [23]. These methods establish the computational feasibility of multiobjective design, but they do not determine which trade-offs medicinal chemistry should accept. Search and judgment therefore remain distinct analytical operations.

The distinction becomes critical when optimization is performed against predictive models whose reliability changes across chemical space. Multiobjective generators can exploit weaknesses in their scoring functions by proposing molecules with apparently exceptional predicted properties in regions where the underlying predictors are poorly supported. Reliability-aware work on reward hacking shows that restricting optimization according to multiple applicability domains can materially change which candidates appear attractive [24]. An objective value should therefore be interpreted jointly with evidence that the model producing it remains informative for the candidate under consideration. An extreme prediction outside a defensible domain should not compensate automatically for weaknesses elsewhere simply because a weighted reward function assigns it a large numerical contribution.

A second requirement is preservation of explicit trade-offs. Scaffold-aware generative models demonstrate that several objectives can be pursued while maintaining structural constraints relevant to molecular design [25]. Pareto-oriented latent-space optimization provides a complementary principle: candidates can be compared according to whether one solution improves an objective without requiring deterioration in another, rather than according to a permanently fixed composite score [26]. Pareto efficiency does not identify a therapeutically optimal molecule. Its theoretical value lies in keeping conflicts visible. A candidate can be attractive because it occupies a defensible trade-off position while still having clearly stated weaknesses that require experimental resolution.

This article therefore proposes a challenge–revision–retention logic for interpreting generated candidates. A proposed molecule first inherits the objectives used to generate it, but those objectives do not acquire equal status merely because they appear in the same computational reward function. Synthetic feasibility must remain constrained by what route-planning and execution evidence can support [11]. Developability claims remain conditional on endpoint-specific and prospectively validated predictions [21]. High objective values are qualified when the relevant predictor is operating outside a reliable chemical domain [24]. Trade-offs may be examined through Pareto-style comparison when reducing them to a single weight would hide what is gained and what is sacrificed [26]. Candidate retention is therefore conditional: a molecule remains credible only when weaknesses are either acceptable for the intended program or can be revised without destroying the advantage that made the candidate interesting.

Figure 2 represents this decision logic as a field of repeated challenge and revision rather than a linear selection pipeline.

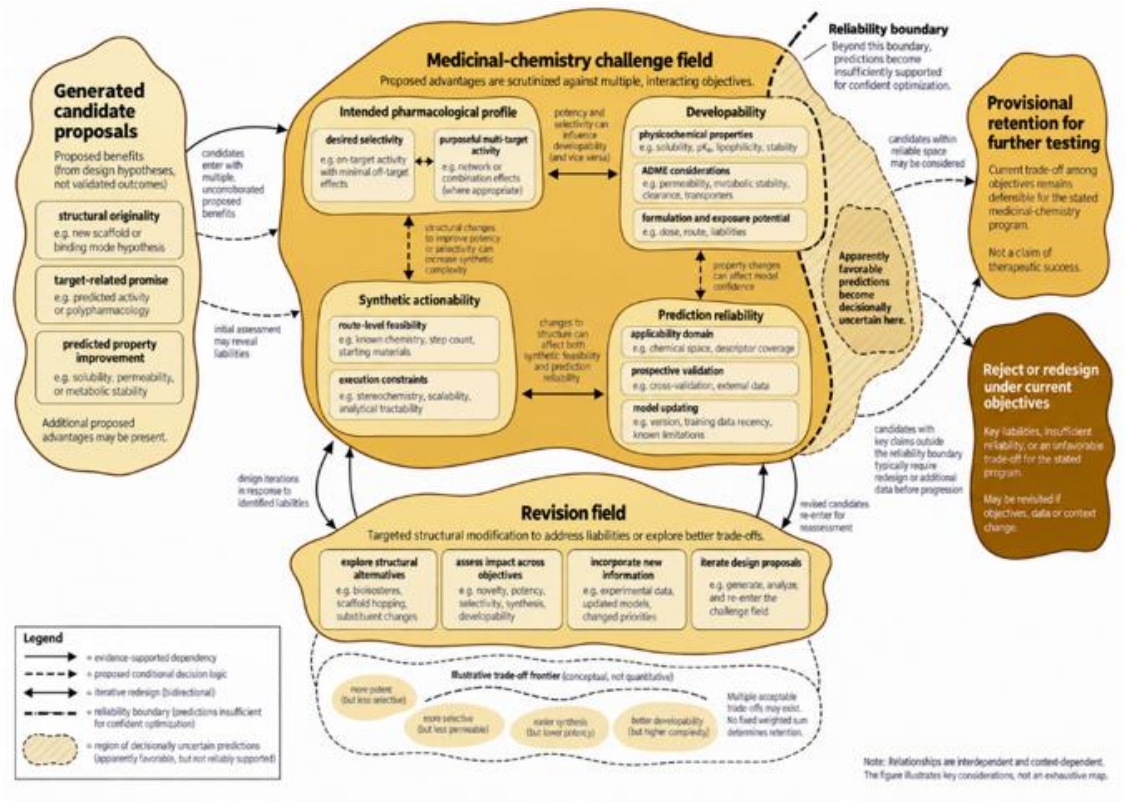


Figure 2. What survives when a generated molecule is challenged by competing medicinal-chemistry objectives.

Implications for Medicinal Chemistry and AI Evaluation

The first implication is that generative-model evaluation should separate dimensions that answer different medicinal-chemistry questions. Scaffold-sensitive evaluation can identify structural relationships that conventional novelty statistics miss.[5] Three-dimensional benchmarking adds information about conformation and protein–ligand interactions that cannot be recovered from two-dimensional validity alone [6]. Synthetic actionability introduces route and execution constraints that structural metrics do not address [17]. Developability requires endpoint-specific evidence whose prospective performance may change with model updating and chemical context [21]. A single leaderboard that compresses these dimensions into one score can therefore obscure why a model succeeds and where it remains unsafe to trust.

Second, prediction reliability should qualify every apparently favorable computational objective. Prospective ADME evaluation demonstrates that predictive performance can differ across endpoints and over time [21]. Reward-hacking analyses show that generative optimization can exploit extrapolative regions in which the objective model is less trustworthy [24]. Evaluations should therefore report not only whether a model generates high-scoring molecules, but whether those scores occur in regions supported by relevant training, validation, or prospective evidence. The practical question is not simply "how high is the predicted value?" but "how much confidence is justified in using that value to make the next medicinal-chemistry decision?"

Third, model rankings should be stress-tested under alternative objective priorities. Multiparameter reinforcement learning [22], multiobjective tree search [23], scaffold-aware generation [25], and Pareto-oriented latent optimization [26] each show that candidate sets depend on how objectives are represented and pursued. The theoretical account developed here predicts that rankings will sometimes reverse when synthetic feasibility, intended selectivity, developability, or prediction reliability receives greater decision weight. Such reversals should not automatically be interpreted as instability of the generative model. They may reveal that an earlier ranking concealed assumptions about what medicinal chemistry was willing to sacrifice.

Finally, the relevant standard for generative AI should shift from how novel a molecule appears at generation to which medicinal-chemistry challenges it can survive without its original advantage becoming irrelevant. Novelty may justify investigation, particularly when known chemical space is exhausted or structurally constrained. It cannot by itself justify progression. A useful generative model should therefore expose the trade-offs it creates, make clear when predictions exceed their reliable domains, and support comparison among candidates without implying that every property is freely compensatory.

Conclusion

Chemical novelty is an important capability of generative artificial intelligence, but it is not a therapeutic endpoint. A generated molecule becomes medicinally meaningful only when its originality remains compatible with feasible chemistry, the intended pharmacological profile, development-relevant properties, and sufficiently reliable evidence for the predictions used to judge it. These dimensions should not be compressed automatically into a universal scalar definition of molecular quality.

The appropriate role of multiobjective generative design is therefore to make trade-offs explicit rather than to make them disappear. Synthetic accessibility may invalidate an otherwise attractive proposal; intended polypharmacology may make narrow selectivity undesirable; a favorable developability prediction may require prospective confirmation; and an apparently exceptional objective value may lose significance outside a reliable prediction domain. The strongest candidates are consequently not those that maximize novelty, nor necessarily those that maximize a fixed aggregate reward, but those whose advantages remain defensible when competing medicinal-chemistry requirements are examined together.

Generative AI can expand what medicinal chemists are able to propose. Its therapeutic value depends on whether the resulting molecules withstand the constraints that make a chemical idea experimentally actionable and developmentally credible.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA Jr, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-64. doi:10.1038/s41573-019-0050-3
2. 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
3. Sun Y, Chen L, Jing Z, Li YY, Kim D, Gao JY, et al. Generative AI for the design of molecules: Advances and challenges. J Chem Inf Model. 2025;65(23):12668-90. doi:10.1021/acs.jcim.5c02234
4. Lavecchia A. Navigating the frontier of drug-like chemical space with cutting-edge generative AI models. Drug Discov Today. 2024;29(9):104133. doi:10.1016/j.drudis.2024.104133
5. Fil V, Van Den Broek RL, Šícho M, Čmelo I, Agea MI, Jespers W, et al. Scaffold-based evaluation metrics for fair comparison of molecular generators. J Cheminform. 2026;18(1):84. doi:10.1186/s13321-026-01213-6
6. Sanjrani N, Coupry DE, Pogány P, Palmer DS, Pickett SD. Benchmarking 3D structure-based molecule generators. J Chem Inf Model. 2025;65(15):8006-21. doi:10.1021/acs.jcim.5c01020
7. Renz P, Luukkonen S, Klambauer G. Diverse hits in de novo molecule design: Diversity-based comparison of goal-directed generators. J Chem Inf Model. 2024;64(15):5756-61. doi:10.1021/acs.jcim.4c00519
8. 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
9. Xie W, Wang F, Li Y, Lai L, Pei J. Advances and challenges in de novo drug design using three-dimensional deep generative models. J Chem Inf Model. 2022;62(10):2269-79. doi:10.1021/acs.jcim.2c00042
10. Wang M, Wu Z, Wang J, Weng G, Kang Y, Pan P, et al. Genetic algorithm-based receptor ligand: A genetic algorithm-guided generative model to boost the novelty and drug-likeness of molecules in a sampling chemical space. J Chem Inf Model. 2024;64(4):1213-28. doi:10.1021/acs.jcim.3c01964
11. 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
12. Thakkar A, Chadimová V, Bjerrum EJ, Engkvist O, Reymond JL. Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning. Chem Sci. 2021;12(9):3339-49. doi:10.1039/D0SC05401A
13. 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
14. Li X, Walsh R, Abbas W, Pascual-Diaz S, Hand C, Garland R, et al. Multiclass synthetic accessibility prediction. J Chem Inf Model. 2025;65(3):1155-65. doi:10.1021/acs.jcim.4c01663
15. Gao S, Zhou X, Liang L, Lin J. RetroScore: Graph edit distance-guided retrosynthesis for accessibility scoring with route metrics. J Cheminform. 2026;18(1):10. doi:10.1186/s13321-025-01138-6
16. Jiang X, Lu L, Li J, Jiang J, Zhang J, Zhou S, et al. Synthetically feasible de novo molecular design of leads based on a reinforcement learning model: AI-assisted discovery of an anti-IBD lead targeting CXCR4. J Med Chem. 2024;67(12):10057-75. doi:10.1021/acs.jmedchem.4c00184
17. Simonetti SO, Moroni AB, Zanardi MM, Kozielski F, Wells G, Porta EOJ. From chemical space to synthetic reality: A framework for makeable-by-design generative AI in drug discovery. Drug Discov Today. 2026;31(6):104794. doi:10.1016/j.drudis.2026.104794
18. Wu K, Xia Y, Deng P, Liu R, Zhang Y, Guo H, et al. TamGen: Drug design with target-aware molecule generation through a chemical language model. Nat Commun. 2024;15(1):9360. doi:10.1038/s41467-024-53632-4

19. Isigkeit L, Hörmann T, Schallmayer E, Scholz K, Lillich FF, Ehrler JHM, et al. Automated design of multi-target ligands by generative deep learning. Nat Commun. 2024;15(1):7946. doi:10.1038/s41467-024-52060-8
20. Ziv Y, Imrie F, Marsden B, Deane CM. MolSnapper: Conditioning diffusion for structure-based drug design. J Chem Inf Model. 2025;65(9):4263-73. doi:10.1021/acs.jcim.4c02008
21. Fang C, Wang Y, Grater R, Kapadnis S, Black C, Trapa P, et al. Prospective validation of machine learning algorithms for absorption, distribution, metabolism, and excretion prediction: An industrial perspective. J Chem Inf Model. 2023;63(11):3263-74. doi:10.1021/acs.jcim.3c00160
22. Ståhl N, Falkman G, Karlsson A, Mathiason G, Boström J. Deep reinforcement learning for multiparameter optimization in de novo drug design. J Chem Inf Model. 2019;59(7):3166-76. doi:10.1021/acs.jcim.9b00325
23. Suzuki T, Ma D, Yasuo N, Sekijima M. Mothra: Multiobjective de novo molecular generation using Monte Carlo tree search. J Chem Inf Model. 2024;64(19):7291-302. doi:10.1021/acs.jcim.4c00759
24. Yoshizawa T, Ishida S, Sato T, Ohta M, Honma T, Terayama K. A data-driven generative strategy to avoid reward hacking in multi-objective molecular design. Nat Commun. 2025;16(1):2409. doi:10.1038/s41467-025-57582-3
25. Dong T, You L, Chen CYC. Multi-objective drug design with a scaffold-aware variational autoencoder. Chem Sci. 2025;16(29):13352-67. doi:10.1039/D4SC08736D
26. Abeer ANMN, Urban NM, Weil MR, Alexander FJ, Yoon BJ. Multi-objective latent space optimization of generative molecular design models. Patterns (N Y). 2024;5(10):101042. doi:10.1016/j.patter.2024.101042